

Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions

Joel G. Kingsolver · Sarah E. Diamond · Adam M. Siepielski ·
Stephanie M. Carlson

Received: 21 November 2011 / Accepted: 13 February 2012
© Springer Science+Business Media B.V. 2012

Abstract There are now thousands of estimates of phenotypic selection in natural populations, resulting in multiple synthetic reviews of these data. Here we consider several major lessons and limitations emerging from these syntheses, and how they may guide future studies of selection in the wild. First, we review past analyses of the patterns of directional selection. We present new meta-analyses that confirm differences in the direction and magnitude of selection for different types of traits and fitness components. Second, we describe patterns of temporal and spatial variation in directional selection, and their implications for cumulative selection and directional evolution. Meta-analyses suggest that sampling error contributes importantly to observed temporal variation in selection, and indicate that evidence for frequent temporal changes in the direction of selection in natural populations is limited. Third, we review the apparent lack of evidence for widespread stabilizing selection, and discuss biological and methodological explanations for this pattern. Finally, we describe how sampling error, statistical biases, choice of traits, fitness measures and selection metrics, environmental covariance and other factors may limit the inferences we can draw from analyses of selection coefficients. Current

Electronic supplementary material The online version of this article (doi:[10.1007/s10682-012-9563-5](https://doi.org/10.1007/s10682-012-9563-5)) contains supplementary material, which is available to authorized users.

J. G. Kingsolver (✉)
Department of Biology, University of North Carolina, Chapel Hill, NC 27516, USA
e-mail: jgking@bio.unc.edu

S. E. Diamond
Department of Biology, North Carolina State University, Raleigh, NC 27695, USA
e-mail: sediamon@ncsu.edu

A. M. Siepielski
Department of Biology, University of San Diego, San Diego, CA 92110, USA
e-mail: adamsiepielski@sandiego.edu

S. M. Carlson
Department of Environmental Science, Policy, and Management, University of California, Berkeley,
CA 94720, USA
e-mail: smcarlson@berkeley.edu

standardized selection metrics based on simple parametric statistical models may be inadequate for understanding patterns of non-linear selection and complex fitness surfaces. We highlight three promising areas for expanding our understanding of selection in the wild: (1) field studies of stabilizing selection, selection on physiological and behavioral traits, and the ecological causes of selection; (2) new statistical models and methods that connect phenotypic variation to population demography and selection; and (3) availability of the underlying individual-level data sets from past and future selection studies, which will allow comprehensive modeling of selection and fitness variation within and across systems, rather than meta-analyses of standardized selection metrics.

Keywords Fitness · Meta-analysis · Natural selection · Phenotypic selection

Introduction

Natural and sexual selection are the primary mechanisms that cause adaptive evolution within natural populations (Darwin 1859). Despite the centrality of selection to Darwin's theory of evolution, he never quantified selection in the wild. In the century following the publication of *The Origin of Species*, selection was generally regarded as too weak to be observed directly in natural populations (but see Bumpus (1899) and Weldon (1901) for early exceptions).

Studies in the past four decades have shown that phenotypic selection—differences in fitness (or its components) associated with phenotypic variation among individuals—in natural populations can be stronger and more dynamic than Darwin and other early evolutionary biologists imagined. Following the development of standard methods for detecting and quantifying selection, and a theoretical framework for selection and evolution of multiple phenotypic traits, the number of estimates of selection on quantitative traits in natural populations has increased more than tenfold in the past 25 years (Lande 1979; Lande and Arnold 1983; Endler 1986). Selection has now been detected in hundreds of populations in nature, with thousands of estimates of the strength and form of selection on numerous phenotypic traits (Kingsolver et al. 2001; Siepielski et al. 2009).

In the past decade, this research has produced several synthetic analyses that explore patterns of selection in the wild. Here we summarize the general lessons and surprises that have emerged from these analyses, and present new meta-analyses to evaluate the statistical support for these previous results. Second, we describe some of the important limitations of current methodologies and frameworks for quantifying selection, and how these alter our interpretation of past analyses and syntheses of selection. Finally, we identify promising directions for new field studies, modeling approaches, and syntheses that will be essential for deepening our understanding of selection in the wild.

Lessons from past and new analyses

Metrics of phenotypic selection

The work of Lande, Arnold and Wade provided both a theoretical framework for selection and evolution of multiple, quantitative traits, and simple methods for estimating selection (Lande 1979; Lande and Arnold 1983; Arnold and Wade 1984). These methods focus on linear and quadratic components of selection, using regression models to relate variation in

(potentially correlated) traits to variation in relative fitness (or fitness components: see below) among individuals in a population. The linear (s) and quadratic (g) *selection differentials* estimate the total selection on the trait via simple univariate regression, reflecting both direct selection on the trait and indirect selection due to correlated (and measured) traits. The linear (β) and quadratic (γ) *selection gradients* estimate selection directly on the trait of interest, controlling statistically (via partial regression coefficients from a multiple regression) for indirect selection due to correlated traits. The linear coefficients quantify directional selection on the trait, whereas the quadratic coefficients quantify the curvature of the fitness function near the population mean trait value. These latter regressions also allow a test of correlational selection (e.g., selection for or against trait combinations) by considering the sign and magnitude of bivariate coefficients. To allow comparisons among different types of traits and organisms, one can standardize selection gradients and differentials by the amount of variation in the trait (i.e., by the standard deviation) to obtain standardized measures of selection (Arnold and Wade 1984; Lande and Arnold 1983). Unless otherwise noted, for simplicity we use the terms selection gradients and selection differentials to refer to these variance-standardized measures throughout the paper.

These standardized measures can be viewed as metrics of effect size, and have served as the basis for multiple synthetic analyses of selection. These analyses have focused on studies of selection that consider natural variation in quantitative phenotypes within populations under natural field conditions. Of course, these criteria exclude many excellent and informative studies that use phenotypic manipulations, experimental genetic lines, or controlled environmental conditions. Recent reviews have tabulated more than 4,500 distinct estimates of phenotypic selection from more than 140 field studies that meet the criteria listed above (Siepielski et al. 2009; Kingsolver and Diamond 2011). The open availability of these datasets has also enabled a variety of additional analyses (Hereford et al. 2004; Hersch and Phillips 2004; Morrissey and Hadfield 2012).

Patterns of directional selection

Reviews of phenotypic selection have documented several major patterns. First, there is abundant evidence for directional selection on morphology and life history (Endler 1986; Kingsolver et al. 2001; Hereford et al. 2004). The magnitude of directional selection (e.g., $|\beta|$) varies widely: $|\beta|$ follows an exponential distribution, with a long tail of values exceeding 0.5, and a median of ~ 0.16 (but see next section). As a result, the magnitude of directional selection is sufficient to produce rapid microevolutionary changes in many populations (Grant and Grant 1989; Hendry and Kinnison 1999; Kingsolver and Pfennig 2007; Reznick and Ghalambor 2001).

Second, there are differences in the direction and magnitude of directional selection for different types of phenotypic traits. For example, directional selection on size is significantly shifted towards positive values, whereas selection on phenology (seasonal timing) is significantly shifted towards negative values. These qualitative patterns hold both within and across studies, and for different taxa and fitness components. This suggests that there is directional selection for increasing size and for earlier phenological timing in many study systems (Kingsolver and Pfennig 2004; Kingsolver 2009). However, it is unclear how much of this pattern is due to conditioning or environmental covariation: individuals in better body condition or better microenvironments may be larger, reproduce sooner and have greater reproductive success, inflating the magnitude of selection on size and timing (see “[Environmental covariance](#)”).

Third, the magnitude of directional selection differs among fitness components. In particular, median estimates of $|\beta|$ and $|\text{s}|$ are significantly larger for selection via fecundity and mating success than for selection via survival. Again, this pattern appears to hold both within and across different studies, and for different trait types. This suggests that viability selection may typically be weaker than other components of selection in most natural populations (Hoekstra et al. 2001; Siepielski et al. 2011). However, demographic analyses of elasticities indicate that fitness (e.g., λ , the asymptotic rate of increase) depends more strongly on fecundity than on survival in field studies across a range of taxa and study systems (Crone 2001), emphasizing the importance of demographic context in understanding selection (see “Future directions: what should we do now?”).

Fourth, comparisons of linear selection gradients (β) and differentials (s) show that for most traits, β and s values are similar, suggesting that indirect selection due to correlated traits generally has modest effects on the total selection that traits experience. In contrast, total selection (s) is typically smaller than direct selection (β) for studies of size, suggesting that indirect selection on correlated traits may reduce the total selection on size in many populations. We note that the criteria for choosing traits used by researchers may influence this pattern. In particular, if researchers select traits to reduce the well-known problems of inference and interpretation caused by strong correlations (Lande and Arnold 1983; Mitchell-Olds and Shaw 1987), this will reduce the inferred importance of indirect selection. Indeed, many researchers have used principal components analyses to produce orthogonal axes of trait combinations and estimate selection based on the resulting PC scores as a means of avoiding the difficulties associated with correlated traits.

Meta-analyses of selection

Formal meta-analyses of selection have previously been difficult to perform on available datasets of phenotypic selection due to the limited number of studies which report standard errors of selection coefficients, and the difficulty in accounting for potential autocorrelation among estimates of selection based on spatial, temporal, and phylogenetic similarity (Gurevitch and Hedges 1999). However, recent advances and statistical models have made such analyses more feasible (e.g., Hadfield 2010), facilitating meta-analyses that explicitly account for effects of sampling error on the magnitude and patterns of selection (Morrissey and Hadfield 2012).

We have re-analyzed a dataset of estimates of phenotypic selection in natural populations (combined datasets from Kingsolver et al. 2001 and Siepielski et al. 2009; see Kingsolver and Diamond 2011 for details) with a formal meta-analysis for both linear and quadratic estimates of selection. We were restricted to using the subset of the full dataset that included standard errors for the selection metrics (e.g., for β , $n = 1396$, or 49% of available estimates; for γ , $n = 686$, or 55%). We conducted random-effect meta-analyses in a Bayesian modeling framework using the package MCMCglmm (Hadfield 2010) in R (version 2.13.0; R Development Core Team 2011). This framework allowed us to account for both the effects of sampling error and study- and species-level autocorrelation. We were unable to incorporate phylogeny into the meta-analysis (e.g., see Hadfield and Nakagawa 2010) because of the widely heterogeneous taxa in our dataset. To explicitly address the influence of sampling error and autocorrelation, we compared models of selection estimates that do and do not include these effects (see Supplemental Appendix for model details).

We also considered models that included several moderator variables (see Kingsolver and Diamond 2011): trait type (size, non-size morphology, phenology, or non-phenology

life history); fitness component (survival, mating success, fecundity, or total fitness); and taxonomic group (plant, invertebrate, or vertebrate). Preliminary analyses that included year of publication (Simmons et al. 1999; Poulin 2000) as a continuous or categorical moderator resulted in a poorer model fit, so year of publication was not included as a moderator. We modeled variation in both linear selection (β and s) and quadratic (γ and g) estimates of selection, as well as in the magnitude of selection (e.g. $|\beta|$ and $|\gamma|$). The magnitude (absolute value) of selection coefficients follows a folded normal distribution (Hereford et al. 2004; Morrissey and Hadfield 2012). Our meta-analyses of the magnitude of selection incorporate the folded normal distribution (see Supplemental Appendix), and so unless indicated otherwise, subsequent textual references to, e.g., $|\beta|$ indicate the magnitude of β under the folded normal distribution.

In general, our formal meta-analyses revealed comparable results to previous visual and non-parametric analyses of linear and quadratic selection estimates. Overall, estimates of linear selection were marginally right-shifted, while overall estimates of quadratic selection were essentially centered about zero (Figs. 1a, 2; Tables S1-S2). Among trait types, we found that directional selection (β) for body size was right-shifted (towards positive values), while phenology was left-shifted (negative values); other morphological and life history traits were closer to zero (Fig. 1b). Among fitness components, we found that the magnitude of directional selection ($|\beta|$) via survival was relatively weak compared with fecundity and mating success; total fitness also tended to be stronger than viability selection, but was quite variable, likely owing to the comparatively few available estimates via total fitness (Fig. 1c). Morrissey and Hadfield (2012) reported a similar result using a smaller dataset. Little consistent variation among taxonomic groups was detected in the magnitude of directional selection, though plants tended to experience stronger directional selection compared with vertebrates and invertebrates (Fig. 1d; Table S1). These meta-analytic results generally confirm patterns reported in previous synthetic analyses (Kingsolver and Diamond 2011; Kingsolver et al. 2001; Kingsolver and Pfennig 2004; Siepielski et al. 2009; Siepielski et al. 2011).

Finally, we found that failure to account for sampling error and study- and species-level autocorrelation can greatly influence estimates of linear and quadratic selection, especially for the magnitude of selection ($|\beta|$ and $|\gamma|$) (Fig. 2) (see Hereford et al. 2004).

Patterns of quadratic selection

The quadratic selection gradient γ reflects the curvature of the fitness surface near the population mean phenotype: negative values of γ are necessary (but not sufficient) for stabilizing selection, whereas positive values of γ are necessary (but not sufficient) for disruptive selection. If mean trait values in most populations are near local fitness peaks, we would predict estimates of γ to be strongly shifted towards negative values (Estes and Arnold 2007).

There are clear demonstrations of stabilizing selection in some natural systems (e.g., Weis and Abrahamson 1986; Egan et al. 2011). However, the available evidence from natural populations indicates that positive and negative values of γ are equally common, with an overall mean and median near zero (Kingsolver and Diamond 2011). This pattern holds for different fitness components and for most trait types. One notable exception is for body size, which shows a marked shift (>65% of reported γ estimates) towards negative values; implying the operation of stabilizing selection on body size. Joint estimates of β and γ for size-related traits are consistent with stabilizing selection on size in many

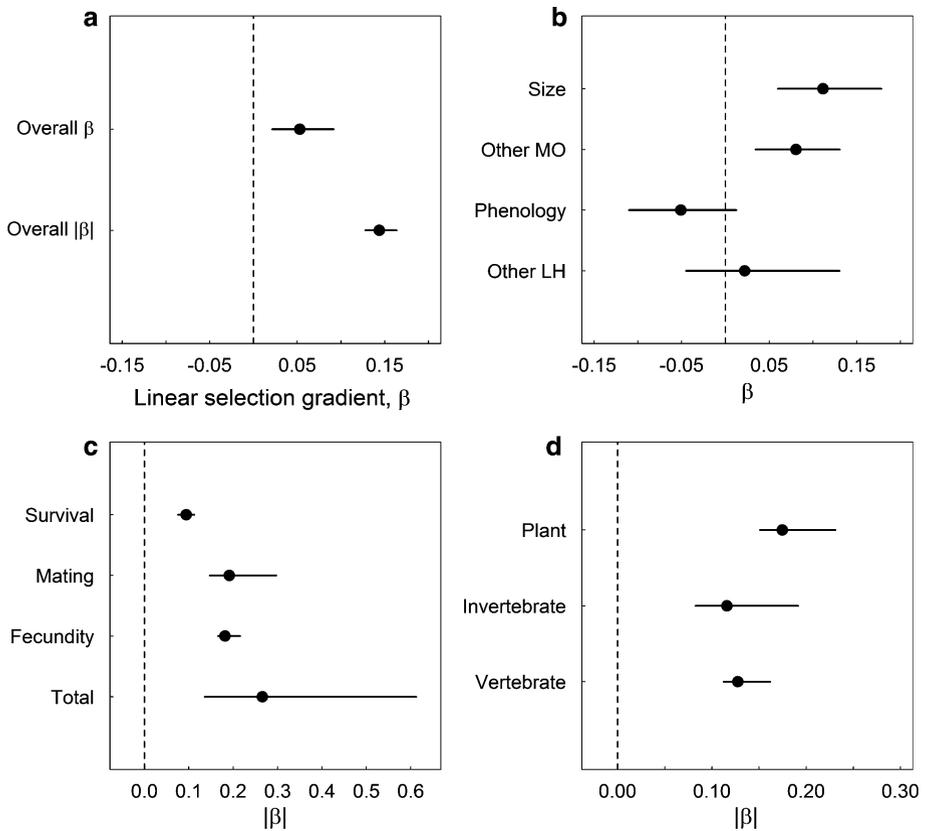


Fig. 1 Summary of linear selection gradients (β), presented as posterior modes and 95% credible intervals; estimates account for sampling error and study- and species-level autocorrelation. **a** Overall β and overall $|\beta|$ (magnitude), **b** β as a function of the trait type moderator, **c** β as a function of the fitness component moderator, and **d** β as a function of the taxonomic group moderator. Overall and moderator-specific sample sizes: overall $\beta = 1,341$; trait class: size = 176, other morphology = 566, phenology = 338, other life history = 261; fitness component: survival = 764, mating success = 221, fecundity = 309, total fitness = 47; taxonomic group: plant = 190, invertebrate = 176, vertebrate = 975. There are substantial differences in the posterior mode for β among different types of traits and in the posterior mode for $|\beta|$ among fitness components (see text). See Table S1 for actual values

populations. Reported estimates of γ vary widely between -1 and 1 (with more than 80% between -0.5 and 0.5), but standard errors on most estimates are quite large (see below). In addition many published values underestimate the magnitude of $|\gamma|$ by a factor of two (Stinchcombe et al. 2008). But for most traits in most studies, the available estimates suggest that stabilizing selection is no more common than disruptive selection. Our formal meta-analyses produced similar results to the previous synthetic analyses: the posterior mode of estimates of quadratic selection was near zero (Fig. 2; Table S2).

As stabilizing selection is widely anticipated theoretically (Estes and Arnold 2007), but underrepresented empirically, this begs the question of what we are missing, either in terms of theory or empirically. One possibility is that there are considerable constraints on adaptation. If populations are relatively far from their fitness optima, this could suggest potentially important roles for tradeoffs among fitness components (e.g., mating success

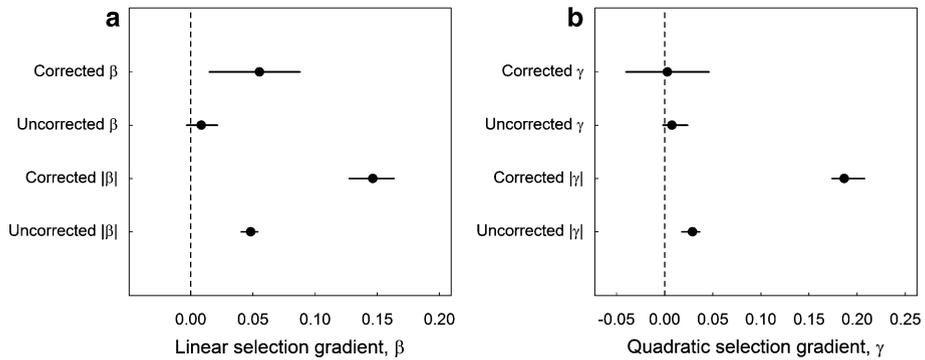


Fig. 2 Effects of performing formal meta-analyses versus descriptive statistics (using kernel density estimation) on linear and quadratic selection gradients (β ($n = 1,341$) and γ ($n = 686$), respectively). “Corrected” estimates are obtained from formal meta-analyses which account for study- and species-level autocorrelation, and are presented as posterior modes with 95% credible intervals; note that corrected estimates of the magnitude of selection are applied to the folded normal distribution. “Uncorrected” estimates are presented as modes of the kernel density estimate with 95% confidence intervals (obtained via bootstrapping, $n = 10,000$). **a** Corrected and uncorrected values of β and $|\beta|$; and **b** Corrected and uncorrected values of γ and $|\gamma|$. There are substantial differences among corrected and uncorrected values of linear and quadratic selection gradients (see text). See Table S2 for actual values

may come at the cost of viability), indirect selection (correlations among traits may reduce total selection), or temporal fluctuations in selection. However, except in a few cases (e.g., indirect selection on body size), these factors were found to impose few limitations on directional selection (Kingsolver and Diamond 2011). In addition, small sample size and/or the limited number of individuals with extreme phenotypic values in samples of natural populations may make it difficult to detect stabilizing selection. Finally, a failure to consider the influence of nonlinear correlational selection acting on pairs of traits can lead to underestimation of the strength of nonlinear selection (though this can be addressed through canonical analysis; see below, and Blows and Brooks 2003). In any case, further evidence quantifying the frequency and strength of stabilizing selection in natural populations is greatly needed.

Selection dynamics in time and space

Natural selection has the potential to vary within populations through time and among populations distributed in space. These two dynamics of selection (coupled with drift, gene flow, and mutation, which may also vary temporally or spatially) have an important role in shaping the evolutionary trajectories of populations. For example, temporal variation in selection may limit divergence, whereas spatial variation in selection may facilitate adaptive divergence. What do phenotypic selection analyses tell us about major patterns of temporal and spatial variation of selection in the wild?

Siepielski et al. (2009, 2011) recently summarized estimates of temporal (annual) variation in selection, compiling 5,519 estimates of β and γ from 89 studies of selection within populations that spanned two or more years. Their analyses suggested substantial variation in the estimated magnitude of selection for both linear and non-linear forms of selection, although much of the variation is due to sampling error (see below). Moreover, the authors report that estimates of the mean and standard deviation of linear selection

gradients were strongly correlated suggesting that there is greater temporal variation in selection for traits that are under stronger average selection, but sampling error could also contribute to this effect. Second, variation in the strength of linear selection also differed among fitness components. Specifically, estimated linear selection through mating success was both stronger (on average) and more variable in strength, compared with selection through either survival or fecundity (Siepielski et al. 2011). Third, changes in the estimated direction of selection were apparently frequent (Siepielski et al. 2009; Siepielski et al. 2011), suggesting the possibility of temporally fluctuating selection in some populations. To explore the evolutionary implications of these patterns, Kingsolver and Diamond (2011) computed how variation (including changes in sign) in selection affected the cumulative selection differential over time. They found little indication that changes in the direction of selection would affect the overall strength of selection that a trait experiences for most populations (Morrissey and Hadfield 2012; see also Bell 2010).

As illustrated above (Figs. 1, 2), sampling error can contribute substantially to variation in selection estimates, especially in studies with small sample sizes (e.g., Siepielski et al. 2009; Siepielski et al. 2011; Hersch and Phillips 2004; Kingsolver et al. 2001). How can we distinguish real temporal variation in selection on a trait from sampling error? Siepielski et al. (2009) tackled this question using variance components analysis, and estimated that the average fraction of total variation that can be attributed to sampling error varies widely within and among studies (average fraction among linear and quadratic gradients and differentials: 0.63–0.92). Morrissey and Hadfield (2012) recently analyzed a subset of the Siepielski et al. (2009) dataset to address this same question, conducting meta-analyses that incorporate sampling error to estimate both between-trait and within-trait (i.e., temporal) variance components for linear selection gradients. Their analyses suggest that ~90% of the temporal variation in directional selection estimates may be accounted for by sampling error alone. Similarly, they concluded that many, perhaps most, apparent changes in the direction of selection can be accounting for by sampling error (Morrissey and Hadfield 2012; see also Siepielski et al. 2011).

Given these results, how important is temporal variation in directional selection? To address this, Morrissey and Hadfield (2012) proposed a consistency or repeatability metric for selection gradients, representing the proportion of total (between-trait plus within-trait) variation explained by between-trait variation. Their estimated consistency values were quite high, typically exceeding 85%. Because within-trait variance is a small component of the total variance, they concluded that the magnitude and direction of selection is 'remarkably constant over time'. However, this interpretation of consistency has several potential limitations. First, as described above, there are important differences in the mean magnitude and direction of selection for different trait types (e.g., Fig. 1). In addition, differences in the direction of selection can arise depending on arbitrary trait definitions (e.g., development time vs. development rate). Similarly, differences in selection for different species and study sites will contribute to the apparent between-trait variation. Each of these factors will increase the estimated between-trait variation, and thus increase the consistency index as defined by Morrissey and Hadfield (2012). The observation that within-trait variation is relatively small compared to the sum of these other factors does not necessarily imply that directional selection is remarkably constant. However, as Morrissey and Hadfield's (2012) analyses suggest, there is little empirical support for the importance of fluctuating selection—temporal changes in the direction of selection—in most natural populations (see also Kingsolver and Diamond 2011), despite clear demonstrations of fluctuating selection in some systems (e.g., Darwin's finches, Grant and Grant 2002).

Spatial variation in environmental features is a ubiquitous feature of the landscape, suggesting that selection should also vary spatially. Spatial variation in selection is key to many evolutionary patterns and processes, including coevolution, geographic variation, local adaptation, maintenance of genetic variation, and speciation (Thompson 2005). A recent meta-analysis of reciprocal transplant experiments suggests that local adaptation of geographic populations is widespread (Hereford 2009), but the traits underlying such adaptation are typically unknown. In addition, there may be common features of the environment that correlate with spatial variation in selection, including both abiotic and biotic factors (Thompson 2005; Urban 2011).

We are currently in the process of assembling a database of all spatially replicated studies (Gotanda, DiBattista, Carlson, and Siepielski, unpub. data). With this database, we hope to address a number of outstanding questions concerning the spatial dynamics of selection. For example, does selection often vary in direction, form, and strength among populations? What is the magnitude of variation in selection among populations? How does spatial variation in selection compare to temporal variation? Our preliminary results suggest that variation in selection among populations is stronger than temporal variation in selection within populations, which would accentuate population divergence. Quantifying the spatial scales over which selection varies will be essential for understanding geographic variation, local adaptation, and evolutionary responses to environmental changes in natural populations.

Limitations

The thousands of standardized estimates summarized above provide a useful window into patterns of selection in nature. But there are important limitations to the Lande and Arnold (1983) methodological approach and how it has been applied in most selection studies, which impacts our interpretations of these patterns. These limitations also suggest alternative approaches and practices that may improve our future understanding of selection. We highlight four main issues: the choice of traits, fitness measurements, and selection metrics; availability of key data and information; the effects of environmental covariance; and characterizing non-linear components of selection.

Choice of traits, fitness, and selection metrics

Organisms are complex integrations of multiple traits. Yet most studies of selection consider only a few (typically 1–4) traits. A key question for interpretation, common to all regression analyses, is whether a measured trait is the actual target of selection, or whether other correlated, but unmeasured, traits are the true targets of selection (Lande and Arnold 1983; Endler 1986; Mitchell-Olds and Shaw 1987). This problem is most acute for interpreting selection gradients, which estimate direct selection on an individual trait after statistically controlling for the effects of correlation of other measured traits. To address this issue, researchers might consider measuring selection on suites of traits. However, incorporating large numbers of traits into selection analyses will reduce statistical power and increase the problems with colinearity among traits (Mitchell-Olds and Shaw 1987). Principal components (PC) analyses and scores are often used to address this issue, but PCs can suffer from problems of biological interpretation (Mitchell-Olds and Shaw 1987). Subsampling of traits for studies that incorporate a larger number of traits could be useful in exploring how the choice and number of traits influence estimates of selection gradients.

In addition to logistical decisions regarding the number and types of traits to include in analyses, researchers are also confronted with similar questions regarding fitness components. The underlying goal of studies of selection is to evaluate how variation in a particular phenotype affects fitness and yet the vast majority of selection estimates ($\sim 98\%$) reflect individual fitness components rather than more integrated measures of total fitness (Kingsolver et al. 2001; Siepielski et al. 2009). The majority of studies focus on selection through either survival or mating success, often using just one fitness component (Siepielski et al. 2011), and many studies quantify selection during only one part of the life cycle. Such studies implicitly assume that individual fitness components are correlated with true fitness, an assumption that is rarely evaluated (McPeck 2010; Thompson et al. 2011). When they analyzed data from studies that quantified selection on the same trait using more than one fitness component, Kingsolver and Diamond (2011) did not detect tradeoffs in selection through different fitness components for most traits in most studies (with the exception of body size), but clearly such tradeoffs are important in some systems (e.g., Price and Grant 1984). Numerous authors have grappled with the problem of combining estimates of multiple bouts of selection through different fitness components (e.g., Arnold and Wade 1984; Wade and Kalisz 1990; Shaw et al. 2008; McGlothlin 2010). Most recently, Shaw et al. (2008) introduced aster modeling, which allows one to combine estimates of selection acting through different fitness components in a unifying statistical framework to gain insight into overall fitness (see below).

While most studies have focused on fitness components rather than true fitness due to the logistical challenges of tracking the reproductive success of large numbers of individuals across their life time, evolutionary biologists might benefit from considering a demographic approach to this same question. Indeed, a more explicit understanding of the feedback between selection, adaptive evolution, and population dynamics may also help with lack of information on true fitness (Gomulkiewicz and Holt 1995; Saccheri and Hanski 2006; Kokko and López-Sepulcre 2007). Elasticities provide a means to quantify the relative contribution of a fitness component on λ (the asymptotic rate of population growth), as they indicate the proportional change in λ from a proportional change in a fitness component. Elasticities are ideal for relating demography and selection because they are identical to selection differentials of vital rates when scaled to have a mean value of zero and variance of one (van Tienderen 2000). Interestingly, although studies of selection consistently show that selection through variation in fecundity is stronger than selection through survival (Endler 1986; Kingsolver and Diamond 2011; Siepielski et al. 2011), studies of elasticities and population demography instead suggest that survival makes a much more important contribution to population growth rates (Crone 2001). Modeling approaches that explicitly link variation in traits, vital rates and elasticities, and λ will be important in resolving these issues (Shaw et al. 2008; Horvitz et al. 2010).

Finally, it is worth noting that the selection results summarized above (see Lessons) are based on variance-standardized selection gradients and differentials, as proposed by Lande and Arnold (1983). Hereford et al. (2004) suggested that traits should instead be standardized by the mean for the trait, as opposed to its standard deviation, which they argue conflates selection and variation. One useful property of mean-standardization is that the gradient for fitness itself is one, which provides a useful and intuitive benchmark for comparing the strength of selection across studies. Using this standardization, Hereford et al. (2004) concluded that the magnitude of directional selection was often quite strong, with a median (bias-corrected) value of $|\beta| = 0.28$, and some values exceeding 1—i.e., stronger than fitness itself. The interpretation of mean-standardized estimates of β greater than 1 is problematic, and by definition must result from methodological or statistical (e.g.,

sampling error or covariance among parameter estimates) rather than biological factors (Hereford et al. 2004). Mean-standardization is only useful for traits that have a ‘natural’ zero and non-negative values, so it cannot be used for phenological dates or times, PC scores, or similar traits (Kingsolver and Diamond 2011). In addition, large values of mean-standardized β are consistently associated with small values of the coefficient of variation (CV), the ratio of the standard deviation to the mean of the trait (Kingsolver and Pfennig 2007). There is no obvious biological reason for very strong selection to be associated with small CV values, suggesting that these large estimates of mean-standardized β are misleading (Kingsolver and Pfennig 2007). Despite these limitations, the merits and behaviors of different standardizations is an important issue; unfortunately, the majority of selection studies do not report mean trait values, even since the publication of Hereford et al.’s (2004) study.

Data availability and statistical power

One of the greatest challenges to our understanding of selection in the wild is insufficient access to data or related information. This limits our understanding and interpretation of past results in several important ways. Publication bias, specifically what has been termed the ‘file-drawer problem’, refers to the tendency for non-significant results to not be published, especially for studies with smaller sample sizes, which can lead to overestimation of effects (Palmer 1999). Although statistical methods have been developed to both detect and correct publication-bias (e.g., the trim-and-fill method of Duval and Tweedie 2000; and the calculation of fail-safe numbers for minimal sample sizes without publication bias effects in meta-analyses, Rosenberg 2005), there is considerable debate regarding their utility (reviewed in Harrison 2011). Visual inspection of funnel plots for selection estimates does suggest some publication bias, especially for quadratic selection (Kingsolver and Diamond 2011; Kingsolver et al. 2001).

A related issue involves information on sampling error. The majority of published studies which report estimates of selection do not provide standard errors of those estimates (Kingsolver and Diamond 2011; Kingsolver et al. 2001; Siepielski et al. 2009). Most selection estimates are based on small sample sizes (i.e., less than 200), resulting in substantial sampling error; and the recent meta-analyses by Morrissey and Hadfield (2012) and presented here (Figs. 1, 2) show that sampling error contributes importantly to variation in selection estimates. The lack of this basic information prevents formal statistical analyses for the majority of past studies of selection—a tremendous waste. Similarly, the lack of data on mean trait values for most studies prevents alternative approaches to standardization (Hereford et al. 2004).

Access to the individual-level data (i.e., data on trait values and fitness measures for each individual) for each study could facilitate major advances in our understanding of selection in natural populations. First, this would enable meta-analyses that account for sampling error, and comparisons of different approaches to standardization. Second, this would allow computation of the phenotypic variance–covariance matrix for each study. Within-study covariances among traits could then be modeled directly in a meta-analytical framework (Hadfield 2010; Hadfield et al. 2007). More fundamentally, this would allow researchers to apply and compare different statistical methods and modeling frameworks to the same datasets in a systematic way. This would be particularly valuable in exploring selection and fitness in a demographic framework (see above), and in modeling non-linear selection and fitness surfaces (see below).

Environmental covariance

Variable environmental conditions pose a dilemma: variation in environmental conditions is likely a key driver of selection in many cases (Carlson and Quinn 2007; Siepielski et al. 2009; Calsbeek and Cox 2010; MacColl 2011), yet environmental variation can make it difficult to quantify selection and predict evolutionary responses to selection.

The key underlying assumption of studies of selection is that there is a causal relationship between fitness and phenotype, generating a non-zero covariance between fitness and phenotype. However, when an environmental factor affects both the trait and fitness, this can generate a covariance between the trait and fitness that only exists because of the underlying environmental factor (see, e.g., Figure 1 in Mauricio and Mojonnier 1997). As a result, measures of selection can be biased by the effects of environment, condition, and nutrition (Price et al. 1988; Schluter et al. 1991; Rausher 1992; Stinchcombe et al. 2002; Kruuk et al. 2003). Such biases would likewise result in unrealistic expectations for directional evolution, even if the trait is heritable, because fitness differences are associated only with the environmental component of the trait (Kruuk et al. 2003). The same problems also arise when phenotypes are not randomly distributed across environmental conditions. Analyses using predicted breeding values, as opposed to phenotypic values, suggest that such biases may be quite common (Kruuk et al. 2003; Stinchcombe et al. 2002; Scheiner et al. 2002). However, Hadfield and colleagues have recently illustrated how use of predicted breeding values for this purpose can introduce other statistical biases and problems (Hadfield et al. 2010). An alternative approach is to estimate selection gradients directly from the estimated variances and covariances (Hadfield 2010; Morrissey et al. 2010), which will be a logistical challenge in many field systems (see “[Future directions: what should we do now?](#)”).

A related issue is the coupling of environment, selection and genetic variance, which can limit or accentuate trait evolution (Merilä et al. 2001; Wilson et al. 2006; Husby et al. 2011). For instance, in the Soay sheep (*Ovis aries*), when environmental conditions are particularly harsh, there is a strong selection on birth weight but little genetic variance in this trait, whereas when conditions are favorable, selection is weak and there is ample genetic variance. Environmental coupling such as this can limit the rate of evolution, maintain genetic variation, and favor phenotypic stasis (Merilä et al. 2001; Wilson et al. 2006). By contrast, Husby et al. (2011) recently showed that the strength of selection and expression of genetic variance in great tits (*Parus major*) are positively coupled with environmental conditions (increased spring temperature), which should speed up adaptive evolution.

Characterizing non-linear selection and fitness surfaces

Most studies quantify non-linear selection via quadratic OLS regression analysis to estimate the quadratic selection gradients (γ). But interpreting quadratic estimates is challenging in several ways. First, in general the estimates of the linear and quadratic coefficients are not statistically independent. Second, γ estimates the mean curvature of the fitness function near the population mean trait value; if the population mean is not near the maximum (or minimum) in the fitness function, then the value of γ will not accurately reflect the shape of the fitness curve (Estes and Arnold 2007; Shaw et al. 2008). This problem is exacerbated if the underlying fitness function is not strictly quadratic (see below). As a result, both the magnitude and the sign of quadratic selection gradient estimate can incorrectly characterize non-linear selection on a trait (e.g., Shaw et al. 2008).

The measurement and interpretation of non-linear selection coefficients is even more challenging when multiple traits are considered (Phillips and Arnold 1989). For example, Blows and Brooks (2003) showed that the strength of nonlinear selection on sets of multivariate traits has been regularly underestimated as a consequence of ignoring nonlinear correlational selection acting on pairs of traits. When correlational selection is present, understanding nonlinear selection on that pair of traits requires identifying the major axes of the quadratic response surface, which can be done through canonical analysis (Phillips and Arnold 1989). This approach argues for considering nonlinear selection as a whole, instead of separating univariate estimates of quadratic selection (stabilizing, disruptive) from bivariate (correlational) ones (Blows 2007). Such canonical analyses suggest the strength of nonlinear selection is likely stronger than that indicated by univariate estimates considered in isolation (Blows and Brooks 2003). However, it is often difficult to give biological interpretation to the multivariate canonical axes, suggesting that there is a tradeoff between reducing dimensionality at the cost of interpretability.

One of the major weaknesses of inferring selection from OLS regression is that it assumes that the fitness function can be described by simple linear or quadratic mathematical functions. True fitness surfaces may be much more complex, with multiple peaks and valleys. Schluter (1988) and Schluter and Nychka (1994) introduced general additive models (GAM) (e.g., cubic splines, projection pursuit regression) as a method for estimating the fitness landscape. These nonparametric approaches have greatly facilitated our ability to visualize fitness landscapes, and often reveal more complexity than the results suggested by OLS regression. Still, graphical depictions of the fitness surfaces are limited to two and three-dimensional visualizations. When more than three traits are studied, correlations between traits included and not included in the visualization may lead to erroneous conclusions about the shape of the fitness surface, although canonical analyses may be useful for reducing dimensions (Phillips and Arnold 1989) (but see above). The flexibility of spline-based GAM approaches is critical for visualizing complex fitness surfaces. However, these methods do not produce parameter values that can be readily compared among and within studies. A promising approach to this problem was recently outlined by Calsbeek (2012). The method uses tensor decomposition in combination with projection pursuit regression to evaluate statistically whether complex fitness surfaces vary over time or space.

Recently, aster modeling has been proposed as an alternative parametric approach to the Lande-Arnold method (e.g., Shaw et al. 2008). In brief, this likelihood-based method models components of fitness (which tend to fit standard probability distributions) to infer the compound distribution of overall fitness (which does not tend to conform to a parametric distribution) through the use of the unconditional canonical parameterization. Using simulated data sets to compare the inferred fitness landscape with the known one, Shaw and Geyer (2010) demonstrate that aster modeling yields more accurate estimates of the fitness landscape than the classic OLS method. This modeling framework holds considerable promise both for integrating different components of fitness and for characterizing fitness surfaces and nonlinear selection. However, it is not clear how parameter estimates from this approach could be used as metrics of effect size in meta-analyses of selection.

Future directions: what should we do now?

As we have summarized here, there is an abundance of field studies and synthetic analyses of selection in natural populations, including several recent meta-analyses. But there are also major gaps in our understanding of selection and its evolutionary consequences. Here

we suggest some promising directions for what we should measure, analyze and report for selection studies in the coming decade.

What should we measure?

Patterns of directional selection on size, morphology and phenology are now well-established. In contrast, selection on other important types of traits, including physiology, behavior and dispersal, are poorly represented. Similarly there are few studies of selection on natural variation in phenotypic plasticity, despite widespread plasticity in ecologically relevant traits.

Our understanding of nonlinear selection remains rudimentary at best. Correlational selection is key to the evolution of functional co-adaptation, but only a handful of studies have documented such selection (Sinervo and Svensson 2002). The apparent lack of evidence for widespread stabilizing selection is also a major challenge: careful, replicated studies of traits and systems likely to be under stabilizing or opposing selection are sorely needed.

Another major area where we are lacking in our understanding of selection is in the underlying causes of selection (Wade and Kalisz 1990). Only a handful of studies have tested for the environmental correlates of selection (reviewed in MacColl 2011), including such factors as precipitation (Carlson and Quinn 2007) and temperature (Husby et al. 2011). Even fewer have complemented studies of selection in the wild with key experimental manipulations of phenotypic distributions and putative causal agents of selection (Wade and Kalisz 1990). Understanding the ecological basis of selection and adaptive evolution remains a key challenge in evolutionary ecology (e.g., Wade and Kalisz 1990; Calsbeek and Cox 2010; McPeck 2010; MacColl 2011).

The vast majority of selection estimates are based on measures of individual fitness components that span a fraction of the generation time of the species studied. Snapshots of selection can yield incorrect estimates of the true form of selection (Siepielski et al. 2009; Bell 2010; Cornwallis and Uller 2010); and individual fitness components are often weakly (even negatively) correlated with total fitness. Studies that consider more integrated measures of lifetime fitness are needed to estimate the true magnitude of selection (Arnold and Wade 1984; Wade and Kalisz 1990; Fairbairn and Preziosi 1996; Hunt et al. 2009; Shaw and Geyer 2010), and to evaluate variation in selection within and between-generations (Hedrick 1973; Ellner and Hairston 1994; Sasaki and Ellner 1997; Bell 2010; Punzalan et al. 2010).

How should we analyze data on selection?

Characterizing non-linear components of selection in nature remains a major challenge that standard OLS methods typically cannot address. Canonical axis analysis (e.g., Phillips and Arnold 1989; Blows and Brooks 2003) and spline-based approaches (Schluter 1988; Schluter and Nychka 1994) are valuable tools for describing and visualizing fitness surfaces, but typically do not assess the statistical support for the key features of such surfaces (e.g., location and curvature of ridges or peaks in the surface). Recent approaches such as aster models (Shaw et al. 2008) and Tucker3 tensor decompositions (Calsbeek 2012) hold promise for characterizing and comparing fitness surfaces.

In addition, integrating selection across fitness components through the lifecycle is essential for quantifying the overall selection on traits, but few field selection studies to date have done this (Crone 2001). Recent methods that characterize selection in the context of demographic models (Shaw et al. 2008; Horvitz et al. 2010) suggest a way forward in achieving a more integrated understanding of selection and fitness variation in nature.

Finally, a major rationale for estimating selection gradients is to predict evolutionary responses to selection, typically using the breeder's equation or its multivariate equivalents (Lande and Arnold 1983). However, recent analyses emphasize the important limitations of applying the breeder's equation to predicting evolution in natural populations, because of the complex associations among phenotypes, genotypes and environments in natural settings (Kruuk et al. 2008; Hadfield et al. 2010; Morrissey et al. 2010). Morrissey et al. (2010) propose an interesting framework for addressing these limitations using the Robertson-Price identity, whose assumptions are less restrictive than the breeder's equation.

What should we report?

Meta-analyses require estimates of the standard errors associated with selection coefficients (Morrissey and Hadfield 2012; Siepielski et al. 2009; Kingsolver and Diamond 2011); alternative selection metrics require mean trait values (Hereford et al. 2004). But despite earlier calls for such information (Hereford et al. 2004; Kingsolver et al. 2001), many recent studies do not provide trait means or standard errors (Siepielski et al. 2009). Future selection studies should be required to report these basic summary statistics as a condition of publication. In addition, estimates of phenotypic variances and fitness variance alone (i.e., opportunity for selection, Crow 1958; Arnold and Wade 1984) would be valuable for evaluating temporal and spatial variation in the underlying fitness surfaces, as distinct from changes only in the covariance between phenotypic and fitness variation.

More broadly, we believe our general understanding of patterns of selection is most limited by lack of access to individual-level data (i.e., data on trait values and fitness measures for each individual) of most studies. Consider this: a typical selection study involves hundreds to thousands of person-hours of research time over multiple years. Based on our current datasets of selection gradients and differentials, a typical study contributes only ~30 datapoints (estimates) to our understanding of the generalities of selection (as distinct from knowledge of selection in specific systems). This seems an extremely inefficient use of research time (and funding) towards a general understanding of selection. Access to the individual-level data for each study would enable more sophisticated meta-analyses and alternative approaches to selection standardization and integration of fitness metrics. More fundamentally, it would allow researchers to apply and compare different statistical methods and modeling frameworks to the same datasets in a systematic way, to explore non-linear selection, fitness surfaces, the demographic context of selection, and other important issues (including issues not currently on our radar). Access to 50–100 individual-level datasets in publicly available repositories (e.g., Dryad: <http://datadryad.org/>) by 2020 could facilitate major advances in quantifying the broader patterns of selection in the wild during the coming decade. Recycle your hard-won, slightly used, still precious data today!

Acknowledgments We thank Andrew Hendry, Ryan Martin, Michael Morrissey and two anonymous reviewers for useful suggestions on earlier versions of the manuscript. We thank Jarrod Hadfield for identifying problems and suggesting solutions to an earlier version of the meta-analyses. The Whiteley Center provided an ideal venue for analysis and the initial writing for JGK. Research supported in part by NSF grant IOS-1120500 to JGK.

References

Arnold SJ, Wade MJ (1984) On the measurement of natural and sexual selection: applications. *Evolution* 38(4):720–734

- Bell G (2010) Fluctuating selection: the perpetual renewal of adaptation in variable environments. *Phil Trans R Soc B* 365:87–97
- Blows MW (2007) A tale of two matrices: multivariate approaches in evolutionary biology. *J Evol Biol* 20:1–8
- Blows MW, Brooks R (2003) Measuring nonlinear selection. *Am Nat* 162:815–820
- Bumpus HC (1899) The elimination of the unfit as illustrated by the introduced sparrow, *Passer domesticus*. *Biol Lectures Marine Biol Lab, Woods Hole, MA*, pp 209–226
- Calsbeek B (2012) Exploring variation in fitness surfaces over time or space. *Evolution*. doi: [10.1111/j.1558-5646.2011.01503.x](https://doi.org/10.1111/j.1558-5646.2011.01503.x)
- Calsbeek R, Cox RM (2010) Experimentally assessing the relative importance of predation and competition as agents of selection. *Nature* 465:613–616
- Carlson SM, Quinn TP (2007) Ten years of varying lake level and selection on size-at-maturity in sockeye salmon. *Ecology* 88(10):2620–2629
- Cornwallis CK, Uller T (2010) Towards an evolutionary ecology of sexual traits. *Trends Ecol Evol* 253:145–152
- Crone EE (2001) Is survivorship a better fitness surrogate than fecundity? *Evolution* 55:2611–2614
- Crow JF (1958) Some possibilities for measuring selection intensities in man. *Hum Biol* 30:1–13
- Darwin CD (1859) *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life*. John Murray, London
- Duval S, Tweedie R (2000) Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56:455–463
- Egan SP, Hood GR, Ott JR (2011) Natural selection on gall size: variable contributions of individual host plants to population-wide patterns. *Evolution* 65:3543–3557
- Ellner SP, Hairston NG Jr (1994) Role of overlapping generations in maintaining genetic variation in a fluctuating environment. *Am Nat* 143:403–417
- Endler JA (1986) *Natural selection in the wild*. Princeton University Press, Princeton
- Estes S, Arnold SJ (2007) Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am Nat* 169:227–244
- Fairbairn DJ, Preziosi RF (1996) Sexual selection and the evolution of sexual size dimorphism in the water strider, *Aquarius remigis*. *Evolution* 50(4):1549–1559
- Gomulkiewicz R, Holt RD (1995) When does evolution by natural selection prevent extinction? *Evolution* 49:201–207
- Grant BR, Grant PR (1989) Natural selection in a population of Darwin's finches. *Am Nat* 133(3):377–393
- Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science* 296(5568):707–711
- Gurevitch J, Hedges LV (1999) Statistical issues in ecological meta-analyses. *Ecology* 80:1142–1149
- Hadfield J (2010) MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Softw* 33:1–22
- Hadfield J, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol* 23:494–508
- Hadfield JD, Nutall A, Osorio D, Owens IPF (2007) Testing the phenotypic gambit: phenotypic, genetic and environmental correlations of colour. *J Evol Biol* 20:549–557
- Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB (2010) The misuse of BLUP in ecology and evolution. *Am Nat* 175:116–125
- Harrison F (2011) Getting started with meta-analysis. *Methods Ecol Evol* 2:1–10
- Hedrick PW (1973) Genetic variation and the generalized frequency-dependent selection model. *Am Nat* 107:800–802
- Hendry AP, Kinnison MT (1999) Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53(6):1637–1653
- Hereford J (2009) A quantitative survey of local adaptation and fitness trade-offs. *Am Nat* 173:579–588
- Hereford J, Hansen TF, Houle D (2004) Comparing strengths of directional selection: How strong is strong? *Evolution* 58(10):2133–2143
- Hersch EI, Phillips PC (2004) Power and potential bias in field studies of natural selection. *Evolution* 58(3):479–485
- Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hoang A, Hill CE, Beerli P, Kingsolver JG (2001) Strength and tempo of directional selection in the wild. *Proc Natl Acad Sci USA* 98(16):9157–9160
- Horvitz CC, Coulson T, Tuljapurkar S, Schemske DW (2010) A new way to integrate selection when both demography and selection gradients vary over time. *Int J Plant Sci* 171(9):945–959
- Hunt J, Breuker CJ, Sadowski JA, Moore AJ (2009) Male-male competition, female mate choice and their interaction: determining total sexual selection. *J Evol Biol* 22:13–26

- Husby A, Visser ME, Kruuk LEB (2011) Speeding up microevolution: the effects of increasing temperature on selection and genetic variance in a wild bird population. *PLoS Biol* 9:e1000585
- Kingsolver JG (2009) The well-temperated biologist. *Am Nat* 174:755–768
- Kingsolver JG, Diamond SE (2011) Phenotypic selection in natural populations: what limits directional selection? *Am Nat* 177:346–357
- Kingsolver JG, Pfennig DW (2004) Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution* 58:1608–1612
- Kingsolver JG, Pfennig DW (2007) Patterns and power of phenotypic selection in nature. *Bioscience* 57:561–572
- Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) The strength of phenotypic selection in natural populations. *Am Nat* 157(3):245–261
- Kokko H, López-Sepulcre A (2007) The ecogenetic link between demography and evolution: can we bridge the gap between theory and data? *Ecol Lett* 10:773–782
- Kruuk LEB, Merilä J, Sheldon BC (2003) When environmental covariance short-circuits natural selection. *Trends Ecol Evol* 18:207–208
- Kruuk LEB, Slate J, Wilson AJ (2008) New answers for old questions: the evolutionary quantitative genetics of wild animal populations. *Annu Rev Ecol Evol Syst* 39:525–548
- Lande R (1979) Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402–416
- Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* 37(6):1210–1226
- MacColl AD (2011) The ecological causes of evolution. *Trends Ecol Evol* 26:514–522
- Mauricio R, Mojonier LE (1997) Reducing bias in the measurement of selection. *Trends Ecol Evol* 12:433–436
- McGlothlin JW (2010) Combining selective episodes to estimate lifetime nonlinear selection. *Evolution* 64:1377–1384
- McPeck M (2010) Evolutionary biology: the next 150 years. In: Bell MA, Futuyma DA, Eanes WF, Levinton JS (eds) *Evolution since Darwin: the first 150 years*. Sinauer Press, Sunderland
- Merilä J, Kruuk LE, Sheldon BC (2001) Explaining stasis: microevolutionary studies of natural populations. *Genetica* 112(113):199–222
- Mitchell-Olds T, Shaw RG (1987) Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41(6):1149–1161
- Morrissey MB, Hadfield JD (2012) Directional selection in temporally replicated studies is remarkably consistent. *Evolution* 66:435–442
- Morrissey MB, Kruuk LEB, Wilson AJ (2010) The danger of applying the breeder's equation in observational studies of natural populations. *J Evol Biol* 23:2277–2288
- Palmer AR (1999) Detecting publication bias in metaanalyses: a case study of fluctuating asymmetry and sexual selection. *Am Nat* 154:220–233
- Phillips PC, Arnold SJ (1989) Visualizing multivariate selection. *Evolution* 43:1209–1222
- Poulin R (2000) Manipulation of host behaviour by parasites: a weakening paradigm? *Proc R Soc B* 267:787–792
- Price TD, Grant PR (1984) Life history traits and natural selection for small body size in a population of Darwin's finches. *Evolution* 38(3):483–494
- Price TD, Kirkpatrick M, Arnold SJ (1988) Directional selection and the evolution of breeding date in birds. *Science* 240:798–799
- Punzalan D, Rodd FH, Rowe L (2010) Temporal variation in patterns of multivariate sexual selection in a wild insect population. *Am Nat* 175:401–414
- Rauscher MD (1992) The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46:616–626
- Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. *Genetica* 112(113):183–198
- Rosenberg MS (2005) The file-drawer problem revisited: a general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution* 59:464–468
- R Development Core Team (2011) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL <<http://www.R-project.org/>>
- Saccheri I, Hanski I (2006) Natural selection and population dynamics. *Trends Ecol Evol* 21(6):341–347
- Sasaki A, Ellner SP (1997) Quantitative genetic variance maintained by fluctuating selection with overlapping generations: variance components and covariances. *Evolution* 51:682–696
- Scheiner SM, Donohue K, Dorn LA, Mazer SJ, Wolfe LM (2002) Reducing environmental bias when measuring natural selection. *Evolution* 56:2156–2167

- Schluter D (1988) Estimating the form of natural selection on a quantitative trait. *Evolution* 42(5):849–861
- Schluter D, Nychka D (1994) Exploring fitness surfaces. *Am Nat* 143(4):597–616
- Schluter D, Price TD, Rowe L (1991) Conflicting selection pressures and life-history trade-offs. *Proc R Soc B* 246(1315):11–17
- Shaw RG, Geyer CJ (2010) Inferring fitness landscapes. *Evolution* 64:2510–2520
- Shaw RG, Geyer CJ, Wagenius S, Hangelbroek HH, Etterson JR (2008) Unifying life-history analyses for inference of fitness and population growth. *Am Nat* 172:E35–E47
- Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecol Lett* 12:1261–1276
- Siepielski AM, DiBattista JD, Evans JA, Carlson SM (2011) Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proc R Soc B* 278:1572–1580
- Simmons LW, Tomkins JL, Kotiaho JS, Hunt J (1999) Fluctuating paradigm. *Proc R Soc B* 266:593–595
- Sinervo B, Svensson E (2002) Correlational selection and the evolution of genomic architecture. *Heredity* 89:329–338
- Stinchcombe JR, Rutter MT, Burdick DS, Tiffin P, Rausher MD, Mauricio R (2002) Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Am Nat* 160:511–523
- Stinchcombe JR, Agrawal AF, Hohenlohe PA, Arnold SJ, Blows MW (2008) Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution* 62(9):2435–2440
- Thompson JN (2005) *The geographic mosaic of coevolution*. University of Chicago Press, Chicago
- Thompson DJ, Hasall C, Low CD, Watts PC (2011) Field estimates of reproductive success in a model insect: behavioral surrogates are a poor predictors of fitness. *Ecol Lett* 14:905–913
- Urban MC (2011) The evolution of species interactions across natural landscapes. *Ecol Lett* 14:723–732
- van Tienderen PH (2000) Elasticities and the link between demographic and evolutionary dynamics. *Ecology* 81(3):666–679
- Wade MJ, Kalisz S (1990) The causes of natural selection. *Evolution* 44:1947–1955
- Weis AE, Abrahamson WG (1986) Evolution of a host plant manipulation by gall makers: ecological and genetic factors in the *Solidago-Eurosta* system. *Am Nat* 127:681–695
- Weldon WFR (1901) A first study of natural selection in *Clausilia laminata* (Montagu). *Biometrika* 1:109–124
- Wilson AJ, Pemberton JM, Pilkington JG, Coltman DW, Misfud DV, Clutton-Brock TH, Kruuk LEB (2006) Environmental coupling of selection and heritability limits evolution. *PLoS Biol* 7:e216