Inference of selection gradients using performance measures as fitness proxies

Oliver D. Franklin[1] and Michael B. Morrissey[2]

[1] Department of Integrative Biology
University of Guelph,
Guelph, Ontario, Canada, N1G 2W1
ofrankli@uoguelph.ca

[2] Dyers Brae House, School of Biology
University of St Andrews
St Andrews, Scotland, KY18 9TH
michael.morrissey@st-andrews.ac.uk

Abstract

(1) Selection coefficients, i.e., selection differentials and gradients, are useful for quantifying selection and for making comparisons across traits and organisms, because they appear in known equations for relating selection and genetic variation to one another and to evolutionary change. However, selection coefficients can only be estimated in organisms where traits and fitness (components) can be measured. This is probably a major contributor to taxonomic biases of selection studies. Aspects of organismal performance, i.e., quantities that are likely to be positively related to fitness components, such as body size, are sometimes used as proxies for fitness, i.e., used in place of fitness components in regression-based selection analysis. To date, little theory exists to inform empirical studies about whether such procedures may yield selection coefficients with known relationships to genetic variation and evolution.

(2) We show that the conditions under which performance measures can be used as proxies for fitness are very limited. Such analyses require that the regression of fitness on the proxy is linear and goes through the origin. We illustrate how fitness proxies may be used in conjunction with information about the performance-fitness relationship, and clarify how this is different from substituting fitness proxies for fitness components in selection analyses.

(3) We apply proxy-based and fitness component-based selection analysis to a system where traits, a performance measure (size; similar to proxies that are commonly used in place of fitness), and a more proximate fitness measure, are all available on the same set of individuals. We find that proxy-based selection gradients are poorly reflective of selection

gradients estimated using fitness components, even when proxy-fitness relationships are quite strong and reasonably linear.

(4) We discuss the implications for proxy-based selection analysis. We emphasise that measures of organismal performance, such as size, may in many cases provide useful information that can contribute to quantitative inferences about natural selection, and their use could allow quantitative inference about selection to be conducted in a wider range of taxa. However, such inferences require quantitative analysis of both trait-performance and performance-fitness relationships, rather than substitution of performance for measures of fitness or fitness components.

Key words: quantitative genetics, phenotypic selection, microevolution, natural populations, path analysis

Introduction

Understanding how natural selection operates in the complexity of the wild remains a key challenge for biologists, and data across a broad range of natural systems is necessary to understand both general patterns and causes of differences in selection. However, data on natural selection in the wild comes predominantly from a limited range of taxa (Kingsolver and Diamond 2011; Siepielski *et al*. 2013). A major barrier to quantitative inference of selection in many taxa is determination of individual fitness or fitness components (survival and fecundity) in the wild. While a formal quantitative genetic framework exists for measuring selection via fitness and fitness components, little formal work has been conducted on how quantitative inference of selection might be made via non-fitness (component) variables that might plausibly be highly related to fitness, given knowledge of the biology of a particular taxon. For example, mass-fecundity relationships are often well known at the species or even population level in fishes (e.g. Wootton 1973; Stauffer 1976; Blueweiss *et al*. 1978; Power *et al*. 2005), and thus relationships of traits with mass might profitably be used to make inferences about selection. This work seeks to formalise the theory of quantitative inference of selection via fitness proxies, and to specify the necessary conditions under which fitness proxies can be used for quantitative inference of natural selection.

Fitness is the expected contribution of individuals to future generations. If a trait covaries with fitness, the distribution of that trait weighted by fitness will be different than the unweighted distribution, i.e., the distribution among unselected individuals. If this association has a (partial) genetic basis, that change in the distribution of phenotype will be (in part) propagated into future generations (Robertson 1966; Lynch and Walsh forthcoming). *Fitness components*, which are quantities appearing in a life table (i.e. age-specific survivorship and fecundity) or summaries of life table entries (e.g. survival to maturity), represent the demographic contribution of individuals to the population during periods within the life cycle. Associations of traits with fitness components are associated with changes in the distributions of traits, weighted by those fitness components, during the period to which the fitness components pertain. As for associations with fitness, any genetic basis to these trait-fitness component relationships also generates evolutionary change, providing antagonistic selection does not occur via other fitness components. Importantly, many aspects of the statistical mechanics quantitatively relating trait-fitness (component) relationships and the genetic basis of variation in traits to evolutionary change are known (Robertson 1966; Price 1970; Lande and Arnold 1983; Arnold and Wade 1984; Mitchell-Olds and Shaw 1987; van Tienderen 2000; Shaw *et al*. 2008; Morrissey 2014a; 2015). This body of theoretical work justifies the concept of selection coefficients (Lush 1937; Lande and Arnold 1983; Arnold and Wade 1984) which have proven useful for synthesising general information about selection in the wild (e.g. Endler 1986; Kingsolver *et al*. 2001; Hereford *et al*. 2004; Kingsolver and Pfennig 2004; Siepielski *et al*. 2009; Kingsolver *et al*. 2012; Morrissey and Hadfield 2012). In contrast to fitness (components), *fitness proxies,* such as measures of organismal performance, cannot be assumed to be indicative of demographic contribution to future generations because aspects of performance that are not demographic rates do not reflect the representation of an individual's genes in a population at future times. Here we refer to fitness proxies and performance, the former is a broad term describing any measure used in place of fitness, while the latter refers to a measure of organismal success that is justified by the natural history of a given organism (aspects of size are commonly used as performance-based fitness proxies).

Recognising the practical difficulties associated with measuring both traits and fitness on the same organisms in the wild, Arnold (1983) introduced the concept of performance gradients, calculated in much the same way as selection gradients. It is important to note that a performance gradient, calculated via (multiple) regression of performance on phenotype, is not a selection gradient. Arnold explained that the product of coefficients of trait-performance and performance-(relative) fitness regressions is equivalent to a selection gradient. Importantly, he suggested that the trait-performance and performance-fitness data could come from different studies (e.g., one in the lab and the other in the field). This is particularly useful where it is unfeasible to measure traits and fitness of the same individuals in the wild: measurements of performance (e.g. mass) and focal traits in the wild can be combined with data from studies on other individuals (e.g. a known mass-fecundity relationship from lab studies) to enable calculation of selection gradients. The introduction of path analytical techniques to selection analyses by Arnold (1983) has been important in motivating interest in, and providing a framework for, exploring trait interactions and functional pathways within studies of natural selection (see Kingsolver and Huey 2003). However, despite Arnold's (1983) recommendation that trait-performance and performance-fitness data can be combined to give selection gradients as justified in evolutionary quantitative genetic theory, this approach has rarely been used to investigate taxa for which trait-fitness measurements in the wild are unattainable.

For taxa in which field measurements of fitness are unobtainable, there is a literature that uses performance measures (see Table 1). However, rather than the incorporation of quantitative information about performance-fitness relationships from other studies or individuals than the trait-performance data are obtained, these studies have predominantly used relative performance measures, such as size or growth, as substitutes for relative fitness. Such fitness proxies are usually measures of organismal performance known to be related to the focal trait(s) and believed to causally affect fitness. The inherent, and typically stated, assumption of a positive performance-fitness relationship tends to be convincingly justified in these reports, typically by appealing to knowledge of the organismal biology and ecology of the particular study system. Although the logical connection between traits, performance, and fitness does not differ between the approach in these studies and that proposed by Arnold (1983), these studies have calculated what Arnold termed ‘performance gradients’ but interpreted them as ‘selection gradients’ with their associated evolutionary quantitative genetic justification and the comparisons they make possible. Use of performance in place of fitness extends beyond studies that have self-identified as using proxies (i.e., Table 1), such that a discernible proportion exist in the Kingsolver *et al.* (2001) selection meta-analysis: e.g. territory tenure (Grant 1985; Grether 1996), aspects of mass, nest defence (van den Berghe and Gross 1989; Mitchell-Olds & Bergelson 1990a). The performance gradient is only part of Arnold’s approach, which also requires use of a known performance-fitness relationship. So, although these studies are typically based on sound biological knowledge, it is unknown whether this approach can yield selection gradients that are interpretable in a quantitative genetic framework. To whatever extent different quantitatively-justifiable options exist for using performance in selection studies, the range of taxa in which we can infer quantitative estimates of selection in the wild could be greatly increased.

In this paper, we first analyse a model where traits have direct effects on a performance measure, and that performance measure has a direct effect on fitness. We first analyse a simple case, where all effects are linear. We derive a simple but limited condition under which performance can be substituted for fitness in selection analyses, beyond those typically assumed and stated when performance measures are used in place of fitness components in selection analysis.

Importantly, we confirm that path analysis can be used to construct analyses that use performance data, in conjunction with additional quantitative data about performance-fitness relationships, to make inferences about natural selection (as suggested by Arnold 1983). In the appendix, we extend our analysis to non-linear selection, and show that similar conditions hold for estimation of quadratic and correlational selection. We then apply performance-based (i.e., using performance as a proxy in calculations of relative fitness), fitness-based (i.e., using a fitness component for calculation of relative fitness), and performance-mediated (i.e., using performance in conjunction with data on the performance-fitness relationship) selection analyses in a system in which trait, performance, and fitness data are all available on the same individuals (see Figure 1 for a graphical representation of these terms). We conduct linear and quadratic/correlational analyses using linear models of the performance-fitness relationship (as justified by our theoretical sections), and we describe and perform a more flexible numerical analysis (extending methods in Morrissey and Sakrejda 2013 to performance-mediated analysis), allowing non-linear performance-fitness relationships to be accommodated. We conclude with a discussion of the best means to leverage performance data for quantitative selection analysis, and of ways in which trait-performance relationships can be of use for qualitative inference of natural selection in the absence of quantitative performance-fitness information.

Theory

Here we outline the conditions under which performance measures can be substituted for fitness measures to obtain selection gradients. We derive conditions for their equivalence where the relationship between performance and fitness is assumed to be linear. This provides a useful case for exposing the basic principles, while aspects of analysis of non-linear trait-performance and performance-fitness relationships are detailed in the appendix. We assume that the effect of the trait on fitness is mediated entirely by the performance measure.

Linear performance-fitness relationship

Assume a linear trait-performance function

$$(1) \qquad E[y_i|z_i] = f(z_i) = a + bz_i,$$

where $E[y_i|z_i]$ is expected performance, *y*, given phenotype, *z*, for individual *i*, as a function of an intercept, *a*, and a slope term, *b*, defining the regression of *y* on *z*. Similarly, assume a linear performance-fitness function

$$(2) \qquad E[W_i|y_i] = f(y_i) = m + ny_i,$$

where W_i is individual absolute fitness, and *m* and *n* are the intercept and regression coefficients of the linear regression of fitness on performance.

The selection gradient is the derivative of relative fitness, i.e., $w_i = \frac{W_i}{\bar{W}}$, with respect to phenotype, *z*, averaged over the distribution of phenotype. The first step to obtaining an expression for the selection gradient requires that we relate individual phenotype directly to fitness. We can perform this first step by substituting (1) into (2). Because (2) is a linear function, $\mathrm{E}[f(y_i)] = f(\mathrm{E}[y_i])$ irrespective of the distributions of residuals of performance and fitness; note that this relation would not hold if the performance-fitness function were not linear (see further discussion in the appendix). Therefore, we can write expected fitness given phenotype as

$$(3) \qquad E[W_i|z_i] = W(z_i) = m + n(a + bz_i) = m + na + nbz_i.$$

We can write the derivative of expected fitness with respect to phenotype as

(4) $$\frac{dW}{dz} = nb = \mathrm{E}[\frac{dW}{dz}],$$

which is a constant (both *n* and *b* are constants in equations 1 and 2, to be estimated in practice). In the model developed so far, with a linear trait-performance function and a linear performance-fitness function, the derivative of fitness with respect to phenotype does not depend on phenotype. Therefore (4) also gives the average derivative of (absolute) fitness with respect to phenotype, as depicted in its last relation.

Since the selection gradient depends on relative fitness rather than absolute fitness, we must express the average trait-fitness relationship in terms of relative fitness in order to obtain the selection gradient, i.e.,

(5) $$\beta = \frac{E[\frac{dW}{dz}]}{\bar{W}}$$

Again, because all relationships are linear, the mean fitness is the expected fitness given the mean phenotype. Therefore the selection gradient can be expressed as,

(6) $$\beta = \frac{nb}{m+na+nb\bar{z}},$$

where $\bar{z}$ is mean phenotype. In order to render the implications of this expression more intuitive, and applicable to performance-based selection analysis (where performance will be divided by its mean to derive a proxy for relative fitness), we consider the case where mean performance is 1. Furthermore, since the mean phenotype may be arbitrarily scaled, we consider the case where it is centred to a mean of zero (this is typical, but not necessary, in an analysis of linear selection, and necessary in regression-based analysis of linear and quadratic selection; Lande and Arnold

1983). Under these conditions, the coefficient *a* in equation 1 has a value of one; therefore, treating performance as relative fitness and centring the phenotype, we simplify equation 6 to yield

$$(7) \qquad \beta = \frac{nb}{m+n(1)+nb(0)} = \frac{nb}{m+n},$$

from which we can see that the linear coefficient in a regression of relative performance on phenotype is only interpretable, even if the performance-fitness relationship is strictly linear (i.e., the most commonly-stated assumption in existing empirical performance-based selection analyses, and also an assumption encoded in equation 2), if the regression of fitness on performance passes through the origin i.e., if $m = 0$. We show in the appendix that this pair of conditions for performance-based selection gradients to be correct, i.e., a truly linear regression through the origin of fitness on performance, holds for non-linear selection gradients, and for inference of selection differentials, and for different standardisations of selection gradients and differentials.

It is important to note that these conditions for equivalence rest on the assumption that fitness is independent of the traits, conditional on performance. If the traits affect fitness directly and/or via an additional unmeasured aspect of performance, then any resulting selection gradient must be interpreted as a partial description of the selection gradient, that is, selection mediated by that particular performance trait. A number of authors have invoked positive relationships between performance and fitness as justifying the use of performance as a substitute for fitness or fitness components in selection analysis (Table 1). Given the above analysis, along with the potential for alternative performance pathways, we caution that the conditions for such an interpretation are much stricter. Even where traits are independent of fitness, conditional on measured performance, and the performance-fitness regression is truly linear, the error can be of essentially arbitrary order of magnitude. The error is dependent on the relationship between the intercept and slope, i.e., the proportion by which a performance-based analysis overestimates the selection gradient is given by rearrangement of equation (7) as $\frac{b}{\beta} = \frac{m+n}{n}$ (Figure 2). None of the studies using performance as a fitness proxy (Table 1) considered the intercept of the performance-fitness regression as part of their justification. A statistic that is consistently reported in support of the use of performance measures as substitutes for fitness is the correlation or r^2 of the performance-fitness relationship. The correlation of performance with fitness has no bearing on the adequacy of performance-based selection analysis, when the performance-fitness relationship is linear.

Comparison and interpretation of fitness- and performance-based selection inferences in *Arabidopsis thaliana*

In a single dataset, we conduct fitness-based, performance-based, and performance-mediated selection gradient analysis. This exercise allows us to investigate performance-based and performance-mediated selection gradient analysis, and to investigate the consequences when and if their assumptions are not met. In particular, the performance-based selection gradient analysis assumes that the performance-fitness relationship is linear and intersects the origin, and both the performance-based and performance-mediated analyses assume that there are no alternative pathways through which the focal traits affect fitness. We first assess these assumptions for our focal dataset, before comparing the gradients generated through performance-based and fitness-based analyses, for both linear and nonlinear trait effects.

The focal dataset, provided by H.S. Callahan (Columbia University) and S.M Scheiner (NSF), is an experimental population of *Arabidopsis thaliana*. Path analysis-based inference of selection in this system has been reported in Scheiner *et al*., (2000). Briefly, in this study the plants were monitored every other day to record bolting day (inflorescence initiation) and number of rosette leaves and, after flowering had ended, inflorescence height and fruit number were measured. Further details about the study system are available in Scheiner *et al*. 2000.

For performance-based and performance-mediated analyses, we adopted Scheiner *et al*.'s (2000) *a priori* path model, whereby bolting day and leaf number both directly affect inflorescence height, and inflorescence height directly affects fruit number, the measure of fitness (Figure 1c). We consider inference of direct selection gradients of bolting day and leaf number, and we use inflorescence height as a performance measure. Scheiner *et al.*'s original intention was to consider height a trait; however their path diagram is particularly amenable to demonstrating how an intermediate such as this may be used to generate selection inferences, and what assumptions are required to use such a trait as a substitute for fitness. In fact, as a measure of organismal size, this character is immediately analogous to many measures used in performance-based selection inference (e.g. Mitchell-Olds and Bergelson 1990a, Heschel *et al.* 2005, Martin and Pfennig 2009, Ossler *et al*. 2015, Ramirez-Valiente *et al*. 2015).

The performance-fitness relationship: assessing the potential for substitution of performance for fitness

First we characterise the performance-fitness relationship by regressing fruit number on inflorescence height. Visual inspection suggests that it is reasonable to assume linearity for the height-fruit number relationship, and that the regression line is close to intersecting the origin

(Table 2, Figure 3a). Despite the apparent modest deviation of the intercept from the origin (Table 2), the quantitative consequences of the intercept not intersecting the origin are manifested in terms of the relationship of the intercept, not to the overall range of the performance and fitness data, but to the slope of the regression of fitness on relative performance. The intercept and slope are -2.96 and 9.23, respectively (-0.47 and 1.47 for the regression of relative fitness on relative performance; Table 2), and so correspond to a proportional error of $\frac{m+n}{n} = \frac{-2.96+9.23}{9.23} = 0.68$. In other words, the performance gradient is only 68% the value of the fitness-based selection gradient. Furthermore the height-fruit number relationship is significantly non-linear (Table 2, Figure 3b). It is not clear how to make an *a priori* judgement of the severity of any potential effects of this non-linearity on performance-based selection inferences. See the appendix for a brief discussion on why theoretical treatment of systems with quadratic (or other non-linear performance-fitness) relationships would be complex and unlikely to yield sufficiently simple results to be generally informative. We will nonetheless return to the issue of non-linearity of the performance-fitness relationship.

The other assumption of performance-based and performance-mediated selection analysis - that fruit number is independent of traits, conditional on inflorescence height - determines whether the error associated with the performance-fitness relationship pertains to total selection on the focal traits or only the portion of selection that is mediated by inflorescence height. To test this assumption we regressed focal traits and performance on fitness according to

(8a) $$w = \mu_w + b_1 y + b_2 z_{blt} + b_3 z_{lnr} + e, \mathit{and}$$

(8b) $$w = \mu_w + b_1 y + b_2 z_{blt} + b_3 z_{lnr} + \frac{1}{2} g_1 y^2 + \frac{1}{2} g_2 z_{blt}^2 + \frac{1}{2} g_3 z_{lnr}^2$$
$$+ g_{12} z_y z_{blt} + + g_{13} z_y z_{lnr} + g_{23} z_{blt} z_{lnr} + g_{123} z_y z_{blt} z_{lnr} + e,$$

where w represents relative fitness, y represents inflorescence height, μ represents the intercept, and z represents phenotype (subscripts blt and lnr denote bolting day, and leaf number, respectively). Note that the analogous regression coefficients, e.g., b_1 for the effect of y on w in equations 8a&b are separately estimated. Similarly, the re-use of such coefficients in subsequent models throughout the example analyses (i.e., in equations 9, 10 and 12) is to avoid confusion arising from separately defining many different quantities; throughout, all coefficients are separately estimated by the regression analyses described by each equation. All predictors were variance-standardised and centred on zero ($(z-\bar{z})/\sigma_z$), and fitness was relativized ($w/\bar{w}$), as in typical selection analyses (Lande and Arnold 1983). Statistically significant direct effects of bolting day and leaf number on fitness in the analyses that include inflorescence height as a predictor (equations 8a&b), indicate that fruit number is not independent of these traits, conditional on inflorescence height (Table 3). The error associated with using performance measures in place of fitness in this example will therefore stem from both the nature of the performance-fitness relationship and the existence of alternative paths through which selection of focal traits are mediated.

We have now ascertained that for the *Arabidopsis* data, performance-based and fitness-based estimates of selection gradients will not be equivalent, and that differences will be due, in part, to the existence of alternative causal pathways of traits effects on fitness and, in part, to a performance-fitness relationship that may be reasonably approximated with a linear function, but fails to intersect the origin. Next we calculate the scale of this total error, and estimate the contribution of the performance-fitness relationship, by comparing selection gradients calculated in three different ways. First, selection gradients (considering bolting time and leaf number as the focal trait vector) are calculated with a multiple regression of relative fitness on traits (following

Lande and Arnold 1983; see also Stinchcombe *et al.* 2008), and we consider these ‘fitness-based selection gradients’ to be the most theoretically justifiable selection gradients and thus the reference against which we will judge estimates based on the performance measure (height, in our example). Next, we calculate gradients with multiple regression of relative performance on traits, as is the use of fitness proxies in the literature (e.g. Table 1), and refer to these as ‘performance-based selection gradients’. The difference between these two forms of gradients provides an estimate of total error associated with using performance as fitness in this system. Finally, we calculate gradients using a path analytical approach in order to estimate selection of traits mediated only by our performance measure, and we term these ‘performance-mediated selection gradients’. The difference between performance-mediated selection gradients and performance-based selection gradients is introduced by the performance-fitness relationship.

Fitness-based selection gradient analysis

The equations for the fitness-based selection gradients are

(9a) $$w = \mu_w + \beta_{blt} z_{blt} + \beta_{lnr} z_{lnr} + e,$$

for analysis of directional selection gradients only, and

(9b) $$w = \mu_w + \beta_{blt} z_{blt} + \beta_{lnr} z_{lnr} + \frac{1}{2}\gamma_{blt} z_{blt}^2 + \frac{1}{2}\gamma_{lnr} z_{lnr}^2 + \gamma_{blt,lnr} z_{blt} z_{lnr} + e,$$

for a full analysis of directional and quadratic selection. β and γ represent directional and quadratic selection gradients, subscripted by the traits to which they pertain. Prior to inclusion in the model, focal traits were standardised to mean zero and unit variance, and fitness was relativized. Fitness-based selection gradients from the combined linear and quadratic model indicate directional selection for earlier bolting date (β_{blt} = -0.244, SE: 0.058, P<0.005) and greater leaf numbers (β_{lnr} = 0.451, SE: 0.041, P<0.005), with positive quadratic selection in both instances (γ_{blt} = 0.163, SE: 0.069, P=0.02; γ_{lnr} = 0.162, SE: 0.052, P<0.005, respectively), but no significant correlational selection (Figure 4).

Performance-based selection gradient analysis

To explore the effects of using a performance measure as a proxy for fitness, we substituted the relative inflorescence height for the response variable in equations 9a&b yielding

(10a) $y = \mu_y + b_{blt}z_{blt} + b_{lnr}z_{lnr} + e$, and

(10b) $y = \mu_y + b_{blt}z_{blt} + b_{lnr}z_{lnr} + \frac{1}{2}g_{blt}z_{blt}^2 + \frac{1}{2}g_{lnr}z_{lnr}^2 + g_{blt,lnr}z_{blt}z_{lnr} + e$

where *y* represents relative performance (inflorescence height). Regression coefficients are as for equations 8ab and 9ab, except directional effects are denoted by *b* and quadratic effects by *g*, for consistency with the theory section, and to distinguish these quantities from selection gradients as justified by quantitative genetic theory. Qualitatively, the relationships among focal traits and height are the same as those among focal traits and fruit number, although the positive quadratic value of leaf number is no longer significantly different from zero (Figure 4). However, when considering the magnitude of the coefficients, there is a considerable difference between fitness-based and performance-based selection gradients (from the quadratic model, bolting time: linear -0.114, SE: 0.035, P<0.005, quadratic 0.101, SE: 0.042, P=0.02; leaf number: linear 0.279, SE: 0.025, P<0.005; see Figure 4).

Performance-mediated selection gradient analysis

Our comparisons of performance-based and fitness-based selection gradients illustrate that these quantities can differ; however, the differences in our example analyses will be because of broken assumptions about the performance-fitness relationship (Figure 3, Table 2), and about conditional independence of traits and fitness, given performance (Table 3). To disentangle these two sources of error we must estimate only the portion of selection mediated by the performance trait of inflorescence height. We thus conduct formal performance-mediated selection analyses, such that we can relax assumptions about the performance-fitness relationship. We first relax the assumption that the linear regression of fitness on performance passes through the origin, using results from our analytical theory (see above, and the appendix for non-linear analysis based on quadratic trait-performance regressions).

Linear performance-fitness model

To estimate only the selection mediated by performance we multiply the performance-based selection gradients by the coefficient of the regression of relative fruit number on relative inflorescence height. Thus we obtain the performance-mediated selection gradients as the product of the performance gradients and the regression of relative fitness on relative performance. This is inherent to equation 3, and justified by equation 7 (when traits are mean-centred and performance is relative performance), and equations A4 and A8 (in appendix). These are the path analyses advocated by Arnold (1983). Note that we conduct this analysis based on relative performance. The results of the path analysis-based performance-mediated selection analysis are identical regardless of whether absolute or relative (or some other linear re-scaling of) performance is used, so long as the same scaling of performance is used in the analysis of the trait-performance and the performance-fitness relationships. We used standard errors of the performance gradients and the linear performance-relative fitness regression to calculate standard errors of the performance-mediated selection gradients, according to a first order approximation (Lynch and Walsh 1998, Appendix 1)

(11) $$SE[XY] = \sqrt{SE[X]^2 + SE[Y]^2 + \hat{Y}^2 SE[X]^2 + \hat{X}^2 SE[Y]^2}$$

where X and Y represent two quantities, in this case performance gradients and the performance-relative fitness slope, $\hat{X}$ and $\hat{Y}$ are estimated values of X and Y, and $SE[X]$ and $SE[Y]$ represent their corresponding standard errors.

The comparison of performance-mediated with performance-based selection gradients isolates the error associated with the performance-fitness relationship not meeting the conditions of linearity and origin intersection, in this example only the latter. The effect is to underestimate the magnitude of selection by approximately 1/3 (i.e., we expect the true value to be about 50% greater than the performance-based gradients), and this applies to both linear and quadratic components (see Figure 4, see also equations 7 and A8).

We have thus shown that use of a performance measure in place of a fitness measure has the potential to drastically alter the estimate of selection, even under the assumption of a linear performance-fitness relationship. The strength of selection here was underestimated because of the presence of alternative pathways among traits and fitness and because of the performance-fitness relationship not meeting the condition of linearity and intersection of origin. Whereas the former source of error can influence traits differentially, performance-based selection gradients for traits included in the analysis will be affected equally by the proportional error introduced by the performance-fitness relationship, which validates within-study comparisons of selection, if fitness can be assumed to be independent of traits, conditional on performance.

Non-linear performance-fitness model

In this section, we outline a numerical approach to multivariate directional and quadratic performance-mediated selection analysis. The analysis centers on two functions. The first $y(\boldsymbol{z})$ takes (potentially multiple) trait values as predictor variables, and expected values of a performance measure as a response variable. The second $W(y)$ takes the performance measure, and returns expected fitness.

In order to model the trait-performance relationship for the non-linear performance-mediated selection gradients analysis, we fitted a quadratic regression of log inflorescence height (performance) on the traits

$$(12) \qquad y = a + b_1 z_1 + b_2 z_2 + \frac{1}{2} g_1 z_1^2 + \frac{1}{2} g_2 z_2^2 + g_{12} z_1 z_2 + e,\ e \sim N(0, \sigma_e^2)$$

where y is the performance measure (log inflorescence height, in our non-linear analysis), a is an intercept, b_1 and b_2, g_1 and g_2, and g_{12} are linear, quadratic, and correlational regression coefficients for the two traits, bolting day, and leaf number. e are residuals, the variance of which (σ_e^2) are estimated. For comparison, we also used a model for (log) performance given phenotype with only linear effects.

The resulting coefficients are qualitatively similar as when relative performance was the response variable (equation 10b, see Table 4 and Figure 5a). We examined model r^2 values and distributions of residuals for versions of equation 12 applied to both logged and un-logged performance data, and these aspects of model fit were very similar in both cases (not shown). Since logged values of a strictly positive character seem most natural in an additive model, we adopted these (see Table 2). We also adopted a quadratic regression model for $W(y)$ in order to model the non-linear performance-fitness relationship, but in a generalized linear model analysis, with a log link function and assuming Poisson errors, (13a) $$W_i \sim Poisson(E[W_i]),$$

(13b) $$E[W(y)] = \exp(m + ny + oy^2)$$

The coefficients of the model in equations 13a&b are given in Table 2. From these two models, we can construct a function giving expected fitness as a function of individual phenotype

(14) $$W(z)_i = \int W(y)\, N\big(y, y(z_i), \sigma_y^2\big) dy.$$

Because $W(y)$ is a non-linear function, the expected fitness of an individual with a given expected value of performance (based on its phenotype) is not equal to the expected fitness of an individual with that specific value of performance (this is Jensen's 1906 inequality). Consequently, the integration in equation 14 over the distribution of values that performance might take for an individual with a given phenotype is necessary to obtain expected fitness, given phenotype; see Morrissey (2015) for further explanation of this general approach to the inference of quantitative genetic parameters in non-linear systems.

Population mean fitness may be obtained by taking an average of individual expected fitness given by $W(z)_i$, over the distribution of phenotype in a population. We may choose to assume some distribution of phenotype, such as a multivariate normal distribution of the traits, with a mean vector and covariance matrix equal to that estimated directly from the trait data. As such, mean fitness would be given by

(15) $$\bar{W} = \int W(z) p(z) dz,$$

where $p(z)$ is the assumed distribution of phenotype with parameters estimated from the data. Such performance-mediated selection analysis (i.e., using this construction for $\bar{W}$ with the rest of the procedure, below), would exactly follow the mechanics for inference of extended selection gradients in non-linear systems in Morrissey (2015).

However, we may wish to accommodate an analysis that makes fewer assumptions about the distribution of phenotype; in the present example, bolting time is very non-normal. Consequently, we could calculate population mean fitness as

$$\bar{W} = \frac{1}{n}\sum_{i=1}^{n} W(z_i) \tag{16}$$

i.e., an average of expected fitness of all observed phenotypes, where i indexes the n observed individual phenotypes. This approach follows Morrissey and Sakrejda (2014) and Morrissey (2014b). Selection gradients calculated using expected fitness calculated in this way still correspond to changes in the mean and variance of breeding values due to selection (according to expressions given in Lande 1979 and Lande and Arnold 1983), assuming that breeding values are multivariate normal, but not making any parametric assumption about the distribution of environmental effects on phenotype.

Regardless of the choice of function for $\bar{W}$, selection gradients may be calculated using $\beta = \bar{W}^{-1}\nabla\bar{W}$, and $\gamma = \bar{W}^{-1}\nabla^2\bar{W}$, where $\nabla\bar{W}$ and $\nabla^2\bar{W}$ represent the gradient vector and matrix of second partial derivatives of mean fitness with respect to mean phenotype, respectively. Note that these expressions define selection gradients as the derivatives of population mean fitness with respect to population mean phenotype, and are equivalent to the earlier definitions based on the average derivatives of (individual) fitness with respect to (individual) phenotype given earlier. These latter definitions are more directly useful for numerical analysis of selection gradients.

$\nabla\bar{W}$ and $\nabla^2\bar{W}$ are relatively easily calculated numerically. Define a vector of perturbations of mean phenotype x, and a function for mean fitness accommodating these perturbations

$$(17) \qquad \bar{W}(x) = \frac{1}{n}\sum_{i=1}^{n} W(z_i + x)$$

values in $\nabla\bar{W}$ are then be calculated numerically, most simply by finite differences. For example the partial derivative of mean fitness with respect to the mean phenotype for the first trait would be approximated by

$$(18) \qquad \frac{d\bar{W}}{d\bar{z}_1} \approx \frac{\bar{W}(\mathrm{x}_1) - \bar{W}(\mathrm{x}_0)}{h}, \mathrm{x}_1 = [h, 0], \mathrm{x}_0 = [0,0]$$

when h is set to a small value, relative to the SD of the distribution of the trait. Calculations of second partial derivatives are simple extensions of this method, and their implementation is detailed in the supplemental R code.

The performance-mediated selection gradients returned from this analysis closely matched those from the path analysis (justified in the theory section, above, and in the appendix for non-linear selection). This analysis' linear coefficients for bolting day, β_{blt} = -0.181 (SE: 0.064, P<0.005), and leaf number, β_{lnr} = 0.446 (SE: 0.041, P<0.005), are close to those of the path analysis: β_{blt} = -0.168 (SE: 0.052) and β_{lnr} = 0.411 (SE: 0.039), respectively. Nonlinear estimates from the analysis accommodating the non-linear performance fitness function differ more (Figure 4). Thus, the simpler analysis (i.e., the path analysis assuming a linear performance-fitness relationship) may generally be quite robust, at least for inference of directional selection.

Discussion

We have demonstrated that it is possible to substitute performance measures for fitness (components) in regression-based analyses of selection to obtain accurate selection gradients, however the conditions are strict. In addition to linearity of the performance-fitness relationship, the linear regression of fitness on performance must pass through the origin, if performance-based selection analyses are to recover selection gradients. Our literature review suggests that this condition is not generally met (see Table 1). Although these studies – and an additional number not self-identifying as using proxies – could misinform subsequent research, our main concern was the ambiguity surrounding the use of fitness proxies, which may have contributed to a disinclination to estimate selection in systems where fitness is not directly measurable in the wild. We hope that, by clarifying the conditions under which performance measures can be used in selection analyses, researchers will be able to reduce taxonomic disparities in our understanding of selection in the wild.

The relationship between inflorescence height and fruit number in the *Arabidopsis* example does not meet the conditions of linearity and intersection of the origin. However, visual inspection suggests the relationship is reasonably linear, and that its intercept is reasonably close to the origin (Figure 3). It appears that the assumption of linearity is in fact sufficiently well-met to allow path analysis-based performance-mediated selection inference: the selection gradient estimates based on path analysis (assuming a linear performance-fitness relationship) and the full non-linear analysis agree quite closely (Figure 4), certainly closely enough that they lead to equivalent biological interpretations. The apparent minor lack of correspondence between the performance-fitness regression and the assumption that it passes through the origin (Figure 3) could be misleading. In fact, the error associated with the regression not passing through the origin is not determined by how close it is, relative to the distributions of performance and fitness, but rather, it depends on the value of the intercept relative to the slope (Figure 2). In the *Arabidopsis* example, this corresponds to approximately a 50% error, which is reflected in differences between the performance-based and performance-mediated selection inferences (Figure 4).

We attempted to use data from performance-based reports of selection gradients, and their associated sources justifying assumptions about the positive relationships of the performance measures with fitness (Table 1), to reconstruct performance-mediated selection gradient estimates. This exercise would have allowed us to test how large errors are in practice as a result of performance-fitness functions not passing through the origin. However, this exercise required that mean absolute performance (in the trait-performance analysis) was known. In some cases we could not find this information. Furthermore, such reconstructions required that the performance measures in the studies reporting performance-based selection gradients were the same quantities as the performance measures involved in the performance-fitness relationships that were invoked to justify the performance-based analysis; this was often not the case (Table 1). Consequently, we were unable to derive the factor by which performance-mediated selection analysis would differ from performance-based selection analysis (i.e., $\frac{m+n}{n}$, when performance is relativized) in most cases. However, we suspect that a general argument can be made that fitness-performance relationships may systematically fail to intersect the origin. If some threshold level of performance is necessary before any fitness is realised (e.g., if it is often the case that only individuals above some minimum size produce any gametes), then intercepts of performance-fitness relationships would generally be negative, at least when fitness components associated with reproduction are considered. In such cases, the intensity of selection would be systematically underestimated. There is one study, Heschel *et al.* (2005), for which we can reconstruct the selection gradient, given the performance gradient and the information available about the performance-fitness relationship. Heschel *et al.* (2005) reported performance gradients based on biomass as a proxy, and the associated performance fitness function is W = 29.72y - 5.90 (Waller 1979). The slope, if the regression were on relative biomass rather than absolute biomass is obtained by multiplying this slope by mean biomass (3.49g; Heschel *et al.* 2005). Proportional error is calculated as $\frac{b}{\beta} = \frac{m+n}{n} = \frac{-5.90+29.72\cdot 3.49}{29.72\cdot 3.49} = 0.94$, demonstrating that selection gradients via the proxy about are 6% smaller than the corresponding gradients. The relatively small magnitude of this error does suggest that performance measures can be reasonable estimates of fitness, and importantly that the direction and magnitude of errors can be checked and corrected, given data that may already be available about many performance-fitness relationships. With only a single reconstructed comparison (and the larger error in the performance-based analysis in our example) we are unable to ascertain the general scale of any potential bias introduced to our general understanding of selection by the use of proxies.

The fitness measure in our example analysis, fruit number, is not immediately proximate to fitness. In other words, it is not completely representative of the demographic representation of individuals in future generations. A more proximate fitness measure, such as number of seeds, could provide further improvements to the inference of selection in this system, if it were available. We suggest that fruit number can be considered a fitness component (rather than a proxy), since it can be mathematically represented as a multiplicative component of net reproductive rate (i.e., total seed production is the product of number of fruits and number of seeds per fruit). However, since fruit production is not total fitness, the selection gradients reported here must be interpreted as those via fruit production (and, for example, evolutionary predictions based on these selection gradient inferences using the Lande equation should be interpreted as that evolution expected through fitness via the corresponding episodes of selection).

In practice, selection via a fitness component can be related to total selection, or selection via a component more proximate to fitness in two ways: (a) via treating different components as multiplicative and using existing theory to combine selection gradients across multiplicative episodes (Arnold and Wade 1984; Wade and Kalisz 1989), or (b) by applying the same relationships as apply for performance-mediated selection gradient analysis route which, as we have discussed, allows use of separate studies and therefore facilitates studies in systems where direct fitness measurements are difficult. The latter is possible for the *Arabidopsis* example due to a relationship between seed number (W) and fruit number (x) established by Westerman and Lawrence (1970): $W = 23.74 \bullet x - 12.31$. Mean fruit number is 5.98, and so by the method described above for obtaining the error in the Heschel *et al.* (2005) study, we find that selection gradients via total seed production would be 9% larger than those via fruit production.

We have focused primarily on the inference of direct selection gradients. The same basic conditions apply to performance-based and performance-mediated analysis of extended selection gradients (Morrissey 2014). The approaches we suggest here for obtaining direct selection gradients can also yield selection differentials, assuming multivariate normality of phenotype, by multiplying gradients by phenotypic variances (or in the multivariate case, pre-multiplying the vector of directional selection gradients by the inverse of the **P** matrix, and similar operations for non-linear selection differentials given in Lande and Arnold 1983). Similarly, while we have focused on the calculation of unit-variance standardised selection gradients (as is most common; Lande and Arnold 1983; Kingsolver *et al.* 2012), known relationships among different standardisations of selection gradients (Hereford *et al.* 2004) are immediately applicable to performance-mediated selection gradient estimates.

We suggest that performance-based selection analyses should not be assumed to have the same justification in evolutionary theory as those based on fitness and fitness components. For example, performance-based gradients should be excluded from meta-analyses, as the broad comparisons across traits, taxa, etc., made in meta-analysis are justified by evolutionary quantitative genetic theory. However, this does not mean that trait-performance relationships should not be considered useful for qualitative inference of natural selection. In particular, insofar as it is reasonable to assume that performance-fitness relationships are monotonic, functions relating traits to performance should be representative of major aspects of the shape of trait-fitness relationships, such as the existence of fitness minima or maxima. We hope that by (*a*) formulating clear conditions under which the use of performance measures as proxies for fitness in selection gradient analysis is justified, and (b) highlighting more general ways of using performance measures, when quantitative data about the performance-fitness relationship are available (including from other individuals or other studies), studies of selection can be expanded to a wider range of taxa.

Acknowledgements

We thank H.S. Callahan and colleagues for their efforts conducting the original research which generated the data used in our example, S.M. Scheiner, M.M. Ferguson, and M.J. Janeiro, for providing constructive feedback and comments on the manuscript. ODF is supported by the Natural Sciences and Engineering Research Council of Canada and the Madame Vigdís

Finnbogadóttir Scholarship. MBM is supported by a University Research Fellowship from the Royal Society (London).

Author contributions

MBM and ODF conceived the ideas, MBM designed the methodology and derived the theoretical results, ODF collected the literature-based data, and both authors both contributed critically to the drafts, and gave final approval for publication.

References

Arnold, S. J. (1983). Morphology, performance and fitness. *American Zoologist*, *23*(2), 347–361.

Arnold, S. J. (2003). Performance surfaces and adaptive landscapes. *Integrative and Comparative Biology, 43*(3), 367-375.

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

Arntz, A. M., DeLucia, E. H., & Jordan, N. (2000). Fitness effects of a photosynthetic mutation across contrasting environments. *Journal of Evolutionary Biology*, *13*(5), 792–803.

Blueweiss, L., Fox, H., Kudzma, V., Nakashima, D., Peters, R., & Sams, S. (1978). Relationships between body size and some life history parameters. *Oecologia*, *37*(2), 257–272.

Bolnick, D. I., & Lau, O. L. (2008). Predictable patterns of disruptive selection in stickleback in postglacial lakes. *The American Naturalist*, *172*(1), 1–11

Carpenter, F. L., Hixon, M. A., Beuchat, C. A., Russell, R. W., & Paton, D. C. (1993). Biphasic mass gain in migrant hummingbirds: body composition changes, torpor, and ecological significance. *Ecology*, *74*(4), 1173–1182.

Donovan, L. A., Dudley, S. A., Rosenthal, D. M., & Ludwig, F. (2007). Phenotypic selection on leaf water use efficiency and related ecophysiological traits for natural populations of desert sunflowers. *Oecologia*, *152*(1), 13–25.

Ehdaie, B., Alloush, G. A., Madore, M. A., & Waines, J. G. (2006). Genotypic variation for stem reserves and mobilization in wheat. *Crop Science*, *46*(5), 2093–2103.

Endler, J. A. (1986). *Natural selection in the wild*. New Jersey: Princeton University Press.

Forseth, T., Ugedal, O., Jonsson, B., & Fleming, I. A. (2003). Selection on Arctic charr generated by competition from brown trout. *Oikos*, *101*(3), 467–478.

González-Gómez, P. L., & Estades, C. F. (2009). Is natural selection promoting sexual dimorphism in the Green-backed Firecrown Hummingbird (*Sephanoides Sephaniodes*)? *Journal of Ornithology*, *150*(2), 351–356.

Grant, B. R. (1985). Selection on bill characters in a population of Darwin's finches: *Geospiza conirostris* on Isla Genovesa, Galapagos. *Evolution*, *39*(3), 523–532.

Greenberg, C. H. (2000). Individual variation in acorn production by five species of southern Appalachian oaks. *Forest Ecology and Management*, *132*(2), 199–210.

Grether, G. F. (1996). Intrasexual competition alone favors a sexually dimorphic ornament in the rubyspot damselfly *Hetaerina americana*. *Evolution*, *50*(5), 1949–1957.

Hakes, A. S., & Cronin, J. T. (2011). Resistance and tolerance to herbivory in *Solidago altissima* (Asteraceae): Genetic variability, costs, and selection for multiple traits. *American Journal of Botany*, *98*(9), 1446–1455.

Healey, M. C. (1982). Timing and relative intensity of size-selective mortality of juvenile chum salmon (*Oncorhynchus keta*) during early sea life. *Canadian Journal of Fisheries and Aquatic Sciences*, *39*(7), 952–957.

Hereford, J., Hansen, T. F., & Houle, D. (2004). Comparing strengths of directional selection: how strong is strong? *Evolution*, *58*(10), 2133–2143.

Heschel, M. S., Hausmann, N., & Schmitt, J. (2005). Testing for stress-dependent inbreeding depression in *Impatiens capensis* (Balsaminaceae). *American Journal of Botany*, *92*(8), 1322–1329.

Hiebert, S. (1993). Seasonal changes in body mass and use of torpor in a migratory hummingbird. *The Auk*, *110*, 787–797.

Hoffmann, A. A., Ratna, E., Sgro, C. M., Barton, M., Blacket, M., Hallas, R., De Garis, S. & Weeks, A. R. (2007). Antagonistic selection between adult thorax and wing size in field released Drosophila melanogaster independent of thermal conditions. *Journal of evolutionary biology*, *20*(6), 2219-2227.

Hughes, T. P., & Connell, J. H. (1987). Population dynamics based on size or age? A reef-coral analysis. *American Naturalist*, *129*(6), 818–829.

Janzen, F. J., & Stern, H. S. (1998). Logistic regression for empirical studies of multivariate selection. *Evolution*, *52*(6), 1564–1571.

Jensen, J. L. W. V. (1906). On the convex functions and inequalities between mean values. Acta Mathematica. Djursholm, 30, 175-193.

Jordan, N. (1996). Effects of the triazine-resistance mutation on fitness in *Amaranthus hybridus* (smooth pigweed). *Journal of Applied Ecology*, *33*, 141–150.

Kingsolver, J. G., & Diamond, S. E. (2011). Phenotypic selection in natural populations: What limits directional selection? *The American Naturalist*, *177*(3), 346–357.

Kingsolver, J. G., Diamond, S. E., Siepielski, A. M., & Carlson, S. M. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, *26*(5), 1101–1118.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., Hoang, A., Gibert, P. & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*, *157*(3), 245–261.

Kingsolver, J. G., & Huey, R. B. (2003). Introduction: the evolution of morphology, performance, and fitness. *Integrative and Comparative Biology*, *43*(3), 361–366.

Kingsolver, J. G., & Pfennig, D. W. (2004). Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution*, *58*(7), 1608–1612.

Lande, R., & Arnold, S. J. (1983). The measurement of selection on correlated characters. *Evolution*, *37*(6), 1210–1226.

Ludwig, F., Rosenthal, D. M., Johnston, J. A., Kane, N., Gross, B. L., Lexer, C., Dudley, S.A., Rieseberg, L.H. & Donovan, L. A. (2004). Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. *Evolution*, *58*(12), 2682–2692.

Lush, J. L. (1937). *Animal breeding plans.* Ames, Iowa.

Lynch, M., & Walsh, B. (1998). Genetics and analysis of quantitative traits (Vol. 1). Sunderland, MA: Sinauer.

Martin, C. H. (2012). Weak disruptive selection and incomplete phenotypic divergence in two classic examples of sympatric speciation: Cameroon Crater Lake cichlids. *The American Naturalist*, *180*(4), E90–E109.

Martin, R. A., & Pfennig, D. W. (2009). Disruptive selection in natural populations: The roles of ecological specialization and resource competition. *The American Naturalist*, *174*(2), 268–281.

Mitchell-Olds, T., & Bergelson, J. (1990a). Statistical genetics of an annual plant, *Impatiens capensis*. II. Natural selection. *Genetics*, *124*(2), 417–421.

Mitchell-Olds, T., & Bergelson, J. (1990b). Statistical genetics of an annual plant, *Impatiens capensis*. I. Genetic basis of quantitative variation. *Genetics*, *124*(2), 407–415.

Mitchell-Olds, T., & Shaw, R. G. (1987). Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution*, *41*(6), 1149–1161.

Monro, K., Poore, A. G. B., & Brooks, R. (2007). Multivariate selection shapes environment-dependent variation in the clonal morphology of a red seaweed. *Evolutionary Ecology*, *21*(6), 765–782.

Morrissey, M. B. (2014a). Selection and evolution of causally covarying traits. *Evolution*, *68*(6), 1748–1761.

Morrissey, M. B. (2014b). In search of the best methods for multivariate selection analysis. *Methods in Ecology and Evolution*, *5*(10), 1095–1109.

Morrissey, M. B. (2015). Evolutionary quantitative genetics of nonlinear developmental systems. *Evolution*, *69*(8), 2050–2066.

Morrissey, M. B., & Hadfield, J. D. (2012). Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, *66*(2), 435–442.

Ossler, J. N., Zielinski, C. A., & Heath, K. D. (2015). Tripartite mutualism: Facilitation or trade-offs between rhizobial and mycorrhizal symbionts of legume hosts. *American Journal of Botany*, *102*(8), 1332–1341.

Persson, L., Byström, P., & Wahlström, E. (2000). Cannibalism and competition in Eurasian perch: population dynamics of an ontogenetic omnivore. *Ecology*, *81*(4), 1058–1071.

Pfennig, D. W., Mabry, A., & Orange, D. (1991). Environmental causes of correlations between age and size at metamorphosis in *Scaphiopus multiplicatus*. *Ecology*, *72*(6), 2240–2248.

Power, M., Dempson, J. B., Reist, J. D., Schwarz, C. J., & Power, G. (2005). Latitudinal variation in fecundity among Arctic charr populations in eastern North America. *Journal of Fish Biology*, *67*(1), 255–273.

Price, G. R. (1970). Selection and covariance. *Nature*, *227*, 520–21.

Procter, D. S., Moore, A. J., & Miller, C. W. (2012). The form of sexual selection arising from male-male competition depends on the presence of females in the social environment: Social influences on sexual selection. *Journal of Evolutionary Biology*, *25*(5), 803–812.

Ramírez-Valiente, J. A., Valladares, F., Delgado, A., Nicotra, A. B., & Aranda, I. (2015). Understanding the importance of intrapopulation functional variability and phenotypic plasticity in *Quercus suber*. *Tree Genetics & Genomes*, *11*(3), 1–11.

Robertson, A. (1966). A mathematical model of the culling process in dairy cattle. *Animal Production*, *8*(1), 95–108.

Scheiner, S. M., Mitchell, R. J., & Callahan, H. S. (2000). Using path analysis to measure natural selection. *Journal of Evolutionary Biology*, *13*(3), 423–433.

Schluter, D. (1988). Estimating the form of natural selection on a quantitative trait. *Evolution*, *42*(5), 849–861.

Schmitt, J., Eccleston, J., & Ehrhardt, D. W. (1987). Dominance and suppression, size-dependent growth and self-thinning in a natural *Impatiens capensis* population. *The Journal of Ecology*, *75*, 651–665.

Shaw, R. G., Geyer, C. J., Wagenius, S., Hangelbroek, H. H., & Etterson, J. R. (2008). Unifying life-history analyses for inference of fitness and population growth. *The American Naturalist*, *172*(1), E35–E47.

Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, *12*(11), 1261–1276.

Siepielski, A. M., Gotanda, K. M., Morrissey, M. B., Diamond, S. E., DiBattista, J. D., & Carlson, S. M. (2013). The spatial patterns of directional phenotypic selection. *Ecology Letters*, *16*(11), 1382–1392.

Stauffer, T. M. (1976). Fecundity of coho salmon (*Oncorhynchus kisutch*) from the Great Lakes and a comparison with ocean salmon. *Journal of the Fisheries Board of Canada*, *33*(5), 1150–1155.

Stinchcombe, J. R., Agrawal, A. F., Hohenlohe, P. A., Arnold, S. J., & Blows, M. W. (2008). Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution*, *62*(9), 2435–2440.

Svanbäck, R., & Persson, L. (2009). Population density fluctuations change the selection gradient in Eurasian Perch. *The American Naturalist*, *173*(4), 507–516.

Tucić, B., Tomić, V., Avramov, S., & Pemac, D. (1998). Testing the adaptive plasticity of *Iris pumila* leaf traits to natural light conditions using phenotypic selection analysis. *Acta Oecologica*, *19*(6), 473–481.

Valluru, R., Link, J., & Claupein, W. (2011). Natural variation and morpho-physiological traits associated with water-soluble carbohydrate concentration in wheat under different nitrogen levels. *Field Crops Research*, *124*(1), 104–113.

van Tienderen, P. H. (2000). Elasticities and the link between demographic and evolutionary dynamics. *Ecology*, *81*(3), 666–679.

van den Berghe, E.P. & Gross, M.R. (1989). Natural selection resulting from female breeding competition in a Pacific salmon (coho: *Oncorhynchus kisutch*). *Evolution.* 43(1), 125-140.

Wade, M. J., & Kalisz, S. (1990). The causes of natural selection. Evolution, 44(8), 1947-1955.

Waller, D. M. (1979). The relative costs of self-and cross-fertilized seeds in *Impatiens capensis* (Balsaminaceae). *American Journal of Botany*, *66*, 313–320.

Westerman, J. M., & Lawrence, M. J. (1970). Genotype-environment interaction and developmental regulation in Arabidopsis thaliana. I. Inbred lines; description. *Heredity*, *25*, 609–627.

Winn, A. A., & Miller, T. E. (1995). Effect of density on magnitude of directional selection on seed mass and emergence time in *Plantago wrightiana* Dcne.(Plantaginaceae). *Oecologia*, *103*(3), 365–370.

Wootton, R. J. (1973). Fecundity of the three-spined stickleback, *Gasterosteus aculeatus* (L.). *Journal of Fish Biology*, *5*(6), 683–688.

Table 1: Publications containing selection gradients calculated using performance measures in place of fitness, including details of purported relationship between performance/fitness proxy and fitness (component). We reviewed the literature citing Lande and Arnold (1983) and/or Arnold (1983) and containing keyword ‘proxy’ to identify these 19 publications, and consulted citations within to explore justifications for substituting performance measures for direct fitness measures. Most studies refer to multiple sources of evidence, both qualitative and quantitative, to justify use of performance as a fitness proxy. For brevity we have selected only the most direct correlational data, showing the greatest fit. It should also be noted that while some performance-fitness associations were measured on the same populations as the selection gradients, it is not uncommon for evidence of the association to stem from different genera or using different performance measures.

Selection gradients published in:	**Study species**	**Proxy estimated**	**Purported proxy-fitness relationship**	**Citation for relationship**	**Relationship details,** from citation unless otherwise stated		
					Fitness measure / w	**Proxy measure / y**	**Association / details of fit (where available)**
Ramirez-Valiente *et al*. 2015	Cork oak (*Quercus suber*)	Aboveground growth rate	linear (via size) with total fruit number	Greenberg 2000	Mean fruit number	Basal area (m2) of white oak (*Quercus alba*)	Linear fit: W = 5239.61y - 26.88 ($r^2 = 0.2677$)
Heschel *et al*. 2005	Orange jewelweed (*Impatiens capensis*)	Biomass	linear with reproductive fitness	Waller 1979	Total seed number	Estimated biomass (g)	Linear fit: Estimated from his Table 1: W = 29.724y -5.9019 (r^2=0.8419)
Mitchell-Olds and Bergelson 1990a and 1990b	Orange jewelweed (*Impatiens capensis*)	Relative adult size	Linear with total seed production	Waller 1979	Total seed number	Estimated biomass (g)	Linear fit: Estimated from his Table 1: W = 29.724y -5.9019 (r^2=0.8419)
Tucic *et al.* 1998	Pygmy iris (*Iris pumila*)	Biomass	linear with survival	Schmitt *et al.* 1987	Last date alive (days)	Estimated biomass (g) of Orange jewelweed (*Impatiens capensis*)	Linear fit: Estimated from their Figure 6: W =246.67y + 204 (stated r_s=0.75):
Donovan *et al*. 2007	Western/desert sunflower (*Helianthus anomalus* and *H. deserticola*)	Vegetative biomass	assumed linear with number of reproductive units	Own study	Number of reproductive units	Vegetative biomass	Linear fit assumed: stated r^2=0.96
Winn and Miller 1995	Wright’s plantain (*Plantago wrightiana*)	Biomass	assumed linear with fecundity	Own study	Seed number	Biomass	Linear fit assumed: r^2=0.97, n=14
Ossler *et al*. 2015	Partridge pea (*Chamaecrista fasciculata)*	Aboveground biomass	linear with seed number	Own study	Seed number	Aboveground biomass	Linear fit: r = 0.75, n=100

Arntz *et al*. 2000	Green amaranth *(Amaranthus hybridus)*	Reproductive mass	linear with seed number	Jordan 1996	Seed production	Reproductive biomass	Linear fit: r=0.98, n=51
Ludwig *et al*. 2004	Western sunflower (*Helianthus anomalus*)	Vegetative biomass	linear (via reproductive biomass) with fecundity	Own study	Reproductive biomass	Vegetative biomass	Linear fit: r=0.75, n=104
Martin and Pfennig 2009	Spadefoot toad (*Spea multiplicata*)	Larval body size	linear with survival, and indirectly via adult body size	Pfennig *et al*. 1991	Age at death (days)	Wet mass at metamorphosis (g)	Linear fit: Estimated from their Figure 3: W = 29.234y + 15.765 (stated r_s=0.83)
Bolnick and Lau 2008	Three-spine stickleback (*Gasterosteus aculeatus*)	Growth rate	linear with survival and reproductive success	Wootton 1973	Number of eggs	Total length (mm)	Linear fit: W = 8.089y - 301 (r=0.89)
Martin 2012	Cameroon cihlid complexes (*Stomatepia mariae*/*pindu,* and *Tilapia fusiforme/deckerti /ejagham*)	Growth rate	linear with survival	Healey 1982	Survival	Scale circuli width (mm) of chum salmon (*Oncorhynchus keta*)	Linear fit: Estimated from his Table 4: W = 67.5y - 1.195 (r^2= 0.9075)
Monro *et al*. 2007	Red algae (*Asparagopsis armata*)	Clonal growth	monotonic (via size) with survival	Hughes and Connell 1987	Survival %	Surface area (cm2) of coral from genera *Acropora, Porites, Pocillopora*	Logarithmic fit: Estimated from their Table 2: W = 14.861 ln(y) + 31.812 (r^2=0.90)
Svanback and Persson 2009	Eurasian perch (*Perca fluviatilis*)	Condition factor	assumed monotonic with survival and reproduction	Persson *et al*. 2000	Citation provides population level associations between mortality and condition		
Gonzalez-Gomez & Estades 2008	Firecrown hummingbird (*Sephanoides Sephaniodes*)	Condition factor	assumed monotonic with survival	Carpenter *et al.* 1993; Hiebert 1993	Citations provide population-level associations among body mass and thermoregulatory or migratory traits of Rufous hummingbird (*Selasphorus rufus*).		
Valluru *et al.* 2011	Einkorn (*Triticum Monococcum*) and Emmer wheat (*Triticum*	Energy stores	assumed monotonic with growth, phenology, yield	Ehdaie *et al.* 2006	Citation reviews contribution of energy stores to yield (5 to 20% non-stressed) and dry grain mass (22 to 60% under stress)		

	dicoccum)			
Procter *et al*. 2012	Leaf-footed cactus bug (*Narnia femorata*)	Male-Male competition	assumed monotonic (via territory) with mating success	Not provided
Forseth *et al*. 2003	Arctic charr (*Salvelinus alpinus*)	Growth rate	unknown	Not provided
Hoffmann *et al*. 2007	Common fruit fly (*Drosophila melanogaster*)	Dispersal distance	unknown	Not provided

Table 2: Performance-fitness functions in the *Arabidopsis* data, detailing linear and non-linear fits and relative and absolute performance. Linear fits were calculated as simple linear regressions; non-linear fits were calculated as either quadratic regression models or as generalized linear models assuming Poisson errors and using a log link function. Standard errors in parentheses. The coefficients of the linear regression of fitness on relative performance (i.e., height divided by mean height) in (a) describe the error in performance-based selection analysis. The quadratic generalised linear model regression in (b) is the regression of absolute fitness on log height, as used in the non-linear performance-mediated selection analysis.

	Absolute performance	**Relative performance**	
(a) linear fits	Fruit number regressed on inflorescence height (cm)	Fruit number regressed on relative height	Relative fruit number regressed on relative height
Intercept (*m*)	– 2.968 (0.348; P<0.005)	– 2.968 (0.35; P<0.005)	– 0.474 (0.056; P<0.005)
Slope (*n*)	1.246 (0.042; P<0.005)	9.232 (0.31; P<0.005)	1.474 (0.050; P<0.005)
(b) non-linear fits	Quadratic regression of fruit number on inflorescence height (cm)	Quadratic GLM (log-link, Poisson errors)	Quadratic regression of relative fruit number on relative height
Intercept	– 0.074 (0.599; P=0.90)	– 0.639 (0.339; P=0.06)	– 0.012 (0.096; P=0.90)
Linear	0.443 (0.144; P<0.005)	0.919 (0.326; P<0.005)	0.524 (0.171; P<0.005)
Quadratic	0.045 (0.008; P<0.005)	0.132 (0.078; P=0.09)	0.394 (0.068; P<0.005)

Table 3: Test of conditional independence of fitness and traits, mediated by performance, through a multiple regression of relative fitness (fruit number) on standardised traits (leaf number, bolting day), and performance (inflorescence height). Table summarises fitness-based selection gradients from (a) only linear terms, and (b) linear, quadratic, and interaction terms, with standard errors in parentheses. Qualitatively, the models both indicate effects of traits on fitness that are not mediated by performance.

Variance-standardised traits	**Linear terms**	**Linear, quadratic, and interaction terms**
Intercept	1.000 (0.024, P<0.005)	0.893 (0.037, P<0.005)
Bolting day	-0.069 (0.024, P<0.005)	-0.111 (0.034, P<0.005)
Leaf number	0.107 (0.029, P<0.005)	0.111 (0.029, P<0.005)
Inflorescence height	0.652 (0.030, P<0.005)	0.575 (0.032, P<0.005)
Bolting day^2		0.0.037 (0.041, P=0.37)
Leaf number^2		-0.041 (0.045, P=0.36)
Inflorescence height^2		0.098 (0.046, P=0.04)
Bolting day * leaf number		-0.002 (0.030, P=0.95)
Inflorescence height * bolting day		-0.05 (0.032, P=0.10)
Inflorescence height * leaf number		0.094 (0.034, P=0.01)

Table 4: Trait-performance relationships for non-linear performance-mediated selection gradient analysis. The model is a quadratic regression of log inflorescence height (performance) on variance standardised traits. Values in parentheses are standard errors. See Figure 5a for a visualisation.

Variance-standardised traits	**Linear, quadratic, and interaction terms**
Intercept	1.842 (0.037, P<0.005)
Bolting day	-0.143 (0.038, P<0.005)
Leaf number	0.328 (0.027, P<0.005)
Bolting day^2	0.064 (0.022, P=0.01)
Leaf number^2	-0.032 (0.017, P=0.06)
Bolting day * leaf number	0.051 (0.023, P=0.03)

Appendix: conditions for non-linear selection analysis with fitness proxies

The directional selection gradient is the average derivative of relative fitness with respect to phenotype (see eq. 5 in the main text), and equivalently, the quadratic selection gradient is the average second derivative of relative fitness with respect to phenotype,

A1 $$\gamma = \frac{E[\frac{d^2W}{dz^2}]}{\bar{W}}.$$

Assume a trait-performance relationship of

A2 $$E[y_i|z_i] = f(z_i) = a + bz_i + \frac{1}{2}gz_i{}^2$$

If y was relative fitness, the fit of this equation by least squares would result in b and g being interpretable as directional and quadratic selection gradients, if the z variable was mean-centered and normally distributed. This is thus the regression equation that would be implemented in an analysis substituting a performance measure for fitness by dividing performance by its mean, and using the resulting relativized performance measure as the response variable.

Assume a linear performance fitness relationship, as in eq. 2, and as invoked in studies using performance as a substitute for fitness,

A3 $$E[W_i|y_i] = f(y_i) = m + ny_i.$$

Fitness as a function of phenotype is thus

A4 $$E[W_i|z_i] = W(z_i) = m + n(a + bz_i + \frac{1}{2}gz_i{}^2)$$

The first and second derivatives of absolute fitness, W, with respect to traits, z, are

A5a $$\frac{dW}{dz} = n(b + gz), \text{ and}$$

A5b $$\frac{d^2W}{dz^2} = ng$$

Since A4 is quadratic, the expectation taken over some distribution of phenotype can be obtained by taking a Taylor series up to second order, since derivatives of higher than second order of W with respect to z are zero. Thus, mean fitness is given using A4 and A5b

A6 $$\bar{W} = W(E[z]) + \frac{1}{2}W''^{(\bar{z})}E[(z - E[z])^2] = m + \frac{n}{2}\left(2a + 2b\mu + g(\mu^2 + \sigma^2)\right),$$

where μ is mean phenotype and σ^2 is the phenotypic variance.

Since A5a is linear, the expected first derivative is simply the derivative evaluated at mean phenotype, i.e., $E[\frac{dW}{dz}] = n(b + g\mu)$, and since A5b is constant with respect to z, $E\left[\frac{d^2W}{dz^2}\right] = ng$. Thus, when the performance function is evaluated via quadratic regression, and the performance-fitness relationship is linear, the directional and quadratic selection gradients are

A7a $$\beta = \frac{E[\frac{dW}{dz}]}{\bar{W}} = \frac{n(b+g\mu)}{m+\frac{n}{2}(2a+2b\mu+g(\mu^2+\sigma^2))}, \text{ and}$$

A7b $$\gamma = \frac{E[\frac{d^2W}{dz^2}]}{\bar{W}} = \frac{ng}{m+\frac{n}{2}(2a+2b\mu+g(\mu^2+\sigma^2))},$$

If phenotype is mean-centred (μ = 0) and standardized to unit variance (σ^2 = 1), the selection gradients simplify to

A8a $$\beta_s = \frac{nb}{m+\frac{n}{2}(2a+g)}, \text{ and}$$

A8b $$\gamma_s = \frac{ng}{m+\frac{n}{2}(2a+g)},$$

To test whether the assumptions inherent to substituting relative values of a proxy for relative fitness include merely that the performance fitness is linear, or whether there are also assumptions about the slope and/or intercept, it is necessary to obtain a version of the expressions for the selection gradients that reflects the case where the mean performance is one. Because performance is a quadratic function of the trait, setting the intercept in equation A2 to one does not correspond to $\bar{W} = 1$, as in the main text. Rather, mean performance is a function of both a and g. Similarly to equation A6, mean performance is given by

A9 $$\bar{y} = a + \frac{1}{2}g.$$

Using $g = 2(1\text{-}a)$, versions of equations A8a&b when performance is relativized are

A10a $$\beta_s = \frac{nb}{m+\frac{n}{2}(2a+2(1-a))} = \frac{nb}{m+n}, \text{ and}$$

A10b $$\gamma_s = \frac{ng}{m+\frac{n}{2}(2a+2(1-a))} = \frac{ng}{m+n}.$$

From A10a&b, it is clear that the use of a fitness proxy assumes not only that the relative performance-fitness relationship is linear, but that the linear regression goes through the origin. If the regression of (absolute or relative) fitness on relative performance does not go through the origin, analysis of both directional and non-linear selection using performance proxies will be incorrect, by a factor of $\frac{m+n}{n}$ (as in the strictly linear case).

One possible analysis advocated in Arnold (2003) for inference of non-linear selection in the general trait-performance-fitness framework is the characterisation of both directional (as in Arnold 1983) and quadratic selection gradients as the product of their performance gradients, and the directional selection gradient of performance. In the framework for quantitative genetic interpretation of path coefficients in selection analysis proffered in Morrissey (2014) these coefficients would be considered ‘extended’ selection gradients if performance was conserved a trait (for example, in application of the equation for evolutionary prediction given in Morrissey 2014 simultaneously to traits and performance measures), or they could be considered ‘direct’ selection gradients in the sense of Lande (1979) and Lande and Arnold (1983), and would be applicable in an analysis where performance was not simultaneously considered as a trait.

Under either interpretation and application, it can be seen from equations A7a&b, and more readily from equations A8a&b that this interpretation (i.e., path analysis applied to quadratic analysis of performance gradients, and linear analysis of the selection gradient of performance) is justified. Since the performance-fitness relationship is linear, mean fitness is given as a function of mean performance as $\bar{W} = m + n\bar{y}$. From equation A9, $\bar{y} = \frac{1}{2}(2a + g)$, when the trait is mean-centred and standardized to unit variance (conditions used to simplify equations A8a&b), and so $\bar{W} = m + n(\frac{1}{2}(2a + g))$, which is the function in the denominator of equations A8a and b.

Arnold's (2003) suggestion that both trait-performance and performance-fitness relationships could be characterized by quadratic regressions, and path analysis subsequently applied to both quadratic and correlational terms will yield approximations of selection gradients. Because a function such as A3, giving fitness as a function of phenotype would contain higher polynomial terms if trait-performance and performance-fitness functions were both quadratic, analytical expressions for the resulting selection gradients would be much more complicated, and would depend on higher derivatives of the trait-fitness function, higher moments of the distribution of phenotype (i.e., the skew and the kurtosis), and the variance and other aspects of the distribution of residuals of the regression of performance on traits.

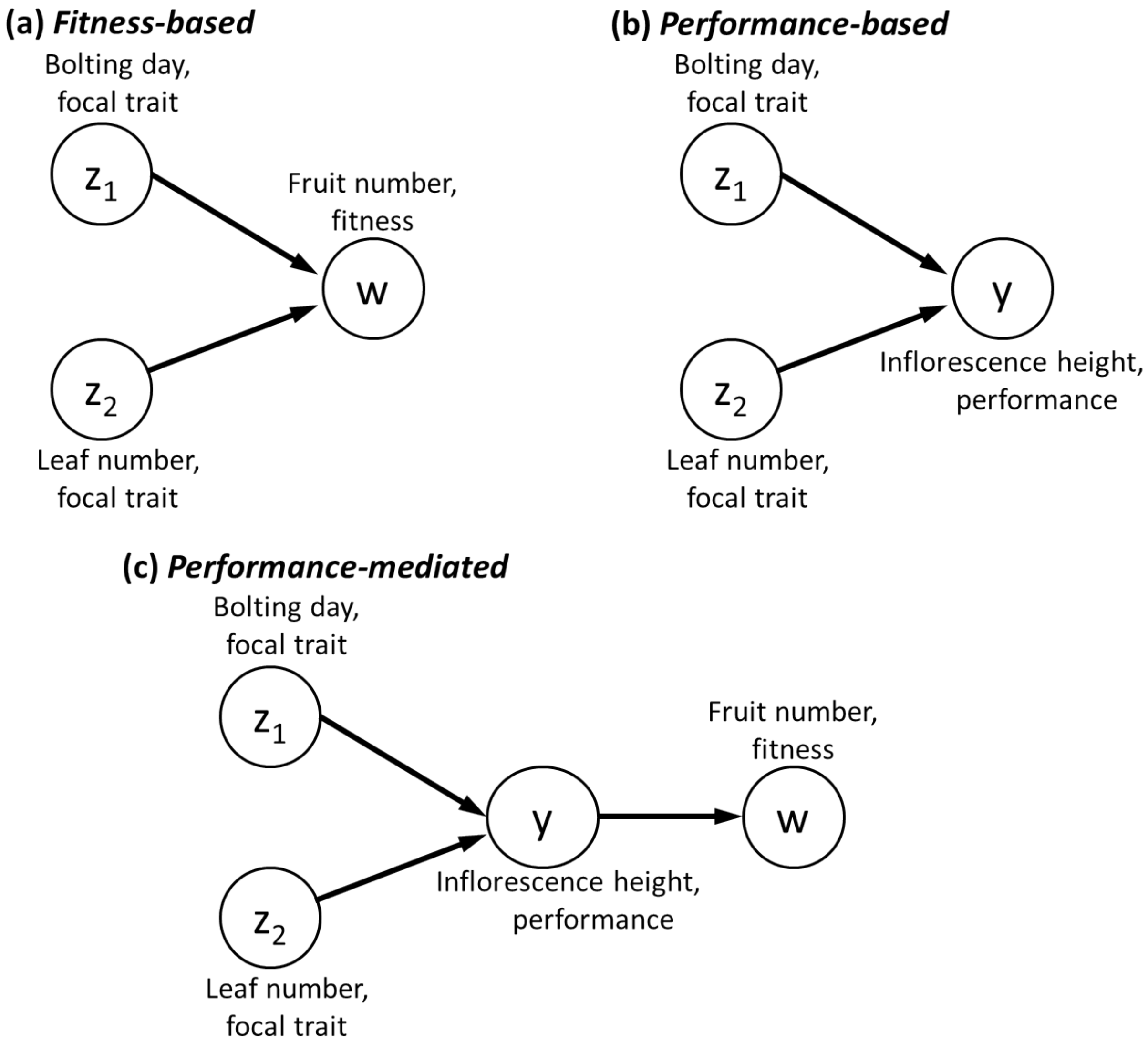

Figure 1: Illustration of fitness-based, performance-based, and performance-mediated selection analyses, as conducted in the *Arabidopsis* system. (a) In a fitness-based analysis, relative fitness, w, is regressed on focal traits, as in Lande and Arnold (1983; see also non-linear extensions therein). (b) In performance-based analyses, a measure of organismal performance is used as a proxy for fitness, and used directly to calculate relative fitness. (c) In performance-mediated selection analysis, path analysis is conducted, as in Arnold (1983; see Arnold 2003 and Morrissey 2015 and text here for non-linear extensions), based on a causal structure where focal traits influence the performance measure, which in turn influences relative fitness.

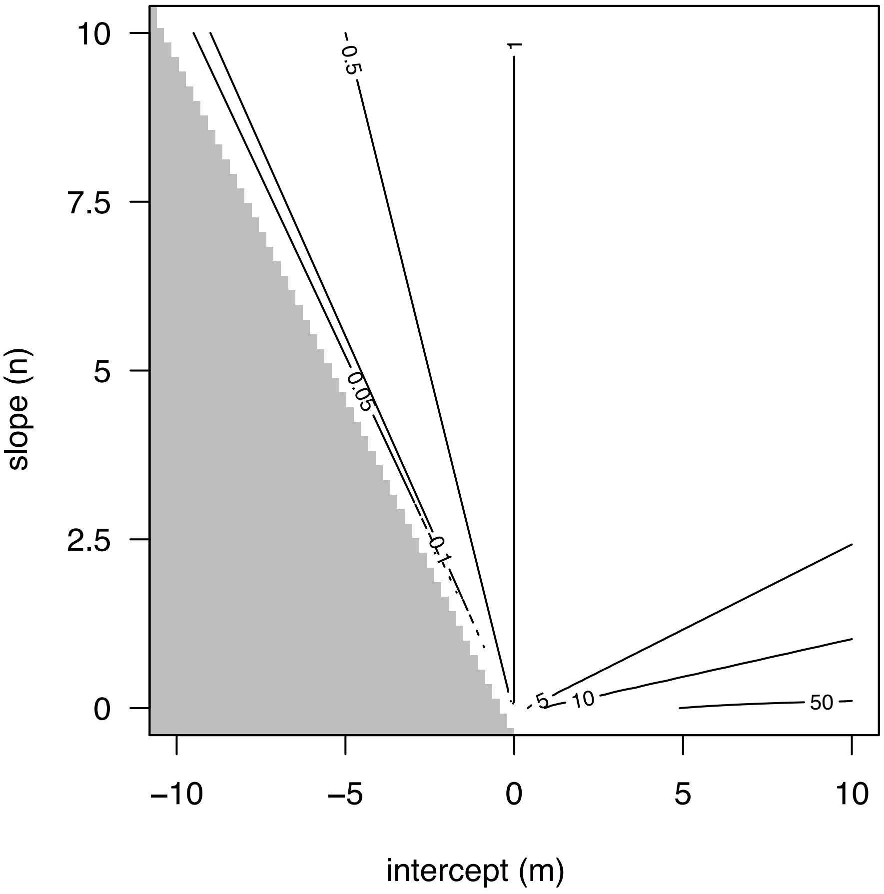

Figure 2: Proportional error (contours) introduced by use of performance-based selection gradients, as a function of the intercept (m) and slope (n) of the function describing a linear performance-fitness relationship. The relationship (equation 7) holds when performance is used on the relative scale (i.e., divided by its mean, as when it is used as a proxy for relative fitness). The proportional error, b/β (performance-based selection gradient / fitness-based selection gradient), increases in magnitude when the deviation of the intercept from 0 is large relative to the slope. When the intercept is the origin (m=0), there is no error (β/b=1), providing the performance-fitness relationship is linear. The grey area represents impermissible combinations of the slope and intercept; these parameters would correspond to negative mean fitness. When fitness is also expressed as relative fitness, then values of m and n are constrained to fall on the line m+n=1.

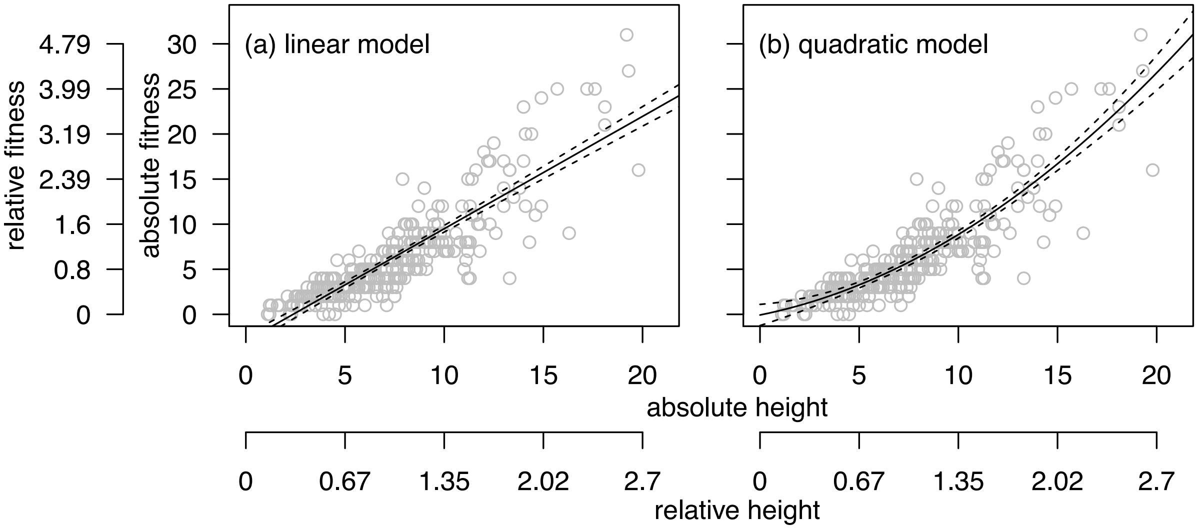

Figure 3: Visualisations of the performance-fitness relationship with (a) linear and (b) quadratic fits, where inflorescence height (cm) and fruit number are the performance and fitness measures, respectively, corresponding to tests of assumptions of performance-based selection analysis. (a) provides a test of the assumption in performance-based selection analysis that the regression of fitness on performance passes through the origin, and (b) tests the assumption of a linear performance-fitness relationship. Coefficients of the regressions in (a) and (b) are given in Table 2.

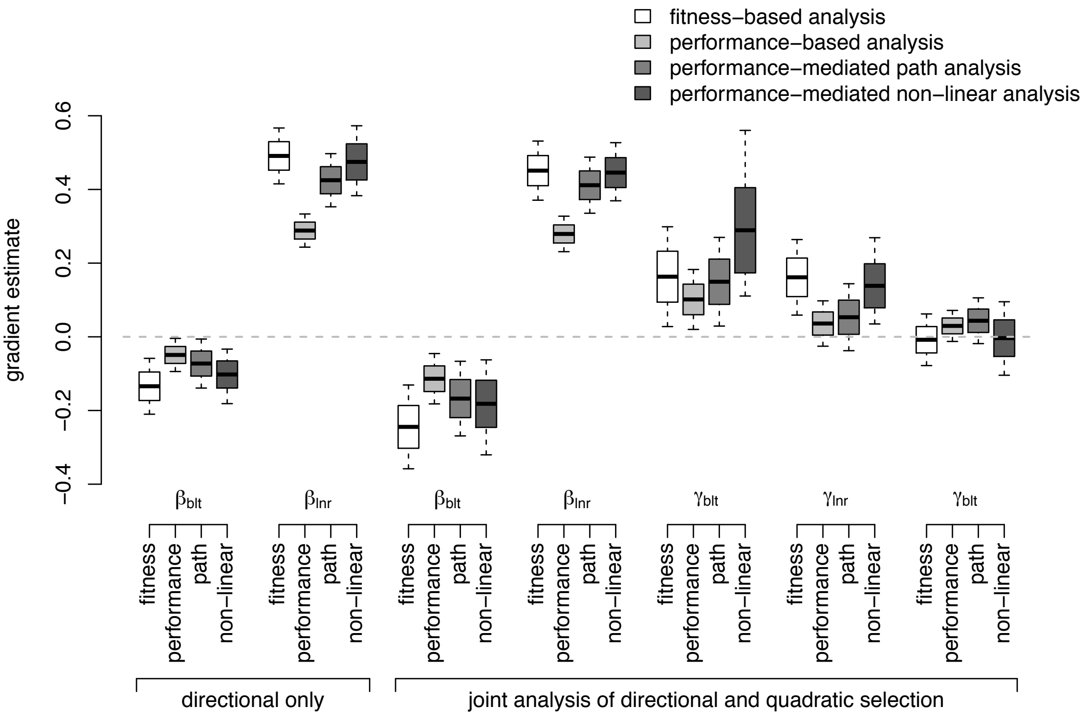

Figure 4: Comparison of fitness-based, performance-based and linear and non-linear performance-mediated selection gradients, for directional only and full linear & nonlinear selection analyses. Boxes depict estimate ± 1 standard error; whiskers depict 95% confidence intervals (± 1.96 x standard error, except for the performance-mediated non-linear analysis where CIs are based on the quantiles of the bootstrap distributions of the parameters). Traits are bolting day (*blt*) and leaf number (*lnr*), with *β* and *γ* depicting linear and quadratic selection gradients, respectively. Multiple regressions are of relative fruit number (for selection gradients) or relative inflorescence height (for performance-based selection gradients) on variance-standardised focal traits.

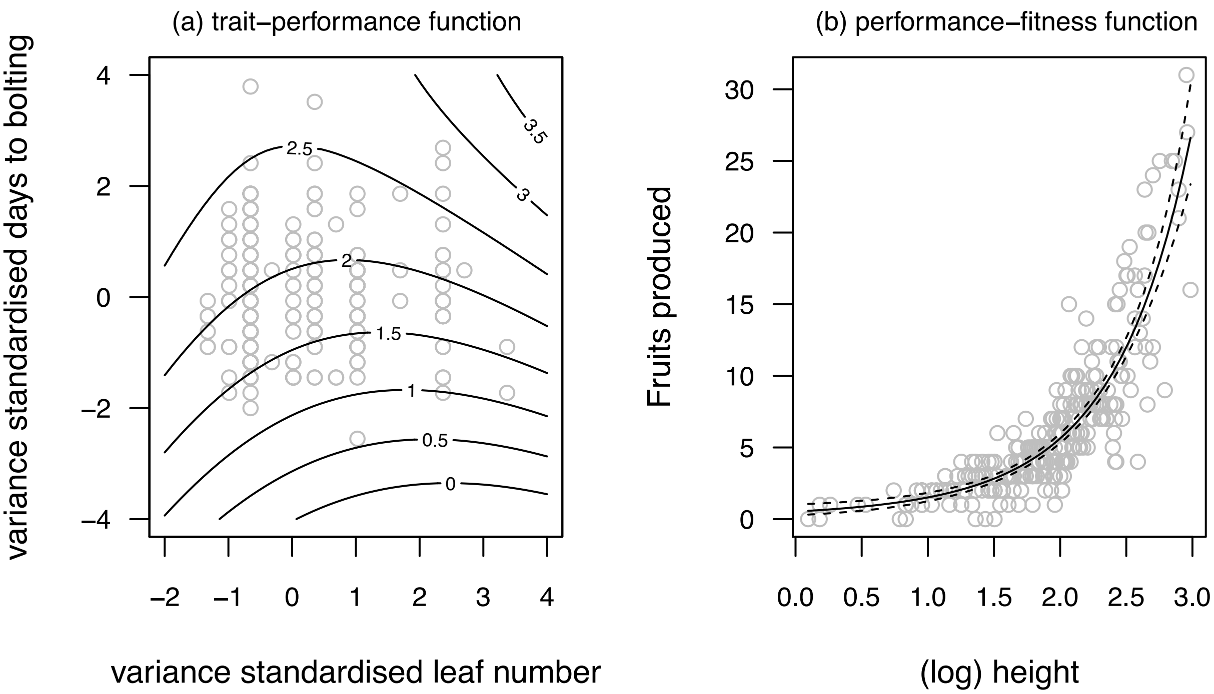

Figure 5: Non-linear trait-performance and performance-fitness functions used in the performance-mediated selection analysis allowing arbitrary functional forms. **(a)** The effect of traits on log inflorescence height (with this response variable visualised with contours). Coefficients corresponding to this figure can be found in Table 4. **(b)** The relationship between log inflorescence height (performance) and fruit number (fitness), demonstrating the appropriateness of a non-linear fit.