

Standardizing selection strengths to study selection in the wild: A critical comparison and suggestions for the future

Matsumura, S., Arlinghaus, R. and Dieckmann, U.

**IIASA Interim Report
2012**



Matsumura, S., Arlinghaus, R. and Dieckmann, U. (2012) Standardizing selection strengths to study selection in the wild: A critical comparison and suggestions for the future. IIASA Interim Report. IIASA, Laxenburg, Austria, IR-12-051 Copyright © 2012 by the author(s). <http://pure.iiasa.ac.at/10224/>

Interim Reports on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work. All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting repository@iiasa.ac.at



International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

Interim Report

IR-12-051

Standardizing selection strengths to study selection in the wild: A critical comparison and suggestions for the future

Shuichi Matsumura
Robert Arlinghaus
Ulf Dieckmann (dieckmann@iiasa.ac.at)

Approved by

Pavel Kabat
Director General and Chief Executive Officer

February 2015

Standardizing selection strengths to study selection in the wild:

A critical comparison and suggestions for the future

Shuichi Matsumura^{1,2,3,*}, Robert Arlinghaus^{2,4}, and Ulf Dieckmann¹

¹ Evolution and Ecology Program, International Institute for Applied Systems Analysis, Schlossplatz 1, 2361 Laxenburg, Austria

² Department of Biology and Ecology of Fishes, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany

³ Faculty of Applied Biological Sciences, Gifu University, Yanagido 1-1, 501-1193 Gifu, Japan

⁴ Inland Fisheries Management Laboratory, Department for Crop and Animal Sciences, Faculty of Agriculture and Horticulture, Humboldt-University of Berlin, Philippstrasse 13, Haus 7, 10115 Berlin, Germany

* Corresponding author: Shuichi Matsumura, Faculty of Applied Biological Sciences, Gifu University, Yanagido 1-1, 501-1193 Gifu, Japan, Phone +81-58-293-2874, Email matsumur@gifu-u.ac.jp

Abstract

We critically review the main approaches for standardizing and comparing selection differentials and gradients among traits, populations, and species and assess their differential merits. In particular, we explain why the most widespread approach to standardizing selection, which measures selection differentials in units of a trait's phenotypic standard deviation, is less appropriate for characterizing strengths of natural selection in wild populations. In contrast, an alternative standardization approach, which uses a trait's phenotypic mean in addition to its standard deviation, results in another dimensionless measure – the mean-standardized selection gradient. This measure offers some key benefits over alternative approaches – such as ease of interpretation, independence of a trait's variance, and the natural interfacing of ecological and evolutionary dynamics – and yet remains rarely used by evolutionary biologists. We explain how the more routine application of this measure will facilitate comparisons of selection strengths in the wild among traits, populations, and species.

Keywords

elasticity, natural selection, selection differential, selection gradient, standardization

Our world is filled with astonishingly diverse forms of organisms. Adaptation by natural selection to spatially heterogeneous and temporally changing environments is a critical driving force for bringing about this biodiversity (Darwin 1859). Phenotypes of organisms change according to both natural and anthropogenic selection pressures. Rapid phenotypic evolution resulting from human-induced environmental change – such as global warming, and the harvesting of wild animals and plants – can be particularly troublesome, because such evolution may have undesirable consequences, for example, population declines, ecosystem instabilities, and forgone yield (Kinnison and Hendry 2001, Palumbi 2001, Jørgensen et al. 2007, Darimont et al. 2009). Anthropogenic selection pressures are likely to be very strong (Darimont et al. 2009); yet, few studies have quantified and compared their strength, for example in the context of fisheries (Law 2007, Hutchings and Fraser 2008, Hard et al. 2008). To understand how various selective forces change phenotypes of animals and plants in the wild, an important first step is to develop robust methods that quantify and compare the power of natural and anthropogenic selection pressures acting on a diverse range of adaptive traits (Hendry 2005).

The general methodological framework for estimating the strength of selection has been developed by Lande and colleagues in the late 1970s and early 1980s (e.g., Lande and Arnold 1983). Because of its conceptual appeal and ease of application, a large number of empirical studies have subsequently estimated the strength of natural selection on adaptive traits in a wide range of animals and plants using the regression-based methods pioneered by Lande and colleagues (reviewed by Kingsolver et al. 2001, Hereford et al. 2004). Their main innovation was to show how selection strength can be measured through a univariate or multivariate linear regression of relative fitness, or components of fitness (such as survival at a particular life stage or the number of off-

spring; Arnold and Wade 1984), on the quantitative trait(s) of interest. Using this compelling approach, important insights have been generated into the mechanisms and patterns governing the direction and speed of evolutionary change, including eco-evolutionary feedback, (Hendry 2005, Kingsolver and Pfennig 2007). The accumulation of data on selection pressures has in turn enabled researchers to compare selection strengths among different traits, populations, and species (e.g., Endler 1986, Kingsolver et al. 2001, Hereford et al. 2004, Hendry 2005, Kingsolver and Pfennig 2007). Yet, the utility of such comparisons is still crucially limited by two fundamental problems. First, as different phenotypic traits have different units, any direct comparison of selection strengths among them is limited. Second, reported indicators of selection strength are often confounded by population- or trait-specific properties, such as a population's generation time or a trait's heritability, phenotypic variance, or phenotypic mean. These two problems hamper a complete understanding of selection responses of different traits in different populations experiencing different selective environments. To enable meaningful comparisons, selection strengths have to be adequately standardized in a way that ensures all confounding factors are removed.

The objective of this article is to review the differential merits of alternative ways of standardizing selection strengths. In particular, we highlight the important advantages of the mean-standardized measure, which render this approach the most useful approach for comparing selection strengths in wild populations of animals and plants. Complementing the commendable pioneering discussion by Hereford et al. (2004), our exposition is meant to provide an easily accessible one-stop overview of these matters for the benefit of all those interested in comparing strengths of selection in natural populations. We also provide a worked-out example and an overview of main steps to show

how to obtain and standardize measures of selection strength from theoretical and empirical data, respectively.

Selection differentials, responses, and gradients

One of the central measures of selection in traditional quantitative genetics and artificial breeding is the selection differential S . It measures the mean phenotypic value of individuals selected as parents, expressed relative to the mean phenotypic value of all individuals in the parental generation before selection (Falconer and Mackay 1996). In selection experiments or artificial breeding efforts, selection often operates through truncation, by admitting for reproduction only individuals whose phenotypic trait values exceed a given threshold (Fig. 1a). A form of viability selection, such truncation selection is recognized as the most efficient implementation of artificial directional selection (Crow and Kimura 1979). By contrast, truncation selection rarely occurs in natural populations, as there usually exists no threshold phenotypic value that would distinguish individuals that reproduce, or survive, from those that do not. For natural selection, instead, the selection differential S is obtained as the phenotypic mean of individuals weighted by their fitness, expressed relative to the population's phenotypic mean in which all individuals are weighted alike (Fig. 1c; see also the illustrative example in Box 1). The empirical measurement of selection differentials thus requires estimating the differential fitness of individuals in dependence on the phenotypic traits characterizing those individual. This is often difficult, or even practically impossible, especially when a trait's fitness consequences are distributed over an organism's full life cycle. Therefore, selection differentials are frequently calculated based on a fitness component (Arnold and Wade 1984), for example, survival at a particular life stage or the number of offspring produced over a certain period (e.g., a year). Such differentials more nar-

rowly measure the strength of selection acting on an adaptive trait at a particular stage of the life cycle.

A more readily observable measure of selection strength is the selection response R . This quantifies the amount of evolutionary (i.e., genetic) change in a phenotypic trait value between generations, and is thus defined as the change in a population's mean phenotypic value during one generation. In traditional quantitative genetics, the selection response is associated with the selection differential according to the breeder's equation,

$$R = h^2 S, \tag{1}$$

where $h^2 = \sigma_A^2 / \sigma_P^2$ is the narrow-sense heritability, and σ_A^2 and σ_P^2 are a population's additive genetic variance and its phenotypic variance, respectively (Falconer and Mackay 1996). It is thus evident that selection responses depend on population- and trait-specific properties (h^2 , or σ_A^2 and σ_P^2). As we will see further below in this section, an analogous conclusion applies to selection differentials, even though this is less commonly emphasized.

Methods for measuring natural selection on multiple traits in natural populations have been developed in the late 1970s and early 1980s by Lande and coworkers (e.g., Lande and Arnold 1983). These methods measure the selection gradient β , which is the average slope (or gradient) of a trait's fitness landscape experienced by a population in a given ecological environment. Such a fitness landscape represents how the fitness of individuals changes with a change in their trait value. It has been shown that β can be estimated as the coefficient of a linear regression of relative fitness on a phenotypic trait, or more generally, as a partial regression coefficient in the case of multivariate traits (Lande 1979, Lande and Arnold 1983). In this framework, the selection differential S is the covariance between the phenotypic trait and fitness, while the selection gradient β

is the covariance divided by the variance of the phenotypic trait. Generally, β , S , and R are related as follows,

$$\beta = S / \sigma_p^2 = R / \sigma_A^2. \quad (2)$$

Since β measures the sensitivity of fitness to changes in a trait's value (which can be approximated as a regression coefficient of relative fitness on the trait), it is a more direct measure of selection strength in natural populations than S or R . From equation 2 we see that the selection differential

$$S = \sigma_p^2 \beta, \quad (3)$$

confounds the selection gradient β with a population's phenotypic variance, which means that, just like R , S is not an ideal measure for comparing selection strengths among different traits, populations, or species.

In multivariate cases, selection is acting on n different traits, and equation 2 generalizes to

$$\mathbf{R} = \mathbf{G}\boldsymbol{\beta}, \quad (4)$$

where $\mathbf{R} = (R_1, R_2, \dots, R_n)^T$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_n)^T$ denote the vectors of multivariate selection responses and selection gradients, respectively. The $n \times n$ matrix \mathbf{G} is called the additive genetic variance-covariance matrix and comprises the additive genetic variances of all n traits, as well as the covariances for all $n(n-1)$ pairs of different traits (Lande and Arnold 1983). In this manner, the equation above captures both the direct evolutionary trait changes resulting from selection on a given focal trait and the indirect trait changes resulting from selection on other traits that are genetically correlated with the focal trait (for a more detailed discussion of multivariate cases, see Hansen and Houle 2008). In the present article, we will focus on univariate selection, since our main

interest is not in selection responses but in selection differentials and gradients for particular traits.

Standardizing selection strength

When researchers equipped with the aforementioned concepts, methods, and equations strive to compare selection strengths among different traits, a fundamental problem arises: neither the selection differential S , nor the selection response R , nor the selection gradient β are dimensionless. If the trait of interest is measured in kg, for instance, the selection differential, selection response, and selection gradient have units of kg, kg, and kg^{-1} , respectively. We illustrate the resultant difficulties by focusing on the selection differential S , the most traditional measure among these three. Let us consider an example from the literature on fisheries-induced evolution, which has attracted considerable attention in recent years (e.g., Jørgensen et al. 2007, Kuparinen and Merilä 2007): Hard et al. (2008) reviewed empirical estimates of selection differentials imposed by fishing, reporting selection differentials on body length in Atlantic cod *Gadus morhua* (ranging from -1 to $+2$ cm), on body weight in Atlantic salmon *Salmo salar* (-0.08 to -0.52 kg), and on body girth in sockeye salmon *Oncorhynchus nerka* (-0.6 to -3.6 mm). From an absolute comparison of these values, there is no way one could tell which trait is under the strongest selection. Therefore, it is evident that S – and analogously R and β – must be standardized before attempting any comparison of selection strengths.

There are two obvious approaches to this standardization. First, selection differentials can be measured in units of the phenotypic mean of the trait under selection, by considering the ratio S / μ_p . Representing in this manner a trait's proportional change relative to its mean is probably the most intuitive approach to standardization. Indeed, when the selection response R is measured in units of the phenotypic mean, R / μ_p , the

resultant ratio usefully describes the proportional change in the trait value from one generation to the next (Roff 2002). Interestingly, this type of standardization is rarely reported for selection differentials. Alternatively, selection differentials can be expressed in units of the phenotypic standard deviation of the trait under selection, by considering the ratio S / σ_p . As we will explain below, this currently most popular approach to standardizing selection strengths originated in the context of controlled selection experiments typical in artificial breeding studies of animals and plants, but unfortunately is not optimally suited to standardizing the strength of natural selection.

Recognition of a third standardized measure is a relatively recent development, since this additional measure is not as obviously constructed as the two described so far. Specifically, the selection differential can be standardized by jointly using the phenotypic mean and the phenotypic variance of the trait under selection, considering the ratio $\mu_p S / \sigma_p^2$. The motivation for this particular, at first sight seemingly odd, construction is to remove the phenotypic variance's confounding effect on the selection differential (equation 3), and yet obtain a dimensionless measure.

As already explained above, the selection gradient β measures the sensitivity of fitness to changes in a trait's value and therefore is considered a more direct measure of selection strength than S . According to equation 2, the second and third standardized measures of selection strength are related to the selection gradient β as

$$\beta_\sigma = S / \sigma_p = \sigma_p \beta \tag{5}$$

and

$$\beta_\mu = \mu_p S / \sigma_p^2 = \mu_p \beta. \tag{6}$$

We thus see that these measures can be interpreted as a fitness landscape's slope measured in units of, respectively, the phenotypic standard deviation σ_p or the phenotypic

mean μ_p . For this reason, the two measures were termed variance-standardized selection gradient (β_σ) and mean-standardized selection gradient (β_μ) by Hereford et al. (2004). In the present article, we call the former measure the SD-standardized selection gradient, because it actually is standardized by the phenotypic standard deviation (SD), and not by the phenotypic variance.

Standardizing selection strength by the phenotypic standard deviation

The SD-standardized selection gradient $\beta_\sigma = \sigma_p \beta = S / \sigma_p$ is also known as the selection intensity (or intensity of selection) i in traditional quantitative genetics (Falconer 1960, Bulmer 1980, Falconer and Mackay 1996, Roff 1997). In the context of truncation selection, the selection intensity i has a particularly compelling interpretation, since it then directly measures the proportion of individuals that are not selected for reproduction. If this proportion increases, the selection intensity i increases. What is important for comparative purposes, especially in artificial breeding efforts, is that the selection intensity is independent of the variability of the trait considered. If the selected proportion is the same, the selection intensity i will be the same, regardless of the variability of the trait (Fig. 1a,b; Falconer and Mackey 1996, Roff 1997). Because of its utility for describing truncation selection, i has long been considered a good measure of selection strength in general (Falconer and Mackey 1996).

This tradition of standardizing selection differentials by phenotypic standard deviations has continued even after measurements of selection strength were extended to natural populations. When Lande and Arnold (1983) applied their newly developed regression method to empirical data, they acknowledged the established tradition by explaining that the “observed selection differential on a particular character is often expressed in terms of phenotypic standard deviation”. Since then, the SD-standardized se-

lection gradient β_σ has been widely used in comparative studies on natural selection in the wild. For example, Kingsolver and colleagues (Kingsolver et al. 2001, Hoekstra et al. 2001, Kingsolver et al. 2011) reviewed an extensive body of empirical studies and conducted a meta-analysis of selection strengths in natural populations of animals and plants using the SD-standardized selection gradient.

Even outside the context of truncation selection, the SD-standardized selection gradient β_σ has a simple interpretation: it measures the change in relative fitness that results from increasing the trait value by 1 phenotypic standard deviation. For example, for $\beta_\sigma = 0.4$, shifting trait values up by 1 standard deviation from their phenotypic mean increases relative fitness by 40%. (One must keep in mind, however, that this interpretation is only permissible when selection is frequency-independent, or when the fraction of individuals whose trait values are shifted is negligible relative to the considered population as a whole, as otherwise the trait shift may reshape the fitness landscape.) Unfortunately, this intuitive interpretation does not bring out the fact that the SD-standardized selection gradient remains confounded by the variance of the focal trait under selection, and thus by an idiosyncratic property of the selected population. We will explain and discuss this limitation further below.

Standardizing selection strength by the phenotypic mean

Increased attention to the mean-standardized selection gradient $\beta_\mu = \mu_p \beta = \mu_p S / \sigma_p^2$ is a relatively recent development. Mean-standardized selection responses were already reported by Johnson et al. (1955). To our knowledge, Morgan and Schoen (1997) might have been the first to give systematic consideration to the specific advantages of using the mean-standardized selection gradient in quantitative genetics. Further discussion about the prospects of mean-standardized selection gradients can be found in van

Tienderen (2000) and Hereford et al. (2004). The latter study reanalyzed the database of Kingsolver et al. (2001) using the mean-standardized measure, and concluded that it was considered superior to the SD-standardized measure when selection strength on different traits was to be compared in natural populations. To understand this claim, it is helpful to look at mean-standardized measures in the context of the breeder's equation. Using equation 2 and applying mean-standardization not only to the selection differential, but also to all other variables, the breeder's equation 1 can be rewritten as

$$R / \mu_p = (\sigma_A / \mu_p)^2 (\mu_p \beta) \text{ or } r = a^2 \beta_\mu, \quad (7)$$

where the proportional selection response $r = R / \mu_p$ is predicted by the mean-standardized selection gradient $\beta_\mu = \mu_p \beta$ and the coefficient of genetic variation $a = \sigma_A / \mu_p$. The latter has been suggested as a suitable measure of evolvability (Houle 1992, Hansen et al. 2003), because it characterizes a population's ability to respond to selection. This can easily be understood since a larger genetic variation (SD, σ_A) relative to a small phenotypic mean (μ_p) implies a higher potential for evolution, hence evolvability. In equation 3, all variables are dimensionless, and are measured in units of the trait's phenotypic mean. As a result, selection strength (β_μ) and evolvability (a) are cleanly separated. On this basis, selection strengths, evolvabilities, and proportional selection responses can consistently be compared among different traits, populations, and species (Hansen et al. 2003, Hereford et al. 2004).

Analogous to the SD-standardized selection gradient β_σ , the mean-standardized selection gradient β_μ has a simple interpretation: it measures the change in relative fitness that results from doubling the trait value. For example, for $\beta_\mu = 0.4$, shifting trait values up by 100% increases relative fitness by 40% (the aforementioned caveat about

potential changes in the fitness landscape resulting from such a shift of course applies here, too).

Which measure of selection strength is more suitable for natural populations?

The meta-analyses of selection strengths in natural populations by Kingsolver et al. (2001) and Hereford et al. (2004) investigated similar data sets, but used different standardizations of selection strength and thereby came to qualitatively different conclusions. The former study concluded that natural selection was generally not strong, whereas the latter study suggested the opposite. With such fundamental conclusions at stake, the choice of standardization certainly is no trifling matter. The question thus naturally arises which of the two standardized measures, β_σ or β_μ , can provide the more relevant assessment of selection strength in comparative studies of natural selection. Addressing this question, below we appraise three limitations and three advantages of using the mean-standardized selection gradient β_μ .

Three limitations of the mean-standardized selection gradient were pointed out in recent reviews and meta-analyses on selection strength (Kingsolver and Pfennig 2007, Kingsolver et al. 2011), leading the authors of those studies to express their preference for using the SD-standardized selection gradient. First, mean-standardized selection gradients should not be applied to traits whose scale lacks a natural origin. If the origin of a trait's scale is arbitrary, independent studies will make different choices, and the effect of these arbitrary choices on the phenotypic mean will thus contaminate the resultant mean-standardized selection gradients. A related point was raised by Stinchcombe (2005), who highlighted that mean-standardized selection gradients are problematic when applied to fractions, which vary between 0 and 1. While such a trait f and its complement $1-f$ necessarily experience selection of equal strength (but opposite di-

rection), the two corresponding mean-standardized selection gradients differ not only in sign, but also in magnitude (as f and $1-f$ in general have different means), so it is not self-evident which of these two magnitudes to report for the mean-standardized selection gradient. We further discuss, and try to resolve, this point in Box 2. Second, the information needed to compute mean-standardized selection gradients is not always reported in publications. After Lande and Arnold's (1983) influential work, SD-standardized values were reported in many studies, but phenotypic means were often not documented in the corresponding publications. Because of this historical bias, the scope for meta-analyses of earlier studies of selection strengths in natural populations is narrower for the mean-standardized than for the SD-standardized selection gradient. While the last statement seems incontrovertible, we suggest that this historical and practical limitation must not be taken to trump the more conceptual arguments we discuss elsewhere in this article. Third, larger mean-standardized selection gradients appear to occur in populations with smaller phenotypic coefficients of variation. Further investigations will need to test this correlation, and if corroborated as robust, determine the underlying reasons.

We now turn to describing three key advantages of using mean-standardized selection gradients for comparing selection strengths in natural populations. First, these measures are easy to interpret in terms of fitness elasticities. Second, they are not confounded by a population's phenotypic variability. Third, they facilitate the liaison of ecological and evolutionary studies. Although both Kingsolver and Pfennig (2007) and Stinchcombe (2005) acknowledged the first advantage, both studies did not mention the second and third. After describing the first of these advantages briefly, we discuss the second in greater detail, before explaining the third toward the end of this section.

First advantage: Interpretation as elasticities

As we described above, the mean-standardized selection gradient β_μ has a simple interpretation: it measures the change in relative fitness that results from doubling the trait value. More generally, increasing the trait value by a fraction f of the mean trait value causes relative fitness to increase by $\beta_\mu f$. Thus, β_μ measures fitness elasticities, which is why this measure has also been called the elasticity of selection (Morgan and Schoen 1997).

Hereford et al. (2004) pointed out that one of the advantages of using mean-standardized selection gradients, directly associated with their interpretation as elasticities, is the existence of a useful benchmark for relatively strong selection: when the selected trait is fitness, then $\beta_\mu = 1$. Thus, a mean-standardized selection gradient of $|\beta_\mu| > 1$ describes a fitness landscape with a relatively steep slope around the population's phenotypic mean (Hereford et al. 2004), as a change in trait value then results in a substantial change in fitness, which in turn implies relatively strong selection. It is sometimes considered surprising that the elasticity of a trait or fitness component can exceed the elasticity of fitness itself. To escape this pitfall in understanding, it is worth recalling that there is no biological reason that would generally limit the steepness of a fitness landscape's slope β , and in particular, would prevent this slope from exceeding the inverse $1/\mu_p$ of a population's phenotypic mean. For example, $\beta_\mu = 1.2$ means that shifting trait values up by 100% increases relative fitness by 120%, which just shows that fitness is very sensitive to a change in that particular trait. Such high sensitivities are expected, for instance, in traits that rather abruptly enable survival or reproduction when exceeding a certain threshold.

Second advantage: Independence of a population's phenotypic variability

To examine the second advantage, we carried out an illustrative model-based analysis testing the three standardized measures of selection strength (for a description of the underlying mode, see Box 1). We varied the variability of a phenotypic trait that was subject to selection of fixed strength (Fig. 1d). The results demonstrate that the mean-standardized selection gradient β_μ remains almost constant under such variation, as long as the phenotypic standard deviation σ_p is not too large (Fig. 2c). In contrast, the SD-standardized selection gradient β_σ and the proportional selection differential S/μ_p are not invariant, instead appearing to increase linearly and quadratically with σ_p , respectively (Fig. 2b and Fig. 2a).

These features are readily understood by examining the relationship between the selection gradient β and the fitness landscape $\lambda(x)$ of a phenotypic trait x . Assuming a normal distribution of trait values around the phenotypic mean μ_p with variance σ_p^2 , and expanding $\lambda(x)$ around μ_p , Phillip and Arnold (1989) showed that

$$\beta = S / \sigma_p^2 = \left(\frac{d\lambda}{dx} + \frac{1}{2} \frac{d^3\lambda}{dx^3} \sigma_p^2 + \frac{1}{8} \frac{d^5\lambda}{dx^5} \sigma_p^4 + \dots \right)_{x=\mu_p} \approx \left(\frac{d\lambda}{dx} \right)_{x=\mu_p}. \quad (8)$$

Since the higher-order terms are expected to be negligible when σ_p is sufficiently small, the selection gradient β approximately equals the slope $d\lambda/dx$ of the fitness landscape $\lambda(x)$ at the phenotypic mean μ_p . From this, we immediately see that the mean-standardized selection gradient $\beta_\mu = \mu_p \beta \approx \mu_p (d\lambda/dx)_{x=\mu_p}$ is not only dimensionless, but also almost constant under variation of σ_p (as long as σ_p is small), approximating the fitness landscape's slope around the trait's mean in units of the latter. By contrast, the SD-standardized selection gradient $\beta_\sigma = \sigma_p \beta \approx \sigma_p (d\lambda/dx)_{x=\mu_p}$ and the selection

intensity, or proportional selection differential, $S / \mu_p = \mu_p^{-1} \sigma_p^2 \beta / \approx \mu_p^{-1} \sigma_p^2 (d\lambda / dx)_{x=\mu_p}$ are shown to increase approximately linearly and quadratically with σ_p , respectively.

These results highlight a major disadvantage of the SD-standardized selection gradient β_σ as a measure of selection strength. Even when selection pressures and phenotypic means are identical, populations with different degrees of phenotypic variability will exhibit different values of β_σ on the same trait of interest. In other words, the SD-standardized selection gradient varies with a population's ecological state, since a population's phenotypic variance is readily affected, for example, by whether it experiences a wide or narrow range of environmental conditions. Consequently, β_σ is not a pure measure of selection strength, but instead confounds this with a population's phenotypic variability in the considered trait. The benefit of standardization by the phenotypic standard deviation in the context of truncation selection, i.e., its independence of the phenotypic variability of the trait, is thus lost in the context of natural selection. This is because artificial truncation selection often restricts breeding to a specific fraction of individuals, whereas this is almost never the case under natural conditions. By contrast, the mean-standardized selection gradient β_μ is largely independent of a trait's phenotypic variability. Hence, mean-standardized selection gradients are fundamentally superior for comparing strengths of natural selection among traits, populations, and species across different ecological environments.

Some may dismiss the utility of the mean-standardized selection gradient because it varies with the phenotypic mean of a trait instead of with its standard deviation even when the slope of a fitness landscape is kept constant. In our assessment, however, this feature just further underscores the utility of the mean-standardized selection gradient, because it reflects a real biological difference. To see this, imagine a population of

zebrafish (*Danio rerio*; with a mean length of 5 cm) and a population of white sturgeon (*Acipenser transmontanus*; with a mean length of 500 cm). When selection on length is described in these two populations by fitness landscapes that have the same slopes of $\beta = 0.01 \text{ cm}^{-1}$ around the corresponding phenotypic means, the relative fitness of a 6 cm long zebrafish (which is 20% larger than their population's phenotypic mean) is 1.01 ($= 1 + (6 - 5) \times 0.01$), whereas that of 600 cm sturgeon (which is 20% larger than their population's phenotypic mean) is 2 ($= 1 + (600 - 500) \times 0.01$). It would presumably not occur to a field biologist to claim that length selection in these two populations were of equal strength. This intuitive understanding is well reflected when comparing β_μ (which yields $5 \times 0.01 = 0.05$ for zebrafish and $500 \times 0.01 = 5$ for white sturgeon), but not when comparing β directly (which is identical for zebrafish and sturgeon). In other words, the same slope of the fitness landscape will commonly be interpreted as representing different selection strengths depending on the considered trait's phenotypic mean. Thus, the mean-standardized selection gradient, i.e., the slope of a fitness landscape at a population's phenotypic mean measured in units of the latter, often serves as a biologically more appropriate measure of selection strength than the slope itself. In conjunction with the coefficient of genetic variation divided by the phenotypic mean, i.e. evolvability σ_A / μ_P , it is also a good predictor of the proportional selection response (Equation 3).

Third advantage: Interfacing ecology and evolution

We conclude this section by highlighting a third fundamental advantage associated with the use of the mean-standardized selection gradient: it naturally interfaces ecological and evolutionary dynamics through its decomposition into constituent components. Fig-

ure 3a is a simplified elasticity path diagram (Van Tienderen 2000) based on our illustrative example (Box 1). In this diagram, arrows indicate how a proportional change of one variable results in a proportional change of another variable. These elasticity values can be obtained either as coefficients of the corresponding linear regressions (in data-based empirical analyses) or as derivatives (in model-based theoretical analyses). The left-hand side of the diagram represents a classical selection analysis in the spirit of Lande and colleagues (e.g., Lande 1979, Lande and Arnold 1983; Box 4), as described above. The right-hand side represents a demographic analysis in which elasticities characterize the effect of a proportional change in fitness components or vital rates (e.g., survival and reproduction) on a population's growth rate. The latter analyses have, for example, been found useful in conservation biology, to identify critical aspects for increasing the growth rate of endangered populations (e.g., Benton and Grant 1999, Caswell 2001). By marrying these normally disparate analyses – the former being typically more oriented towards evolution, the latter more towards ecology – mean-standardized selection gradients assume an important integrative function.

This integrative function extends to a comparative dimension. In subpopulation A of our illustrative example (Box 1, Fig. 3), the elasticity with respect to the annual survival probability of juvenile fish (0.75) is greater than those with respect to the other fitness components (0.25), indicating that the former is the most influential determinant of this subpopulation's growth. This is in contrast to the analogous situation in subpopulation B. Likewise, we see that the effect of the annual juvenile growth increment on newborn and juvenile survival is slightly weaker in subpopulation A (-0.71) than in subpopulation B (-0.83). Joining both sides of the diagram, we can calculate the elasticities of the population growth rate with respect to the annual juvenile growth increment (0.01) and the date of river entry (0.23) for subpopulation A, which tells us how an in-

crease in these phenotypic trait values results in an increase in the population growth rate. The corresponding values for subpopulation B (0.61 and 0.46, respectively) are very different, and considerably larger. Importantly, these elasticities of fitness with respect to the phenotypic traits are nothing else but mean-standardized selection gradients, because the population growth rate can be interpreted as fitness. In other words, if a small change in a trait could enhance population growth rate greatly, it also implies that selection acting on that trait is strong. Thus, elasticities, i.e., mean-standardized values, work as versatile common “currencies” interfacing ecological and evolutionary dynamics, in a way that is difficult to achieve through other techniques. In fact, the elasticity path diagram has played an important role as a gateway for researchers to become familiar with the mean-standardized selection gradient (Box 3). This perspective has great potential to facilitate comparisons across traits, populations, and species (Box 1, Fig. 3) and to integrate analyses of ecological and evolutionary dynamics (Lande 1982, van Tienderen 2000) in natural populations.

Conclusions

As the selection differential S and the selection gradient β are not dimensionless, they must be standardized before executing any kind of comparison of selection strengths across traits, populations, or species. We believe that applied and basic evolutionary biologists will benefit from routinely using the mean-standardized selection gradient in future studies. The reason for recommending this particular standardization is fundamental: it is the most direct measure of the selective implications of particular ecological settings in nature, and as such is not confounded by a selected trait’s phenotypic variability in the population under consideration.

Figure 4 summarizes the systematic standardization process that translates what can be measured in natural populations to the dimensionless measure most suitable for comparing selection strengths among traits, populations, and species. We start our overview from what is probably the most directly observable quantity informing us about the strength of selection in natural populations: this is a trait's evolutionary rate, i.e., the change in a phenotypic trait value divided by the time it took for the change to occur, in light of constant ecological conditions. Such evolutionary rates have been used to compare the pace of evolutionary changes in response to particular (often anthropogenic) selection pressures affecting specific populations and specific traits (e.g., Jørgensen et al. 2007, Darimont et al. 2009). Despite their accessibility to measurement, these rates as such, however, cannot serve as suitable measures of selection strength. While this conclusion is formally evident from the fact that such rates are not dimensionless, we can also understand it by appreciating that evolutionary rates confound the strength of selection with properties such as generation time, heritability, phenotypic variance, and phenotypic mean (Fig. 4), which all idiosyncratically depend on the populations and traits under consideration. It is thus easy to see that the effects of these population- and trait-specific confounding factors must be removed before comparisons of selection strengths between traits, populations, and species can become meaningful. As shown in Fig. 4, it is the mean-standardized selection gradient β_μ that remains once the aforementioned confounding factors have been removed. In the figure, we show the steps from the evolutionary rate, the most directly observable quantity, to the dimensionless quantity of the mean-standardized selection gradient β_μ , in what to us seems the most easily understandable order. We emphasize that the order of standardization is arbitrary, which implies that quantities other than those shown in the figure can be used as stepping stones along the way. Figure 4 also reminds us of the importance of standardizing selection

strengths by generation time: if a selection gradient is calculated based on a population's intrinsic growth rate (Lande 1982, van Tienderen 2000), the result must still be standardized by the population's generation time before using it for comparing selection strengths across populations in a manner that is independent of their generation times. Box 4 outlines the most important issues to consider in regression-based analyses of empirical data for determining mean-standardized selection gradients.

Unfortunately, for historical reasons, the number of studies reporting the mean-standardized selection gradient is still relatively small (Box 3). Based on this observation, we would like to repeat Hereford et al.'s (2004) appeal: authors, reviewers, and journal editors should insist on the publication of basic, yet fundamental values in studies on selection strengths, including phenotypic means and variances. If phenotypic means and variances are available, any measure of selection strength, reported as a selection differential S or a selection gradient β , can easily be converted into the mean-standardized selection gradient (Fig. 4, equation 6). It is only under this condition that reported selection strengths become comparable across traits, populations, and species, and hence are amenable to future meta-analyses.

To conclude, we agree with Hereford et al.'s (2004) caveat that no standardized measure of selection strength will be satisfactory for all research purposes. The reason for this is conceptually profound, and is intimately related to the notion of a trait's natural scale and measurement theory (Houle et al. 2011; Box 2). At any rate, jointly reporting unstandardized, SD-standardized, and mean-standardized selection gradients in all salient scientific publications will allow readers to conduct their own analyses and draw their own conclusions (Stinchcombe 2005), and thus is the recommended way forward.

Acknowledgements

UD gratefully acknowledges financial support by the European Commission, as part of the Marie Curie Research Training Network on Fisheries-induced Adaptive Changes in Exploited Stocks (FishACE, MRTN-CT-2004-005578) and the Specific Targeted Research Project on Fisheries-induced Evolution (FinE, SSP-2006-044276) under the European Community's Sixth Framework Program. UD received additional support by the European Science Foundation, the Austrian Science Fund, the Austrian Ministry of Science and Research, and the Vienna Science and Technology Fund. RA acknowledges funding by the Gottfried-Wilhelm-Leibniz-Association for the project Adaptfish (www.adaptfish.igb-berlin.de). The finalization of this article was funded by the Federal Ministry for Education and Research (BMBF) within the Program on Social-Ecological Research and the project Besatzfisch (#01UU0907, www.besatz-fisch.de), and by a KAKENHI Grant-in-Aid for Scientific Research (C, 22570087, 2010-2012) to SM.

References

- Arnold SJ, Wade MJ. 1984. On the measurement of natural and sexual selection: theory. *Evolution* 38 709-719.
- Arlinghaus R, Matsumura S, Dieckmann U. 2009. Quantifying selection differentials caused by recreational fishing: development of modeling framework and application to reproductive investment in pike (*Esox lucius*). *Evolutionary Applications* 2: 335-355.
- Benton TG, Grant A. 1999. Elasticity analysis as an important tool in evolutionary and population ecology. *Trends in Ecology and Evolution* 14: 467-471.
- Bolstad GH, Armbruster WS, Pelabon C, Perez-Barrales R, Hansen TF. 2010. Direct selection at the blossom level on floral reward by pollinators in a natural popula-

- tion of *Dalechampia schottii*: full-disclosure honesty? *New Phytologist* 188: 370-384.
- Brodie ED III, Moore AJ, Janzen FJ. 1995. Visualizing and quantifying natural selection. *Trends in Ecology and Evolution* 10: 313-318.
- Bromquist GE, Sade DS, Berard JD. 2011. Rank-related fitness differences and their demographic pathways in semi-free-ranging rhesus macaques (*Macaca mulatta*). *International Journal of Primatology* 32: 193-208.
- Bulmer MG. 1980. *The mathematical theory of quantitative genetics*. Clarendon Press.
- Carlson et al. 2007. Four decades of opposing natural and human-induced artificial selection acting on Windermere pike (*Esox lucius*). *Ecology Letters* 10: 512-521.
- Caswell H. 2001. *Matrix population models* (2nd ed.). Sinauer Associates.
- Caswell H, Naiman RJ, Morin R. 1978. Evaluating the consequences of reproduction in complex salmonid life cycles. *Aquaculture* 43: 123-143.
- Coulson T, Kruuk LEB, Tavecchia G, Pemberton JM, Clutton-Brock TH. 2003. Estimating selection on neonatal traits in red deer using elasticity path analysis. *Evolution* 57: 2879-2892.
- Crow JF, Kimura M. 1979. Efficiency of truncation selection. *Proceedings of the National Academy of Sciences of the USA* 76: 396-399.
- de Kroon H, Plaisier A, van Groenendael J. 1986. Elasticity: the relative contribution of demographic parameters to population growth rate. *Ecology* 67: 1427-1431.
- de Kroon H, van Groenendael J, Ehrlén J. 2000. Elasticities: a review of methods and model limitations. *Ecology* 81: 607-618.

- Darimont CT, Carlson SM, Kinnison MT, Paquet PC, Reimchen TE, Wilmers CC. 2009. Human predators outpace other agents of trait change in the wild. *Proceedings of the National Academy of Sciences of the USA* 106: 952-954.
- Darwin C. 1859. *On the Origin of Species*. John Murray.
- Dibattista JD, Feldheim KA, Gruber SH, Hendry AP. 2007. When bigger is not better: selection against large size, high condition and fast growth in juvenile lemon sharks. *Journal of Evolutionary Biology* 20: 201-212.
- Ehrlen J, Munzbergova Z. 2009. Timing of flowering: opposed selection on different fitness components and trait covariation. *American Naturalist* 173: 819-830.
- Falconer DS. 1960. *Introduction to quantitative genetics* (1st ed.). Oliver and Boyd.
- Falconer DS, Mackay TFC. 1996. *Introduction to quantitative genetics* (4th ed.). Longman Limited.
- Farrell AP, Hinch SG, Cooke SJ, Patterson DA, Crossin GT, Lapointe M, Mathes MT. 2008. Pacific salmon in hot water: applying aerobic scope models and biotelemetry to predict the success of spawning migration. *Physiological and Biochemical Zoology* 81: 697-708.
- Friars GW, Smith PJ. 2010. Heritability, correlation and selection response estimates of some traits in fish populations. Atlantic Salmon Federation Technical Report, March 2010 (<http://asf.ca/docs/uploads/friars-smith.pdf>).
- Hansen TF, Houle D. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* 21: 1201-1219.
- Hansen TF, Pélabon C, Armbruster WS, Carlson ML. 2003. Evolvability and genetic constraint in *Dalechampia* blossoms: components of variance and measures of evolvability. *Journal of Evolutionary Biology* 16: 754-766.

- Hard JJ, Gross MR, Heino M, Hilborn R, Kope RG, Law R, Reynolds JD. 2008. Evolutionary consequences of fishing and their applications for salmon. *Evolutionary Applications* 1: 388-408.
- Houle D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* 130: 195-204.
- Houle D, Pélabon C, Wagner GP, Hansen TF. 2011. Measurement and meaning in biology. *Quarterly Review of Biology* 86: 3-34.
- Hendry AP. 2005. The power of natural selection. *Nature* 433: 695-696.
- Hereford J, Hansen TF, Houle D. 2004. Comparing strengths of directional selection: How strong is strong? *Evolution* 58: 1637-1658.
- 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. *Proceedings of the National Academy of Sciences of the USA* 98: 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. *International Journal of Plant Sciences* 171:945-959.
- Hutchings JA, Fraser DJ. 2007. The nature of fisheries- and farming-induced evolution. *Molecular Ecology* 17: 295-313.
- Janzen FJ, Stern, HS. 1998. Logistic regression for empirical studies of multivariate selection. *Evolution* 52: 1564-1571.
- Jørgensen C et al. 2007. Managing evolving fish stocks. *Science* 318: 1247-1248.
- Johnson HW, Robinson HF, Comstock RE. 1955. Estimates of genetic and environmental variability in soybeans. *Agronomy Journal* 47: 314-318.

- Kelley JL, Stinchcombe JR, Weinig C, Schmitt J. 2005. Soft and hard selection on plant defence traits in *Arabidopsis thaliana*. *Evolutionary Ecology Research* 7: 287-302.
- Kingsolver JG, Diamond SE. 2011. Phenotypic selection in natural populations: what limits directional selection? *American Naturalist* 177: 346-357.
- 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 population. *American Naturalist* 157: 245-261.
- Kinnison MT, Hendry AP. 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica* 112-113: 145-164.
- Kipling C. 1983. Changes in the growth of pike (*Esox lucius*) in Windermere. *Journal of Animal Ecology* 52: 647-657.
- Kuparinen A, Merilä J. 2007. Detecting and managing fisheries-induced evolution. *Trends in Ecology and Evolution* 22: 652-659.
- Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33: 402-416.
- Lande R. 1982. A quantitative genetic theory of life history evolution. *Ecology* 63: 607-615.
- Lande R, Arnold SJ. 1983. The measurement of selection on correlated characters. *Evolution* 37: 1210-1226.
- Law R. 2007. Fisheries-induced evolution: present status and future directions. *Marine Ecology Progress Series* 335: 271-277.

- Matsumura S, Arlinghaus R, Dieckmann U. 2011. Assessing evolutionary consequences of size-selective recreational fishing on multiple life-history traits, with an application to northern pike (*Esox lucius*). *Evolutionary Ecology* 25:711-735.
- Metz JAJ, Nisbet RM, Geritz SAH. 1992. How should we define “fitness” for general ecological scenarios? *Trends in Ecology and Evolution* 7: 198-202.
- Mojica JP, Kelly JK. 2010. Viability selection prior to trait expression is an essential component of natural selection. *Proceedings of the Royal Society B* 277: 2945-2950.
- Morgan MT, Schoen DJ. 1997. Selection on reproductive characters: Floral morphology in *Asclepias syriaca*. *Heredity* 79: 433-441.
- Palumbi SR. 2001. Humans as the world’s greatest evolutionary force. *Science* 293: 1786-1790.
- Raat AJP. 1988. Synopsis of biological data on the northern pike (*Esox lucius* Linnaeus, 1758). *FAO Fisheries Synopsis* 30, Review 2. FAO.
- Reinhold K. 2011. Variation in acoustic signalling traits exhibits footprints of sexual selection. *Evolution* 65: 738-745.
- Roff DA. 1997. *Evolutionary Quantitative Genetics*. Chapman & Hall.
- Roff DA. 2002. *Life History Evolution*. Sinauer Associates.
- Seamons TR, Bentzen P, Quinn TP. 2007. DNA parentage analysis reveals inter-annual variation in selection: results from 19 consecutive brood years in steelhead trout. *Evolutionary Ecology Research* 9: 409-431.
- Simonsen AK, Stinchcombe JR. 2010. Qualifying evolutionary genetic constraints in the ivyleaf morning glory, *Ipomoea hederacea*. *International Journal of Plant Sciences* 171: 972-986.

Stinchcombe JR. 2005. Measuring natural selection on proportional traits: Comparisons of three types of selection estimates for resistance and susceptibility to herbivore damage. *Evolutionary Ecology* 19: 363-373.

Stinchcombe JR, Izem R, Heschel MS, McGoey BV, Schmitt J. 2010. Across-environment genetic correlations and the frequency of selective environments shape the evolutionary dynamics of growth rate in *Impatiens capensis*. *Evolution* 64: 2887-2903.

van Tienderen PH. 2000. Elasticities and the link between demographic and evolutionary dynamics. *Ecology* 81: 666-679.

Figure captions

Fig. 1. Schematic illustration of how a selection differential S is defined as the shift in a trait's phenotypic mean before and after selection. The two columns show this for truncation selection (left) and natural selection (right). The two rows show how S is altered by different degrees of phenotypic variability, here illustrated by two normal distributions, one of which has twice the standard deviation (doubled SD, $2\sigma_p$, bottom) of the other (baseline SD, σ_p , top). (a,b) Under truncation selection, S is the difference between the phenotypic mean μ_p of all individuals in a population (thin line) and the phenotypic mean of those individuals that are selected as parents for reproduction (thick line). When the same proportion of individuals is selected in the baseline population (a) and in the population with double SD (b), the selection differential $S_b = 2S_a$ is twice as large as S_a . This makes it natural to consider the SD-standardized selection gradient $\beta_\sigma = S / \sigma_p$, which conveniently measures the proportion of individuals of the whole population that are selected as parents for reproduction. This gives $\beta_{\sigma,b} = S_b / (2\sigma_p) = (2S_a) / (2\sigma_p) = S_a / \sigma_p = \beta_{\sigma,a}$, reflecting that truncation selection was applied in the same way (described by same truncation proportion) in both populations. (c,d) Under natural selection, S is the difference between the mean of a population's phenotypic distribution (thin continuous line) and of this distribution weighted with the differential fitness of phenotypes (thin dashed line, resulting in the product shown by the thick line). When the same fitness function acts on the baseline population (c) and the population with double SD (d), the selection differential $S_d = 4S_c$ is (approximately) four times as large as S_c . This makes it natural to consider the mean-standardized selection gradient $\beta_\mu = \mu_p S / \sigma_p^2$, which conveniently measures the elasticity of fitness. This gives $\beta_{\mu,d} =$

$\mu_p S_d / (2\sigma_p)^2 = \mu_p (4S_c) / (4\sigma_p^2) = \mu_p S_c / \sigma_p^2 = \beta_{\mu,c}$, reflecting that natural selection applied in the same way (described by the same fitness function) in both populations.

Fig. 2. Dependence of three alternative standardized measures of selection strength on a population's phenotypic variability. The panels (a), (b), and (c) show absolute values of S / μ_p , $\beta_\sigma = S / \sigma_p = \sigma_p \beta$, and $\beta_\mu = \mu_p S / \sigma_p^2 = \mu_p \beta$, respectively. Three different levels of river temperature are considered (low, middle, and high), resulting in selection of increasing strength (see Box 1 for model details). Parameters: $a_m = 2$ yr, $d_{opt} = 50$ d, $g = 0.5$, $h = 25$ cm/yr, $h_{max} = 55$ cm/yr, $T = 1^\circ\text{C}$, $T_u = 1^\circ\text{C}$, $L_u = 1$ cm, $b_1 = 0.1^\circ\text{C/d}$, $b_2 = 0.9$, $b_3 = 0.1$, $b_4 = 0.05$, $b_5 = 3$, $b_6 = 0.00032$, and $b_7 = 0.8$.

Fig. 3. Elasticity path diagram and its comparative application. (a) Elasticity path diagram for the pairwise relationships between two phenotypic traits, four fitness components, and the population growth rate in subpopulations A and B of the illustrative example described in Box 1. Arrows indicate how a proportional change of one variable results in the proportional change of another variable. Elasticities are shown in boldface and italics for subpopulations A and B, respectively. For example, a 100% increase in the annual juvenile growth increment h results in a 287% increase in fecundity (2.87). The fitness elasticity of a phenotypic trait is then calculated by following all left-to-right paths from the trait to the population growth rate, multiplying elasticities along the arrows and summing over all paths: in subpopulation A, h has a fitness elasticity of $(-0.71) \times 0.25 + (-0.71) \times 0.75 + 2.87 \times 0.25 = 0.01$, whereas the date d of river entry has a fitness elasticity of $0.91 \times 0.25 = 0.23$. For further details on elasticity path diagrams, see van Tienderen (2000). (b) Comparison of mean-standardized selection gradi-

ents between subpopulations A and B. Parameters: $a_m = 4$ yr (subpopulation A) or 2 yr (B), $h_{\max} = 60$ cm/yr (A) or 55 cm/yr (B), all other parameters as in Fig. 2.

Fig. 4. Successively stripping away four population- and trait-specific properties (generation time T , heritability h^2 , phenotypic variance σ_p^2 , and phenotypic mean μ_p) translates evolutionary rates r into mean-standardized selection gradients β_μ . This standardization process can be likened to the peeling of an onion, with the directly observable quantity as the outer skin and the dimensionless quantity most useful for comparisons across traits, populations, and species, at the core.

Box 1: An illustrative example for estimating mean-standardized selection gradients using model-based analyses of wild populations

To illustrate the usefulness of the mean-standardized selection gradient, here we show how to model selection differentials in a hypothetical semelparous anadromous fish species, such as a salmonid. We hope this example will help theoretical ecologists wishing to apply the mean-standardized selection gradient to their own models.

Model description

We consider a simple fish life history with individuals that are born in a river, migrate to the ocean at age 1, grow in the ocean, return to the river at age a_m , reproduce, and die after reproduction. We examine the date of river entry (d) and the annual juvenile growth increment (h) as evolving traits.

The success of spawning migration in salmonids has been reported to depend on river temperature, with the optimal temperature differing between populations (e.g. Farrell et al. 2008). Any increase in river temperature – for example, due to global warming – is thus expected to cause selection pressures on the date of river entry. Here we simply assume that river temperature decreases linearly from late summer to autumn. For fish entering the river at day d , the deviation from the optimal temperature is given by $T_{\text{dev}}(d) = -b_1(d - d_{\text{opt}}) + T$, where d_{opt} is the optimal date of river entry without climate change, $b_1 > 0$ is the rate of seasonal temperature change, and T measures the increase of river temperature due to climate change. The survival probability during the spawning migration from the river mouth to the spawning site is

$s_{\text{mig}} = b_2 \exp(-b_3(T_{\text{dev}}(d)/T_u)^2)$, where b_2 is the survival probability at the optimal tem-

perature, $b_3 > 0$ determines the temperature sensitivity of migration survival, and T_u is a unit-standardizing constant.

The growth trajectory of individuals is described using the biphasic somatic growth model by Lester et al. (2004): the length at age a_m is represented as

$$L_{a_m} = \frac{3}{g+3}(L_{a_m-1} + h) = \frac{3a_m h}{g+3},$$

where h is the annual juvenile growth increment and g

is the annual reproductive investment. The fecundity f_{a_m} of fish at age a_m that have successfully arrived at spawning sites is allometrically given by $f_{a_m} = b_4 g (L_{a_m} / L_u)^{b_5}$, where b_4 and b_5 are the allometric parameters, and L_u is a unit-standardizing constant. The basic annual natural mortality is given by $m = h / h_{\max}$, where h_{\max} is the maximum growth capacity beyond which survival is zero (Matsumura et al. 2011). The annual survival probability of hatched larvae (newborn survival, s_0) and fish in the ocean (juvenile survival, s_j) is then represented as $s_0 = b_6(1-m)$ and $s_j = b_7(1-m)$, respectively, where b_6 and b_7 are parameters.

We assume a polymorphic resident population with trait values x being normally distributed around the population mean μ_p with phenotypic variance σ_p^2 . The dynamics of fish with trait values x in this population is described by a population projection matrix

$$\mathbf{A}(x) = \begin{pmatrix} 0 & 0 & \cdots & 0 & s_0(x)f_{a_m}(x)s_{\text{mig}}(x) \\ s_j(x) & 0 & \cdots & 0 & 0 \\ 0 & s_j(x) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & s_j(x) & 0 \end{pmatrix}.$$

Changes in the age structure and density of fish with trait values x are thus described by $\mathbf{N}'(x) = \mathbf{A}(x)\mathbf{N}(x)$, where the vectors $\mathbf{N}(x) = (N_1(x), N_2(x), \dots, N_{a_m}(x))^T$ and

$\mathbf{N}'(x) = (N'_1(x), N'_2(x), \dots, N'_{a_m}(x))^T$ contain the densities of fish across all age classes in years t and $t+1$, respectively.

Calculating selection differentials

The dominant eigenvalue $\lambda(x)$ of the population projection matrix of trait values x is known to represent the annual population growth rate, which is commonly used to measure trait-specific lifetime fitness (e.g., Metz et al. 1992, Arlinghaus et al. 2009).

The selection differential per year is then calculated as

$S = \int x\lambda(x)p(x)dx / \int \lambda(x)p(x)dx - \mu_p$ (Fig. 1), where $\lambda(x)$ and $p(x)$ are the fitness and probability density of trait values x , respectively. The first term in this difference corresponds to the average phenotypic value of the population after selection, i.e., the average phenotypic value of parents of the next generation (Arlinghaus et al. 2009).

In the present example, we consider the effects of an increase T in river temperature (e.g., due to global warming), so that the fitness of fish that enter the river later than at their traditional optimal date is expected to be raised. In other words, the proportion of late-comers will be larger in the spawning population than in the original population, making the first term in the above difference larger than the second, which results in a positive selection differential.

Influence of trait variance on the three standardized measures of selection strength

Using this example, we first examine how three standardized measures of selection strength change for different levels of the population's phenotypic variability. We focus on the date d of river entry and calculate the selection differential S for d caused by the considered increase in river temperature. We vary the phenotypic standard deviation

σ_p and examine the resultant changes of the three standardized measures under fixed strengths of selection (Fig. 2). It is easily seen that only the mean-standardized selection gradient is approximately independent of phenotypic variability.

Interfacing evolution and ecology through elasticity path diagrams

Next, we illustrate the potential of the elasticity path diagram (van Tienderen 2000; Fig. 3) to interface evolution and ecology and to facilitate comparison across traits and subpopulations. We consider two hypothetical endangered fish subpopulations A and B that live in different habitats and differ in their life-history patterns. Considerable heritable variation in the age of maturation has been reported in salmonids (e.g., Friars and Smith 2010), so we assume that the ages of returning to the river and reproducing in subpopulation A and B are $a_m = 4$ yr and $a_m = 2$ yr, respectively. We also assume that the basic annual natural mortality increases less rapidly with the growth increment in subpopulation A ($h_{\max} = 60$ cm) than in subpopulation B ($h_{\max} = 55$ cm).

In subpopulation A, the fitness elasticity of juvenile survival (0.75) is larger than those of any of the other vital rates (0.25). This has an implication for population managers: improving juvenile survival would be a good strategy to increase the population growth rate of subpopulation A. At the same time, it is predicted that natural selection would favor traits improving juvenile survival. Although smaller growth increments lead to higher juvenile survival in our model, it also reduces fecundity. Consequently, selection on growth in subpopulation A is weaker than that on the date of river entry. By contrast, in subpopulation B, the fitness elasticity of juvenile survival is not larger than those of the other vital rates, and the selection acting on growth is stronger than that on the date of river entry. Thus, by providing ecological and evolutionary measures as mean-standardized values and considering them together, the elasticity path diagram

facilitates comparison across traits, subpopulations, and species in natural populations. Also the extrapolation of estimates from a well-known subpopulation or species to others is enabled by using such mean-standardized values, although this must always be done with care, because all empirically estimated elasticities are dependent on the considered ecological environments.

Box 2: Natural scales and their implications

In many contexts, a trait's natural scale would be given by its phenotypic standard deviation. In the context of natural selection, however, this standard deviation is disqualified from serving in this role, since a trait's phenotypic variance is of such central dynamical importance for scaling a trait's selection differential and response (Fig. 2). Since the phenotypic variance is needed for removing this confounding effect of phenotypic variability from selection differentials, it is unsuitable for rendering selection gradients dimensionless. In most circumstances, the most natural scale that remains is given by a trait's phenotypic mean. This is ultimately because the majority of traits have a natural minimum, typically designated by a trait value of 0. The difference between this minimum and a trait's current phenotypic mean establishes a natural scale suitable for standardization.

Against this background, we can now systematically appreciate two exceptions mentioned in the main text. First, when a quantitative trait possesses both a natural minimum and a natural maximum, the two natural scales given by the difference between the trait's current phenotypic mean and its minimum or maximum, respectively, can alternatively be employed for standardization. For such traits, however, it will often be more natural simply to use the difference between the trait's minimum and maximum as the natural scale for standardization, which overcomes the ambiguity that would arise otherwise. For example, Stinchcombe (2005) considered two phenotypic traits of a plant species, resistance and susceptibility to herbivore damage (which both vary between 0 and 1 and are direct complements to each other, by always adding up to 1), and pointed out the problem of obtaining different values for the mean-standardized selection gradient of these traits. Here we argue that this problem can be avoided altogether by using the difference between the traits' minimum (0) and maximum (1) as the natural scale for

standardization (in this example the difference is 1). Second, when a quantitative trait has no natural minimum or maximum, or when the values of these quantities are highly uncertain, the implied natural scales do not exist or are empirically unreliable, and as such should not be used for standardization.

More fundamentally, whether a variable can be standardized or not is related to measurement theory: Houle et al. (2011) highlight that permissible transformations differ depending on a variable's scale types such as nominal, ordinal, interval, or ratio. Another important issue pointed out by Houle et al. (2011) is the importance of concepts and hypotheses for measurement: different contexts assign different meanings (and sometimes different scale types) to the same measurement.

Box 3: Is the mean-standardized selection gradient increasingly used?

As described in the main text, the mean-standardized selection gradient has some fundamental advantages as a measure of the strength of natural selection, relative to the SD-standardized selection gradient: in particular, it does not confound selection strength with phenotypic variability. It is therefore interesting to evaluate the uptake and current prevalence of this measure in the relevant scientific literature. To this end, we carried out a literature survey based on the ISI Web of Science by Thomson Reuters, as follows. First, we chose five primary journal articles that discussed the definition of mean-standardized selection gradients and their calculation (Morgan and Schoen 1997, van Tienderen 2000, Hansen et al. 2003, Hereford et al. 2004, and Stinchcombe 2005). Second, we assembled a list of secondary journal articles included in the ISI Web of Science by the end of March 2011 that cited the primary articles. Thereby, we identified 179 papers (excluding the five primary articles), which in a third step we finally assessed for their usage of different methods for standardizing selection strength.

Surprisingly, among these 179 studies there were only eleven that reported empirical estimates of the mean-standardized selection gradient (Table B1). Three out of these papers calculated the mean-standardized selection gradient in the context of elasticity path diagrams (van Tienderen 2000). This suggests that the elasticity analysis of Leslie-matrix models has played an important role as a gateway for a wider range of biologists to become familiar with the mean-standardized selection gradient. We also found that, except for Arlinghaus et al. (2009), not a single model-based analysis published before the end of March 2011 for estimating selection strengths across a range of scenarios has used the mean-standardized selection gradient. Most importantly, the results of our survey show that the standardization of selection gradients by a population's phenotypic mean is as yet far away from becoming common practice, despite this meth-

od's superior suitability for comparing selection strengths in natural populations. However, our results also indicate that there is a small core group of authors increasingly using the mean-standardized selection gradient during the last few years (Table B1). Moreover, some researchers (e.g., Mojica and Kelly 2010) have started to report both the SD-standardized and the mean-standardized selection gradients, as recommended by Hereford et al. (2004) to enable various types of comparisons in future meta-analyses.

Why is the use of the mean-standardized selection gradient still infrequent in the contemporary literature? A reason might be gleaned from a particular study by Seamons et al. (2007) in which the mean-standardized selection gradient was calculated, but not reported, because the authors thought the values they had obtained were unrealistic. In their own words: “[Mean-standardized selection gradients] calculated for [spawner body] length were unrealistic (much larger than unity), so the results of this analysis are not shown”. This finding might be related to the third point mentioned by Kingsolver and Pfennig's (2007), that large mean-standardized selection gradients are observed in association with small phenotypic coefficients of variation. While a statistical explanation has been suggested for explaining this pattern (Kingsolver and Pfennig 2007), further investigation is certainly required. It is also possible, however, that scientists are liable to misinterpret the comparison between mean-standardized selection gradients for fitness (which equals 1 by definition) and other traits, as nothing at all is a priori wrong if the latter exceed 1.

Through our literature survey, we also encountered an almost “Babylonian confusion” of terminology. The fact that the two common measures of standardized selection strength have been used by different authors in different contexts has increased the number of terms in circulation to a level that we consider problematic (Table B2). The term *intensity of selection* (or simply *selection intensity*) has long been established in

quantitative genetics for referring to the SD-standardized selection gradient, in particular in the context of truncation selection. Since this measure was also simply called *standardized selection differential* in a popular textbook (Falconer and Mackey 1996), several authors have used this rather unspecific term in their articles. In studies of selection on multivariate traits, following Lande and Arnold (1983), the term *standardized selection gradient* has been widely employed. The same measure, however, is sometimes also simply called *selection gradient* (e.g., Kingsolver et al. 2001), because standardization by the phenotypic standard deviation has been, and still is, so common. Given such defaults, the terminological situation evidently was bound to become messy when an alternative approach to standardization, by using units of the phenotypic mean instead of the phenotypic standard deviation, was introduced to the field.

In the future, the two standardized measures of selection strength need to be clearly distinguished by appropriate terminology. Recently, the pairing *variance-standardized selection gradient* and *mean-standardized selection gradient* seems to be becoming popular in review papers and meta-analyses (e.g., Hereford et al. 2004, Stinchcombe 2005, Kingsolver and Pfennig 2007). Strictly speaking, however, the selection gradient in the former case is standardized by the phenotypic standard deviation, and not by the phenotypic variance. Terms adequately reflecting this could be *selection gradient in units of the phenotypic standard deviation* and *selection gradient in units of the phenotypic mean*. The corresponding abbreviated terms could be *SD-standardized selection gradient* and *mean-standardized selection gradient*, which is the convention we have adopted throughout the present article.

Finally, we need to emphasize the importance of better distinguishing the (unstandardized) selection differential and selection gradient. In the common regression-based approach to estimating selection from empirical data, the selection differential is

the covariance between relative fitness and the trait of interest (its unit is [trait]), while the selection gradient is a regression coefficient of relative fitness on the trait, i.e., the covariance/variance ratio of that trait (its unit is [trait⁻¹]) (Lande and Arnold 1983). Yet, we noticed that the term selection differential was used for regression coefficients when univariate regressions of relative fitness on one trait have been applied, while the term selection gradient was used in multivariate regressions jointly including several adaptive traits (e.g., Kingsolver et al. 2001, Carlson et al. 2007). This confusion might have arisen because both the selection differential and a regression coefficient in a univariate regression reflect changes resulting from a direct link between fitness and the trait (direct selection), as well as from correlations with other traits affecting fitness (indirect selection). However, a regression coefficient is not identical to the selection differential defined in traditional quantitative genetics. To avoid further confusion, we propose always to stick to the term selection gradient when referring to coefficients of regressions of relative fitness on phenotypic traits.

Box 4: A roadmap to estimating mean-standardized selection gradients using phenotypic data from the wild

In studies of selection in natural populations, the mean-standardized selection gradient will most commonly be obtained from the unstandardized selection gradient β estimated from a regression-based model of relative fitness in dependence on a phenotypic trait. Here, we briefly describe the procedure for estimating the mean-standardized selection gradient from phenotypic data from natural populations. We focus on the standardization processes; the general principles of regression-based selection estimation can be found elsewhere (e.g., Brodie et al. 1995). We hope this outline will help empirical ecologists wishing to apply the mean-standardized selection gradient to their own data.

Generally, investigators will have to account for four related steps when attempting to arrive at mean-standardized selection gradients through regression models:

1. *Collection of empirical data.* Estimation of the selection gradient from empirical observations using regression models requires individual-level data on phenotypic traits and corresponding fitness. Although lifetime measures of fitness are ideal, in most empirical studies only parts of an individual's fitness, i.e., fitness components, can be empirically assessed (e.g., in a marine fish with external broadcast reproduction, it would be impossible to estimate reproductive output by screening all potential offspring using genetic methods). Typical measures of fitness components include fertility and/or fecundity (e.g., number or biomass of offspring), survival at a given stage, and mating success (Brodie et al. 1995).
2. *Regression.* The type of regression analysis depends on the structure of the chosen fitness measure. In most cases, univariate or multivariate linear regression models will be used when the fitness measure is quantitative. By contrast, logistic regression models will be suitable for binary fitness measures (such as survival; Carlson

et al. 2007). Note that in the latter case the parameter estimate for the logistic regression coefficient will need to be transformed to its linear equivalent before standardization of selection strength (Janzen and Stern 1998).

3. *Removal of non-genetic environmental effects.* If phenotypic data are available in time series, any salient environmental variables (e.g., biomasses or densities of the focal species and its potential prey, predators, or competitors; or relevant abiotic variables) can be included in the regression model to control for their effects on relative fitness experienced by an individual in dependence on its the phenotypic trait and varying environmental conditions.
4. *Standardization.* The unstandardized selection gradient is obtained as a coefficient of the regression of relative fitness on the phenotypic trait (or as a partial regression coefficient in the case of multivariate analyses). To obtain the mean-standardized selection gradient, the unstandardized selection gradient needs to be multiplied by the mean of the phenotypic trait in the population (equation 6). In some applications, the trait values are standardized to a mean of 0 and a standard deviation of 1 (e.g., through z-transformation) prior to the regression analysis (e.g., Carlson et al. 2007). In this case, the regression coefficient obtained represents the SD-standardized selection gradient. Therefore, it must be multiplied by the original mean and divided by the original SD of the phenotypic trait to obtain the mean-standardized selection gradient (equations 5 and 6). Fitness measures are often generation-based, in which case the selection gradient obtained from a regression analyses need not be standardized by generation time. Exceptions occur when fitness is measured in a time unit other than generation length (for example, annual population growth rates; Arlinghaus et al. 2009, in which case multiplication with generation length is required).

Table B1. A sample of studies reporting estimates of selection strengths using the mean-standardized selection gradient.

Reference	Type of study	Subjects
Morgan & Schoen 1997	Empirical	Plants (floral characters in <i>Asclepias syriaca</i>)
van Tienderen 2000	Illustrative	Hypothetical plants
Coulson et al. 2003	Empirical	Animals (neonatal traits in <i>Cervus elaphus</i>)
Hereford et al. 2004	Review of empirical studies	Estimates from 38 studies
Stinchcombe 2005	Empirical	Plants (a resistance trait and a size trait in <i>Ipomoea hederacea</i>)
Kelley et al. 2005	Empirical	Plants (defense traits in <i>Arabidopsis thaliana</i>)
Dibattista et al. 2007	Empirical	Animals (life-history traits including size and growth in <i>Negaprion brevirostris</i>)
Ehrlen & Munzbergova 2009	Empirical	Plants (floral characters in <i>Lathyrus vernus</i>)
Arlinghaus et al. 2009	Modeling	Animals (life-history traits in <i>Esox lucius</i>)
Mojica & Kelly 2010	Empirical	Plants (flower size in <i>Mimulus guttatus</i>)
Stinchcombe et al. 2010	Empirical	Plants (growth rate in <i>Impa-</i>

		<i>tiens capensis</i>)
Bolstad et al. 2010	Empirical	Plants (floral traits in <i>Dalechampia schottii</i>)
Simonsen & Stinchcombe 2010	Empirical	Plants (flowering time and size in <i>Ipomoea hederacea</i>)
Horvitz et al. 2010	Reanalysis of empirical data	Plants (<i>Calathea ovandensis</i>) and animals (<i>Cervus elaphus</i>)
Blomquist et al. 2011	Empirical	Animals (dominance rank in <i>Macaca mulatta</i>)
Reinhold 2011	Reanalysis of empirical data	Animals (acoustic signaling traits in insects and amphibians)

Table B2. Alternative terms used for the two major standardized measures of selection strength across a sample of scientific publications.

Measure	Term	Reference
$\beta_{\sigma} = S / \sigma_P$ $= \sigma_P \beta$	Selection intensity (intensity of selection)	Falconer 1960 and many others
	Standardized selection gradient	Lande & Arnold 1983 and many others
	Variance-standardized selection gradient	Hereford et al. 2004
	Standard-deviation-standardized selection differential	Arlinghaus et al. 2009
	Selection gradient in units of the phenotypic standard deviation	This study (recommended long form)
	SD-standardized selection gradient	This study (recommended short form)
$\beta_{\mu} = \mu_P S / \sigma_P^2$ $= \mu_P \beta$	Proportional sensitivity*	Caswell et al. 1978
	Elasticity*	de Kroon et al. 1986
	Elasticity of selection	Morgan & Schoen 1997
	Mean-standardized selection gradient	Hansen et al. 2003
	Fitness elasticity	Stinchcombe 2005
	Mean-standardized selection coefficient	Dibattista et al. 2007
	Mean-and-variance-standardized selection differential	Arlinghaus et al. 2009

Selection gradient in units of the phenotypic mean	This study (recommended long form)
Mean-standardized selection gradient	This study (recommended short form)

* Defined for elements of a population's projection matrix in the context of matrix models of population dynamics

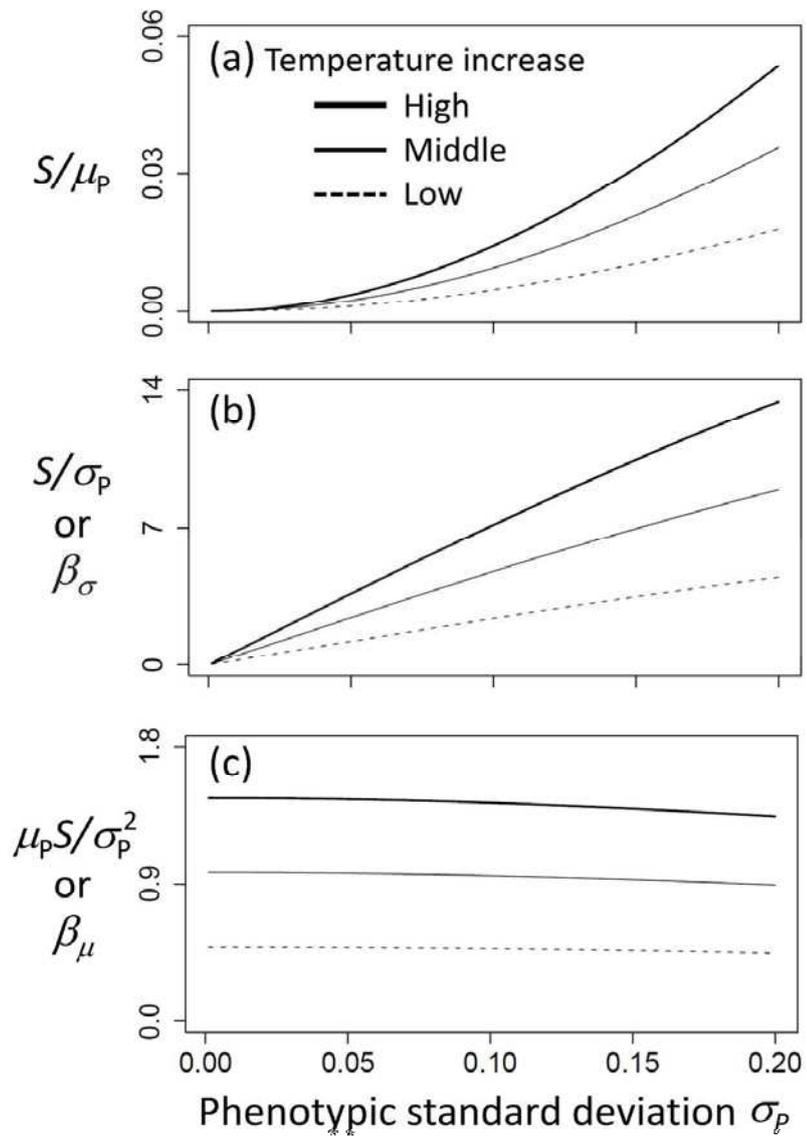


Figure 2

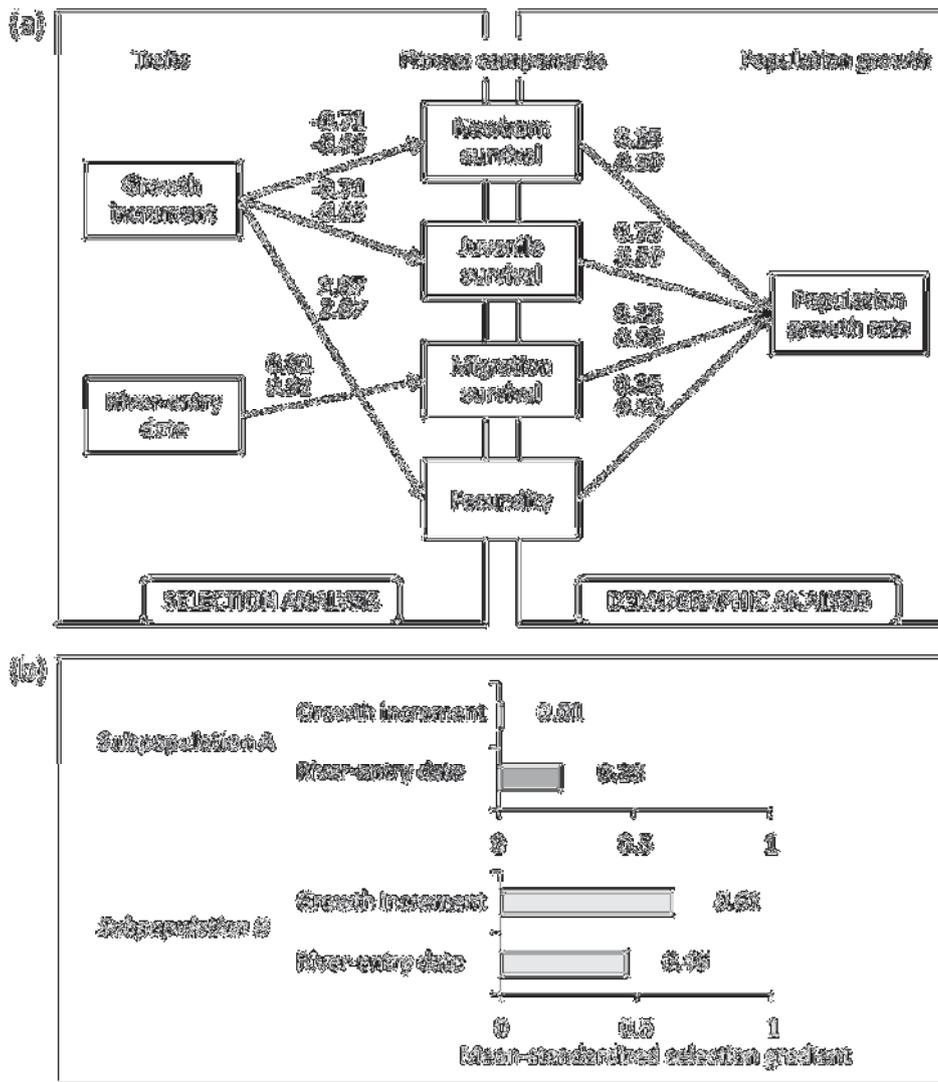


Figure 3

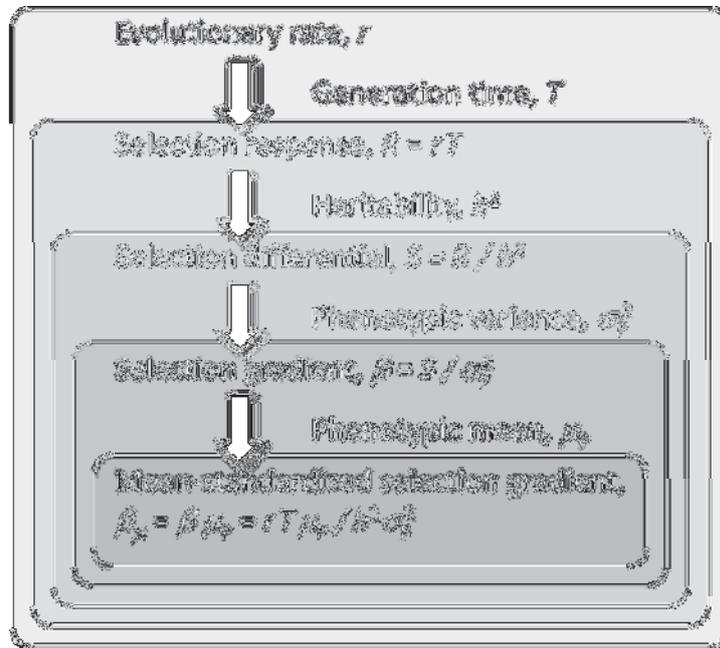


Figure 4