Methods in Ecology and Evolution

BRITISH ECOLOGICAL SOCIETY

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12685

A unified measure of linear and nonlinear selection on quantitative traits

Jonathan M. Henshaw[1]* and Yoav Zemel[2]

[1]Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, 46 Sullivans Creek Road, Acton, Canberra, ACT 02601, Australia; and [2]Chair of Mathematical Statistics, Institute of Mathematics, École Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland

Summary

1. Lande and Arnold's approach to quantifying natural selection has become a standard tool in evolutionary biology due to its simplicity and generality. It treats linear and nonlinear selection in two separate frameworks, generating coefficients of selection (e.g. linear and quadratic selection gradients) that are not directly comparable. Due to this somewhat artificial division, the Lande–Arnold approach lacks an integrated measure of the strength of selection that applies across qualitatively different selection regimes (e.g. directional, stabilizing or disruptive selection).

2. We define a unified measure of selection, the distributional selection differential (DSD), which includes both linear and nonlinear selection. The DSD quantifies total selection on a trait, regardless of the underlying selection regime.

3. The DSD can be partitioned into a directional component, representing selection on the trait mean, and a non-directional component, representing selection on the shape of the trait distribution (e.g. variance, skew or the number of modes). When multiple traits are measured, the DSD can also be separated into direct and correlated effects, analogously to linear selection gradients. As with linear selection differentials, the DSD on a standardized trait is limited in magnitude by the opportunity for selection.

4. The DSD is a general-purpose measure of the total strength of selection. It is particularly valuable where traditional analyses provide limited insight, such as in comparative studies where the shape of selection is variable. Partitioning the DSD into directional and non-directional selection allows biologists to assess whether selection acts consistently in one direction, or in opposing directions over different parts of the trait range.

Key-words: evolutionary biology, population genetics, quantitative genetics

Natural selection is the differential survival and reproduction of individuals with particular traits over their competitors, leading to non-random associations between phenotype and fitness within a generation (Lande & Arnold 1983; Frank 2012; Morrissey 2014). Selection can change not only the means of quantitative traits, but also the shapes of their distributions, including properties like variance, skew and the number and location of modes. For instance, in a study of the medium ground finch *Geospiza fortis*, beak size showed two distinct fitness peaks, indicating selection for greater bimodality of beak size (Hendry *et al*. 2009). Changes in the trait mean are known as linear or directional selection, whereas all other changes in the trait distribution are collectively called nonlinear selection (Phillips & Arnold 1989).

Lande & Arnold's (1983) influential framework quantifies linear and nonlinear selection using two separate regression analyses. This generates coefficients of selection (e.g. linear and quadratic selection gradients: Table 1) that are not directly comparable and that cannot be combined quantitatively to give an overall measure of the strength of selection. We consequently lack an integrated measure of selection that applies regardless of the shape of selection (e.g. directional, stabilizing or disruptive: see Brodie, Moore & Janzen 1995). For instance, we currently cannot compare the strength of selection among traits that experience qualitatively different types of selection, or assess the relative importance of linear and nonlinear selection on a single trait.

We develop a unified index of the strength of selection, the distributional selection differential (DSD) *d*, which incorporates both linear and nonlinear components (Henshaw, Kahn & Fritzsche 2016). The DSD quantifies the total difference in trait distributions between all individuals and those that produce offspring (i.e. the trait distributions before and after selection). It allows the strength of selection to be summarized on a single scale, regardless of the underlying selection regime. The DSD can be partitioned into two components: (i) a directional component, representing the change in the trait mean, which is equal in magnitude to the linear selection differential *s*, and (ii) a non-directional component, representing changes in the shape of the trait distribution (e.g. variance, skew or the number of modes) after accounting for the change in the mean. The DSD is simple to calculate from data on trait values and fitness

*Correspondence author. E-mail: jonathan.henshaw@anu.edu.au

Table 1. Glossary of terms

Term	Description	Definition
Absolute fitness	Unstandardized fitness of individuals (e.g. the number of offspring)	W
Relative fitness	Normalization of absolute fitness so that mean relative fitness equals one	$w = \frac{W}{\mathbb{E}W}$
Linear selection differential	Difference in mean trait values before and after selection	$\boldsymbol{s} = \mathrm{cov}(w, \boldsymbol{Z})$
Phenotypic variance–covariance matrix	Variance–covariance matrix of a trait vector $\boldsymbol{Z}$	Matrix $\boldsymbol{P}$ with entries $\boldsymbol{P}_{ij} = \mathrm{cov}(\boldsymbol{Z}_i, \boldsymbol{Z}_j)$
Linear selection gradient	Partial regression coefficients of fitness on trait values, representing the strength of direct linear selection on standardized traits, assuming that fitness is an additive linear function of trait values	$\boldsymbol{\beta} = \boldsymbol{P}^{-1}\boldsymbol{s}$
Quadratic selection differential	Difference due to selection in the products of pairwise deviations from trait means	$\mathbf{C} = \mathrm{cov}\left(w, (\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})^T\right)$
Quadratic selection gradient	Partial regression coefficients of the quadratic regression of fitness on trait values, providing an approximation of the shape of the fitness surface	$\boldsymbol{\gamma} = \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P}^{-1}$
Gradient condition	Requirement that a function $h(z)$ not change too quickly with trait values z	Grad is the set of functions h such that for all trait values, z_1 and z_2, we have $\lvert h(z_1) - h(z_2)\rvert \leq \lvert z_1 - z_2\rvert$
Distributional selection differential (DSD)	Difference in trait distributions due to selection	*Earth mover's definition:* $d = \min_F \mathbb{E}_F\lvert Z^* - Z\rvert$, where F is any joint probability distribution of the trait values Z and Z* before and after selection *Covariance definition:* $d = \max_{h \in \mathrm{Grad}} \mathrm{cov}(w, h(Z))$ *Cumulative integral definition:* $d = \int_{-\infty}^{\infty} \lvert G^*(z) - G(z)\rvert \mathrm{d}z$, where G and G* are the cumulative distribution functions of trait values before and after selection
Maximizer	Any function of trait values that achieves the maximum possible covariance with relative fitness while satisfying the gradient condition	Any $h \in$ Grad such that $\mathrm{cov}(w, h(Z)) = d$
Maximizer variance–covariance matrix	Variance–covariance matrix of the maximizers $\boldsymbol{h}(\boldsymbol{Z})$ of a trait vector $\boldsymbol{Z}$	Matrix $\boldsymbol{H}$ with entries $\boldsymbol{H}_{ij} = \mathrm{cov}(\boldsymbol{h}_i(\boldsymbol{Z}), \boldsymbol{h}_j(\boldsymbol{Z}))$
Distributional selection gradient	Partial regression coefficients of fitness on the vector of maximizers $\boldsymbol{h}(\boldsymbol{Z})$, representing the total strength of direct selection on standardized traits, assuming that fitness is an additive linear function of maximizers	$\boldsymbol{\delta} = \boldsymbol{H}^{-1}\boldsymbol{d}$

in a population (see eqn 19 below: an implementation in R is available at https://github.com/yoavzemel/dsd).

We provide three mathematically equivalent definitions of the DSD and use them to derive its basic properties. Like the linear selection differential, the DSD is fundamentally limited by phenotypic variance, and for standardized traits, it can be no larger than the square root of the opportunity for selection (i.e. $d \leq \sqrt{\mathrm{var}\, w}$, where w is relative fitness: see Crow 1958; Wade & Arnold 1980; Jones 2009). When selection is purely directional, the DSD and the linear selection differential are equal in magnitude (i.e. $d = \lvert s\rvert$). However, the DSD also accounts for nonlinear selection, which the linear selection differential ignores. For example, under pure stabilizing or disruptive selection, we have $s = 0$, but $d = \lvert\mathrm{cov}(w,\lvert Z\rvert)\rvert$, where Z is a standardized trait value. These two results are particular cases of a general principle: the DSD can always be written as a linear selection differential on a function $h(Z)$ of trait values (i.e. $d = \mathrm{cov}(w,\ h(Z))$), where h meets a technical constraint that ensures it does not vary too steeply with trait values (the 'gradient condition': Table 1).

Selection generally acts on multiple correlated traits simultaneously, and it is useful to separate selection acting directly on a trait from indirect selection due to trait correlations. We consequently also define distributional selection *gradients*, which allow the DSD to be separated into direct and indirect effects, analogously to linear selection gradients in the Lande–Arnold framework.

Our approach to comparing pre- and post-selection trait distributions derives from optimal transport theory and has a long history of application in other disciplines (reviewed in Villani 2009). Similar methods have been applied to categorize images by visual similarity (Rubner, Tomasi & Guibas 1998, 2000), to measure inequalities of wealth or income (the Gini index: Gini 1912; Cowell 2011), to quantify variation in plant size and fecundity (Weiner & Solbrig 1984; Damgaard & Weiner 2000), and to compare anatomical surfaces (Boyer *et al.* 2011) and patterns of animal space use and movement

(Shamoun-Baranes *et al.* 2012; Kranstauber, Smolla & Safi 2016). Closer to the current work, they have also been used to compare trait distributions between populations (Gregorius, Gillet & Ziehe 2003) and to measure the convergence of trait distributions under selection (Rudnicki & Zwoleński 2015; Zwoleński 2015).

Quantifying selection using the Lande–Arnold framework

We first revisit the Lande–Arnold framework for quantifying selection. Trait values before selection can be understood as a random variable Z with probability distribution P. The absolute fitness of an individual (e.g. the number of offspring produced) is a random variable W. Relative fitness w is calculated by normalizing absolute fitness so that mean relative fitness across the population is equal to one:

$$w = \frac{W}{\mathbb{E}W}. \qquad \text{eqn 1}$$

The *fitness function* (or *fitness surface* for multivariate traits) is the mean relative fitness of individuals conditional on their trait values, $\mathbb{E}(w|Z)$. Trait values after selection are represented by a random variable Z^* with distribution $\mathrm{d}P^* = \mathbb{E}(w|Z)\mathrm{d}P$. This means that after selection, the frequency of a trait value equals its frequency before selection times the mean relative fitness of individuals with that trait value. It is important to note that trait values are only measured once: the trait distribution after selection is a hypothetical distribution, calculated from pre-selection trait values and fitness.

For any function $h(Z)$ of trait values, we write $\mathbb{E}h(Z) = \int h(z)\mathrm{d}P(z)$ for its expected value before selection. Its expected value after selection is

$$\mathbb{E}h(Z^*) = \int h(z)\mathrm{d}P^*(z) = \int \mathbb{E}(w|z)h(z)\mathrm{d}P(z) = \mathbb{E}wh(Z). \qquad \text{eqn 2}$$

LINEAR SELECTION DIFFERENTIALS AND GRADIENTS

The change in the mean value of any function of trait values $h(Z)$ due to selection can be written as

$$\mathbb{E}h(Z^*) - \mathbb{E}h(Z) = \mathbb{E}wh(Z) - (\mathbb{E}w)(\mathbb{E}h(Z)) = \mathrm{cov}(w, h(Z)). \qquad \text{eqn 3}$$

This is a generalized form of the Robertson–Price identity (Robertson 1966; Price 1970; Walsh & Lynch 2014). It measures the change in mean function values within a generation between all individuals and those that produce offspring (with the latter weighted by the number of offspring per parent). In particular, the change in the mean trait value is given by the *linear selection differential*:

$$s := \mathbb{E}Z^* - \mathbb{E}Z = \mathrm{cov}(w, Z). \qquad \text{eqn 4}$$

The multivariate version of this definition takes the same form, except that s and Z are replaced by vectors $\boldsymbol{s}$ and $\boldsymbol{Z}$ of selection differentials and trait values, respectively, and the covariance is taken between w and each component of $\boldsymbol{Z}$.

When multiple traits are measured, *linear selection gradients* remove the effects of phenotypic correlations among measured traits, which allows linear selection on each trait to be separated into direct and indirect (i.e. correlated) effects (Lande 1982; Lande & Arnold 1983). Linear selection gradients are defined as the vector of partial regression coefficients $\boldsymbol{\beta}$ of relative fitness on the trait vector $\boldsymbol{Z}$ according to the model

$$w = 1 + \boldsymbol{\beta}^T(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z}) + \varepsilon. \qquad \text{eqn 5}$$

Here, ε is an error term with mean zero. Linear selection gradients and differentials are related by the equation $\boldsymbol{s} = \boldsymbol{P\beta}$, where $\boldsymbol{P}$ is the phenotypic variance–covariance matrix. Direct selection on a trait i is then given by $\boldsymbol{P}_{ii}\boldsymbol{\beta}_i$, whereas indirect selection due to trait correlations is $\sum_{j\neq i}\boldsymbol{P}_{ij}\boldsymbol{\beta}_j$. Note that this interpretation assumes that fitness is an additive linear function of trait values (Mitchell-Olds & Shaw 1987). The term *selection gradients* stems from the property that if the traits follow a multivariate normal distribution, then $\boldsymbol{\beta} = \mathbb{E}\frac{\partial\mathbb{E}(w|\boldsymbol{Z})}{\partial\boldsymbol{Z}}$ is the average gradient of the fitness surface with respect to trait values (Lande & Arnold 1983). This property is no longer valid when the normality assumption is violated (Morrissey & Sakrejda 2013).

QUADRATIC SELECTION DIFFERENTIALS AND GRADIENTS

The (univariate) *quadratic selection differential* represents the change in the average squared deviation from the pre-selection trait mean, $(Z - \mathbb{E}Z)^2$, due to selection. That is,

$$c := \mathrm{cov}(w, (Z - \mathbb{E}Z)^2). \qquad \text{eqn 6}$$

In a multivariate setting, the quadratic selection differential is the matrix $\boldsymbol{C}$ of changes in the average product of pairwise deviations from pre-selection trait means:

$$\boldsymbol{C} := \mathrm{cov}(w, (\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})^T). \qquad \text{eqn 7}$$

Note that the covariance is taken between w and each entry in the matrix $(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})^T$. When there is no directional selection (i.e. when $\boldsymbol{s} = 0$), the quadratic selection differential equals the change in the phenotypic variance-covariance matrix $\boldsymbol{P}$ due to selection (i.e. $\boldsymbol{C} = \boldsymbol{P}^* - \boldsymbol{P}$). More generally, it is given by the less interpretable relationship $\boldsymbol{C} = \boldsymbol{P}^* - \boldsymbol{P} + \boldsymbol{ss}^T$ (Lande & Arnold 1983; Mitchell-Olds & Shaw 1987; Schluter 1988). Linear and quadratic selection differentials are not measured in the same units, and so quantitative comparisons between them are meaningless.

Quadratic selection gradients give an approximate description of the shape of the fitness surface (Phillips & Arnold 1989). They are defined as the partial regression coefficients $\boldsymbol{\gamma}$ corresponding to the quadratic terms in the quadratic regression of relative fitness on trait values (Lande & Arnold 1983; Phillips & Arnold 1989; Morrissey & Sakrejda 2013):

$$w = a + \boldsymbol{b}^T\boldsymbol{Z} + \frac{1}{2}(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z})^T\boldsymbol{\gamma}(\boldsymbol{Z} - \mathbb{E}\boldsymbol{Z}) + \varepsilon. \qquad \text{eqn 8}$$

Note that the partial regression coefficients $\boldsymbol{b}$ for the linear terms in this quadratic regression usually differ from those in the linear regression above (i.e. $\boldsymbol{b} \neq \boldsymbol{\beta}$ in general). Quadratic selection gradients and differentials are related by $\boldsymbol{\gamma} = \boldsymbol{P}^{-1}\boldsymbol{C}\boldsymbol{P}^{-1}$. When trait values are multivariate normal, the quadratic selection gradients $\boldsymbol{\gamma}$ equal the average curvature of the fitness surface (i.e. $\boldsymbol{\gamma} = \mathbb{E}\frac{\partial^2 \mathbb{E}(w|\boldsymbol{Z})}{\partial \boldsymbol{Z}^2}$: see Lande & Arnold 1983; Morrissey & Sakrejda 2013). In any case, the interpretation of $\boldsymbol{\gamma}$ assumes that the fitness surface is a quadratic function of $\boldsymbol{Z}$ (Schluter & Nychka 1994).

Three equivalent definitions of the DSD

We now provide three mathematically equivalent definitions of the DSD (Table 1). This plurality is useful both in visualizing the DSD (Fig. 1) and in deriving its basic properties. A formula for calculating the DSD for empirical applications is given in eqn 19 (see https://github.com/yoavzemel/dsd for an implementation in R).

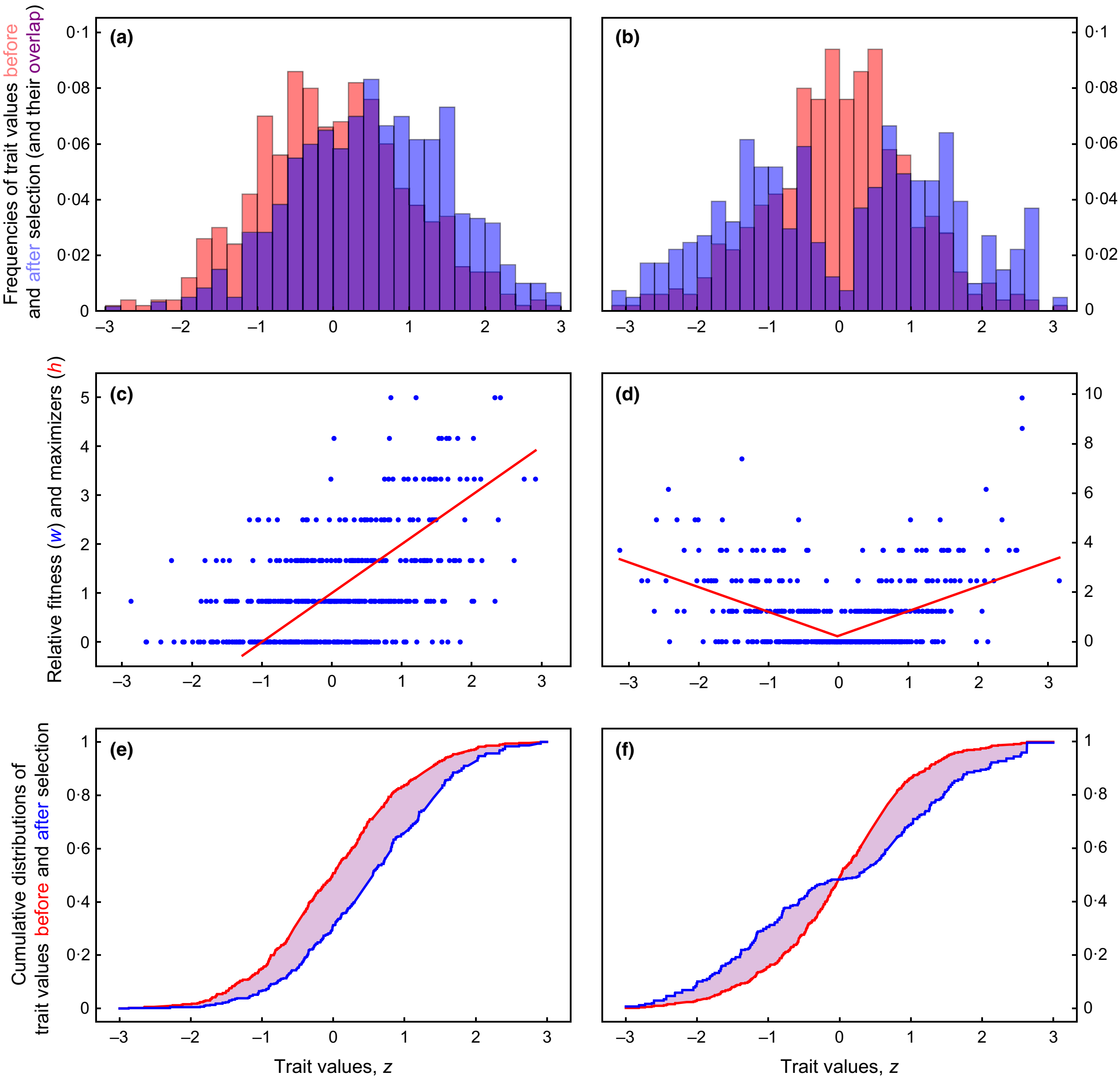

Fig. 1. Illustration of three equivalent definitions of the distributional selection differential (DSD), based on simulations of populations under directional selection (a, c, e) and disruptive selection (b, d, f). The DSDs and linear selection differentials are $d = 0{\cdot}53$, $s = 0{\cdot}53$ (directional selection) and $d = 0{\cdot}54$, $s = 0{\cdot}04$ (disruptive selection). (a, b) *Earth mover's definition*: Binned histograms of trait distributions before selection (red) and after selection (blue); areas where the two distributions overlap are shaded purple. The DSD is approximately equal to the minimum amount of work needed to shift the 'red' mass until it fully coincides with the 'blue' mass ('approximately' because the binned histogram pictured is a simplification of the empirical distribution). (c, d) *Covariance definition*: Relative fitness w (blue dots) and maximizers h (red line; translated so that $\mathbb{E}h(Z) = 1$ for ease of visualization). The DSD is $d = \text{cov}(w, h(Z))$. (e, f) *Cumulative integral definition*: Cumulative distribution functions of trait values before selection (G, red) and after selection (G^*, blue). The DSD is equal to the area between the two curves (shaded purple).

EARTH MOVER'S DEFINITION

Our first definition comes with a neat visual metaphor: if we imagine the distributions of trait values before and after selection as piles of sand, then the DSD is the minimum amount of 'work' needed to transform one pile into the other (Fig. 1a,b). Any difference in the shape of the pre- and post-selection trait distributions will necessitate some amount of sand being shifted, and so this image captures selection of all types. To make the metaphor precise, we need to specify how much work is required to move mass around. We assume that work is equal to the amount of mass times the distance moved. Other assumptions are certainly possible (e.g. using the squared distance), but ours has the advantage of ensuring that the DSD is measured in the same units as the linear selection differential.

To build intuition, we begin with the case where a trait can take only a finite number of values, $z_1, \ldots, z_n$. Before selection, each trait value z_i is obtained by a proportion p_i of the population. After selection, the frequency of each trait value is $p_i^* = p_i w_i$, where w_i is the mean relative fitness of individuals with trait value z_i. In this discrete case, we can visualize a distribution of trait values as n point masses, such that the total mass adds to one.

A *flow F* between trait distributions is an $n \times n$ matrix with non-negative entries such that for all i (Rubner, Tomasi & Guibas 2000; Levina & Bickel 2001),

$$\sum_j F_{ij} = p_i, \qquad \text{eqn 9}$$

and for all j,

$$\sum_i F_{ij} = p_j^*. \qquad \text{eqn 10}$$

We interpret F_{ij} as the amount of mass moved from trait value z_i to z_j. The first condition ensures that the total mass moved from each pile i equals the size of the pile p_i before selection. The second ensures that the total mass moved to each pile j equals its size p_j^* after selection. Note that it is permissible to move mass from a pile to itself at zero cost.

We assume that the work needed to move mass from z_i to z_j is the product of the amount of mass moved F_{ij} and the difference $|z_j - z_i|$ between the two trait values. For a given flow F, the total work needed to transform the trait distribution is then

$$d_F = \sum_{i,j} F_{ij}|z_j - z_i|. \qquad \text{eqn 11}$$

We can reformulate this definition in terms of standard concepts in probability theory (Levina & Bickel 2001). From eqns 9 and 10, we see that a flow can be viewed as a probability mass function over all pairs of trait values, with marginal distributions equalling the pre- and post-selection trait distributions Z and Z*. The amount of work associated with the flow F can then be written as

$$d_F = \mathbb{E}_F|Z^* - Z|. \qquad \text{eqn 12}$$

We define the DSD as the minimum amount of work required to transform the pre-selection trait distribution into its post-selection form, taken over all possible flows:

$$d := \min_F d_F = \min_F \mathbb{E}_F|Z^* - Z|. \qquad \text{eqn 13}$$

This definition generalizes easily to cases where the space of possible trait values is infinite (e.g. continuous distributions of quantitative traits). In this case, a flow is a simply a joint probability distribution of Z and Z*. We refer to eqn 13 as the *earth mover's definition* of the DSD, based on a similar metric used in computer science (Rubner, Tomasi & Guibas 1998, 2000). In mathematics, it is known by many names, but most commonly as the *Wasserstein distance* (Villani 2009).

The minimum value in the earth mover's definition always exists (Villani 2009) and we call a flow *optimal* if it achieves this minimum (i.e. F is optimal if $d_F = d$). Both the definition of work in eqn 11 and the constraints in eqns 9 and 10 are linear in the matrix values F_{ij}, so for empirical applications an optimal flow can be found using standard linear programming software. However, alternative approaches to calculating the DSD are more practical, as they provide explicit expressions for the DSD and related optimizers (see eqns 19 and 20).

COVARIANCE DEFINITION

Linear selection differentials can be written as covariances $s = \text{cov}(w, Z)$ between relative fitness and trait values. We now show that the DSD can be expressed similarly as $d = \text{cov}(w, h(Z))$, where h is a function that is constrained not to vary too quickly with changes in trait values.

A precise definition of 'too quickly' is as follows. A function h satisfies the *gradient condition* if for any two trait values z_1 and z_2, we have

$$|h(z_1) - h(z_2)| \le |z_1 - z_2|. \qquad \text{eqn 14}$$

This means that the gradient of h between any two trait values must always lie between plus and minus one. Consequently, h changes no more quickly than the identity function $f(z) = z$ with trait values. We write Grad for the set of functions satisfying the gradient condition. In mathematics, these are known as Lipschitz (or, more precisely, 1-Lipschitz) functions (Villani 2009).

For any function $h(Z)$ of trait values, we can consider the change in its mean due to selection (i.e. $\mathbb{E}h(Z^*) - \mathbb{E}h(Z)$). The Kantorovich–Rubinstein theorem provides a deep connection between optimal flows and changes in mean function values (Gibbs & Su 2002; Villani 2009). It states that the DSD is equal to the maximum change in mean function values, taken over all functions h that satisfy the gradient condition:

$$d = \max_{h \in \text{Grad}} (\mathbb{E}h(Z^*) - \mathbb{E}h(Z)). \qquad \text{eqn 15}$$

We know from eqn 3 that differences in mean function values can be written as covariances between the function and relative fitness. From this, we obtain the *covariance definition* of the DSD:

$$d = \max_{h \in \text{Grad}} \text{cov}(w, h(Z)). \qquad \text{eqn 16}$$

We refer to any function h that obtains the maximum covariance in this definition as a *maximizer*. A consequence of eqn 16 is that the DSD can always be written as a linear selection differential $d = \mathrm{cov}(w, h(Z))$, where h is a maximizer (Fig. 1c,d).

It is possible to explicitly construct maximizers. We write G and G^* for the cumulative distribution functions of trait values before and after selection [i.e. the functions $G(z) = \mathbb{P}(Z \leq z)$ and $G^*(z) = \mathbb{P}(Z^* \leq z)$]. A maximizer is then any function of the form (see Appendix S1 for derivation):

$$h(z) = h(0) + \int_0^z \mathrm{sgn}(G(x) - G^*(x))\mathrm{d}x. \qquad \text{eqn 17}$$

In simple terms, h increases when $G(x) > G^*(x)$ and decreases when $G(x) < G^*(x)$. In both cases, the change is linear with a slope of one.

Maximizers h can be thought of as rough approximations to the fitness function, but with gradients constrained to lie between plus and minus one (Fig. 2). More precisely, they

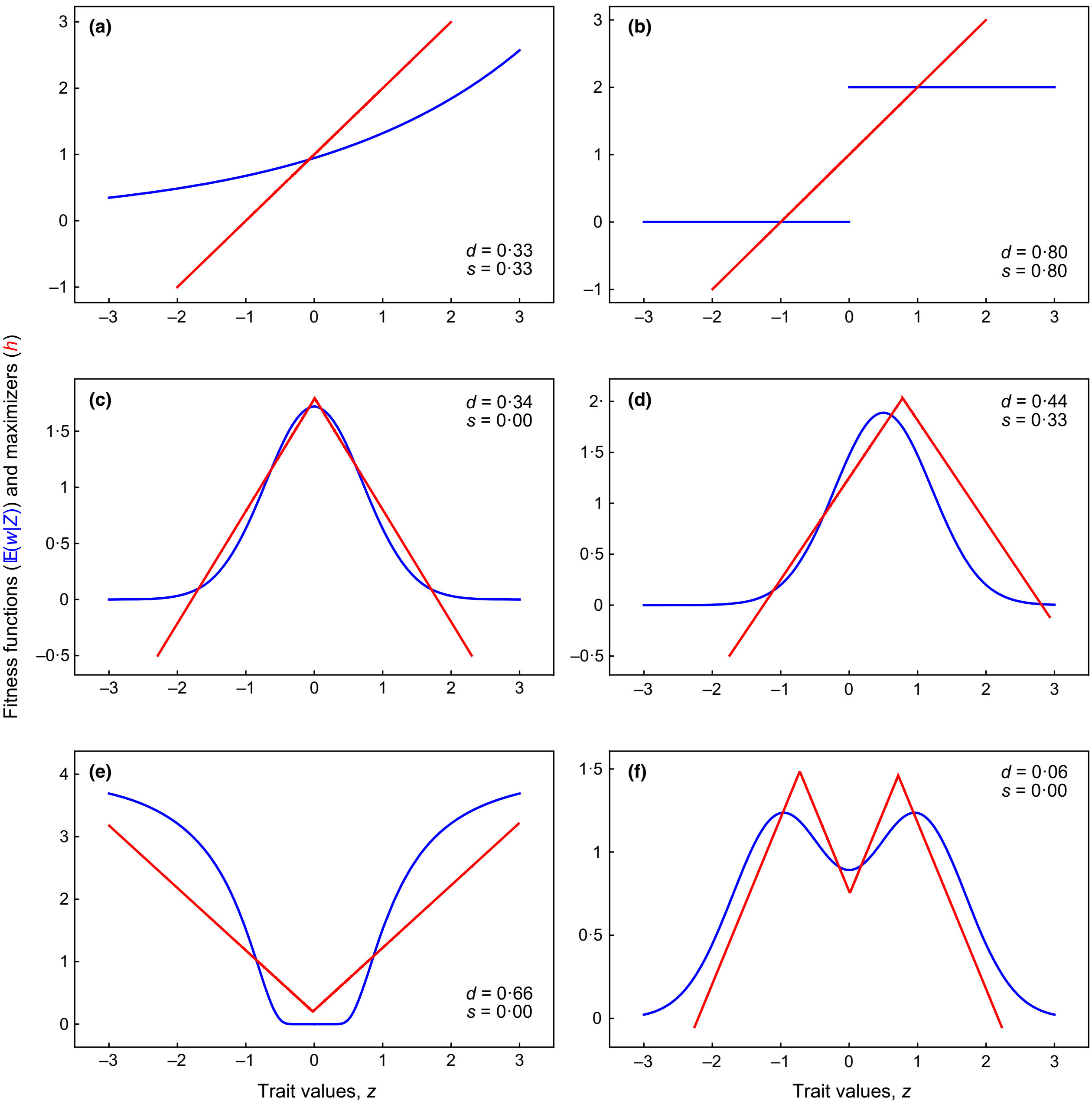

Fig. 2. Fitness functions (blue) and maximizers h (red) for six different selection regimes, shown with the distributional selection differential (DSD), d, and linear selection differential, s: (a) Smooth directional selection for higher trait values, based on mean absolute fitness $W = \exp(Z/3)$. (b) Truncation selection for higher trait values, such that only individuals in the top half of the trait distribution are selected to reproduce. (c) Pure stabilizing selection, based on $W = \exp(-Z^2)$. (d) A combination of directional and stabilizing selection, based on $W = \exp(-(Z-1)^2)$. (e) Pure disruptive selection, based on $W = \exp(-1/Z^2)$. (f) M-shaped selection, based on $W = \exp(-(Z+1)^2) + \exp(-(Z-1)^2)$. Note that different selection regimes may generate the same maximizer h (compare a and b). Trait values are assumed to follow a standard normal distribution.

show in which direction mass is moved to transform the pre-selection trait distribution into the post-selection distribution under an optimal flow (details in Appendix S1). Intervals where h is increasing (with a gradient of +1) correspond to 'uphill' movement from lower to higher trait values, while decreasing h (with a gradient of -1) corresponds to 'downhill' movement. For example, under pure directional selection for higher (lower) trait values, the maximizer h is a simple straight line with positive (negative) gradient. Under pure stabilizing selection, h is a pointed 'hill' centred at the mean trait value, indicating that selection moves mass towards the mean from both directions (Fig. 2).

Maximizers are not uniquely defined, for two reasons. First, the value of $h(0)$ is always arbitrary, so h can only ever be unique up to a constant. Secondly, whenever $G(x) = G^*(x)$, the value of $\mathrm{sgn}(G(x) - G^*(x)) = \mathrm{sgn}(0)$ in eqn 17 is not uniquely determined, but can be chosen freely from the interval $[-1,1]$. This leads to non-unique h in cases where the set $\{x\colon G(x) = G^*(x)\}$ is large (in technical terms, when it has positive Lebesgue measure).

The non-uniqueness of maximizers is no problem for the covariance definition of the DSD, because all maximizers have the same covariance with relative fitness. However, when we define distributional selection gradients below, it becomes necessary to pick out and work with a particular maximizer. In this case, we suggest (i) following the convention $\mathrm{sgn}(0) = 0$, which ensures that maximizers are unique up to an additive constant, and (ii) choosing $h(0)$ so that $\mathbb{E}h(Z) = 0$, which fixes the constant.

CUMULATIVE INTEGRAL DEFINITION

The third definition of the DSD relates it to differences in the cumulative distributions of trait values before and after selection. We write G and G^* for the cumulative distribution functions, as above, and G^{-1} and G^*-1 for their generalized inverses, the quantile functions (Kämpke & Radermacher 2015). The *cumulative integral definition* of the DSD is (Gibbs & Su 2002; Villani 2003; Henshaw, Kahn & Fritzsche 2016):

$$d = \int_{-\infty}^{\infty} |G(z) - G^*(z)|\mathrm{d}z = \int_{0}^{1} \left|G^{-1}(q) - G^{*-1}(q)\right|\mathrm{d}q. \qquad \text{eqn 18}$$

Geometrically, the DSD is equal to the area between the cumulative distribution curves of trait values before and after selection (Fig. 1e,f).

EMPIRICISTS READ HERE: CALCULATING THE DSD FROM FINITE SAMPLES

In empirical applications, trait values and fitness are measured on a finite number of individuals. We can then approximate the trait and fitness distributions of the population by the empirical distributions of the sample. Suppose that n individuals are sampled with trait values $z_1, \ldots, z_n$ and relative fitness $w_1, \ldots, w_n$. We assume that these values are sorted so that $z_1 \leq z_2 \leq \cdots \leq z_n$. Using the cumulative integral definition, the DSD is then

$$d = \sum_{i=1}^{n-1}(z_{i+1} - z_i)\left|\sum_{j=1}^{i}\frac{1 - w_j}{n}\right|. \qquad \text{eqn 19}$$

Confidence intervals for d can be calculated using bootstrapping, assuming sufficient sample sizes.

Using the covariance definition, the DSD is given equivalently by $d = \mathrm{cov}(w, h(Z))$, where the maximizer h is defined iteratively by (see Appendix S1):

$$h(z_{i+1}) - h(z_i) = (z_{i+1} - z_i)\mathrm{sgn}\left(\sum_{j=1}^{i}\frac{1 - w_j}{n}\right). \qquad \text{eqn 20}$$

As noted above, the maximizer h is not uniquely defined. In this formulation, the first value $h(z_1)$ is arbitrary, and when $\sum_{j=1}^{i}\frac{1-w_j}{n} = 0$, the value of $\mathrm{sgn}\left(\sum_{j=1}^{i}\frac{1-w_j}{n}\right)$ can be chosen freely from the interval $[-1,1]$. When it is desirable to work with a particular maximizer, we suggest using the convention $\mathrm{sgn}(0) = 0$ and shifting h so that $\mathbb{E}h(Z) = 0$, as described following eqn 17.

WHAT UNITS IS THE DSD MEASURED IN?

Like the linear selection differential, the DSD is measured in the same units as the original trait values. This is intuitive, as both measures represent changes in trait values due to selection. It is often helpful to standardize the value of each trait (before selection) by first subtracting its mean value and then dividing by its standard deviation:

$$Z' = \frac{Z - \mathbb{E}Z}{\sqrt{\mathrm{var}\,Z}}. \qquad \text{eqn 21}$$

If traits are standardized, then both linear selection differentials and the DSD are in units of trait standard deviations, which aids in comparison among traits (Jones 2009).

SIMULATED EXAMPLES

To illustrate the three equivalent definitions of the DSD, we simulated populations under two selection regimes (Fig. 1). For each population, we drew trait values Z of 500 individuals from a standard normal distribution. We then simulated absolute fitness (representing, e.g. the number of offspring produced) as either:

1 $W \sim \mathrm{Poisson}(\exp(Z/2))$, representing directional selection for higher trait values, or

2 $W \sim \mathrm{Poisson}(|Z|)$, representing disruptive selection for extreme trait values.

We approximated the true distributions of traits and relative fitness by the empirical distributions of the samples and then calculated the DSD and maximizers using eqns 19 and 20, respectively. The DSD was approximately equal to the linear selection differential under directional selection (d, s = 0·53, Fig. 1a,c,e). Under disruptive selection, the linear selection differential was close to zero (s = 0·04) but the DSD still captured the change in the shape of the trait distribution due to selection (d = 0·54, Fig. 1b,d,f).

Fundamental properties

DIRECTIONAL, STABILIZING AND DISRUPTIVE SELECTION

We now formally consider how the DSD behaves under three important types of selection: directional, stabilizing and disruptive (Lande & Arnold 1983; Schluter 1988; Phillips & Arnold 1989). We say that a trait is under pure directional selection if relative fitness w is a monotonic function of trait values (cf. Mitchell-Olds & Shaw 1987; Schluter 1988). If w is non-decreasing, then $G \geq G^*$ and so eqn 17 implies that $h(z) = z$ is a maximizer. Similarly, $h(z) = -z$ is a maximizer when w is non-increasing. In both cases, the covariance definition of the DSD gives us:

$$d = |\mathrm{cov}(w, Z)| = |s|. \qquad \text{eqn 22}$$

Similarly, we say that pure stabilizing selection occurs when (i) the trait distribution before selection is symmetrical about the mean trait value and (ii) relative fitness w is a symmetric non-increasing function of distance from the trait mean (i.e. a non-increasing function of $|Z - \mathbb{E}Z|$). Pure disruptive selection can be defined in the same way, except replacing 'non-increasing' with 'non-decreasing'. A consequence of these definitions is that directional selection is absent (i.e. s = 0). Under pure stabilizing or disruptive selection, the DSD is given by

$$d = |\mathrm{cov}(w, |Z - \mathbb{E}Z|)|. \qquad \text{eqn 23}$$

The quantity $\mathrm{cov}(w, |Z - \mathbb{E}Z|)$ is conceptually similar to the univariate quadratic selection differential c in eqn 6, in that it quantifies a change in dispersion from the pre-selection trait mean. However, it measures the change in the average absolute distance of trait values from the mean, rather than the average squared distance. This is actually an advantage of the DSD: because it is defined via functions that satisfy the gradient condition, it is measured on the same scale as the linear selection differential. In contrast, the magnitudes of linear and quadratic selection differentials in the original Lande–Arnold framework are not directly comparable.

DIRECTIONAL AND NON-DIRECTIONAL COMPONENTS OF THE DSD

The DSD can be partitioned to reflect the relative contributions of directional and non-directional selection. Suppose F is an optimal flow between the trait distributions before and after selection. The redistribution of mass under F is equivalent to the sequential action of two flows, D and N, which can be thought of as two consecutive episodes of selection (Arnold & Wade 1984; see Appendix S1 for details, where we also consider the DSD under multiple episodes of selection more generally). The directional flow D shifts the mean trait value by moving mass in only one direction (i.e. from lower to higher trait values or vice versa), at a cost of $d_D = |s|$. The non-directional flow N reshapes the trait distribution without changing its mean, at a cost of $d_N = d - |s|$.

The DSD can thus be partitioned as $d = d_D + d_N$, where (i) the directional component $d_D = |s|$ represents the change in the trait mean and (ii) the non-directional component $d_D = d - |s|$ represents changes in the shape of the trait distribution (e.g. variance, skew or the number of modes) after accounting for any shift in the mean.

THE DSD IS LIMITED BY VARIANCE IN FITNESS

The linear selection differential on any standardized trait is at most as large as the standard deviation in relative fitness (i.e. $|s| \leq \sqrt{\mathrm{var}\, w}$). For this reason, the variance in relative fitness is sometimes referred to as the *opportunity for selection I* (Crow 1958; Wade & Arnold 1980; Jones 2009). We now show that the DSD on a standardized trait is bounded by the same quantity.

The covariance definition of the DSD implies that for any maximizer h, we have

$$d = \mathrm{cov}(w, h(Z)) \leq \sqrt{\mathrm{var}\, w \cdot \mathrm{var}\, h(Z)}. \qquad \text{eqn 24}$$

The variance of $h(Z)$ is at most as large as the variance in trait values Z, because

$$\begin{aligned} \mathrm{var}\, h(Z) &= \mathbb{E}(h(Z) - \mathbb{E}h(Z))^2 \leq \mathbb{E}(h(Z) - h(\mathbb{E}Z))^2 \\ &\leq \mathbb{E}(Z - \mathbb{E}Z)^2 = \mathrm{var}\, Z. \end{aligned} \qquad \text{eqn 25}$$

The first inequality follows because the average squared deviation from the mean is always smaller than the average squared deviation from any other quantity; the second is from the gradient condition on h. This gives us

$$d \leq \sqrt{\mathrm{var}\, w \cdot \mathrm{var}\, Z}. \qquad \text{eqn 26}$$

In particular, when Z is a standardized trait, we have $d \leq \sqrt{\mathrm{var}\, w} = \sqrt{I}$. Therefore, like the linear selection differential, the DSD on a standardized trait is no larger than the square root of the opportunity for selection.

Direct and indirect selection on correlated traits

When multiple traits are measured, linear selection gradients allow linear selection on a trait to be separated into direct and indirect effects, where the latter arise via correlations among measured traits. Here, we analogously define distributional selection gradients, which allow us to separate the DSD into direct and indirect effects.

We begin by noting the similarity between the covariance definitions of linear selection differentials and the DSD (i.e. $s = \mathrm{cov}(w, Z)$ and $d = \mathrm{cov}(w, h(Z))$, where h is a maximizer). We can use this analogy to construct a regression equation similar to eqn 5, namely:

$$w = 1 + \boldsymbol{\delta}^T \boldsymbol{h}(\boldsymbol{Z}) + \varepsilon. \qquad \text{eqn 27}$$

Here, $\boldsymbol{h}(\boldsymbol{Z})$ is a vector of maximizers of the trait vector $\boldsymbol{Z}$ and ε is an error term with mean zero. We assume that maximizers are standardized using the procedure following eqn 17, which ensures that $\mathbb{E}w = 1$. The *distributional selection gradients* are defined as the vector of partial regression coefficients $\boldsymbol{\delta}$.

For linear selection, the relationship between selection gradients and selection differentials is expressed by the vector equation $\boldsymbol{\beta} = \boldsymbol{P}^{-1}\boldsymbol{s}$, where $\boldsymbol{P}$ is the phenotypic variance–covariance matrix. Similarly, we can write the distributional selection gradients as

$$\boldsymbol{\delta} = \boldsymbol{H}^{-1}\boldsymbol{d}, \qquad \text{eqn 28}$$

where $\boldsymbol{H}$ is the variance–covariance matrix of the maximizers $\boldsymbol{h}(\boldsymbol{Z})$ and $\boldsymbol{d}$ is the vector of DSDs. The direct effect of the ith trait on fitness is then given by $\boldsymbol{H}_{ii}\boldsymbol{\delta}_i$, and the indirect effect due to correlated traits is $\sum_{j \neq i} \boldsymbol{H}_{ij}\boldsymbol{\delta}_j$.

Note that unlike the definition of the DSD, which makes no assumptions about the relationship between fitness and trait values, the interpretation of distributional selection gradients as the direct components of selection assumes that fitness is a linear and additive function of the maximizers. As with all multivariate regressions, small deviations from these assumptions will likely be tolerated, but large deviations will hinder the interpretability of the results.

An alternative approach to defining distributional selection gradients involves fewer assumptions about the distributions of $\boldsymbol{h}(\boldsymbol{Z})$ and fitness. First, the fitness surface on $\boldsymbol{h}(\boldsymbol{Z})$ can be approximated using a semi-parametric method such as cubic splines or a generalized additive model (Schluter 1988; Schluter & Nychka 1994; Morrissey & Sakrejda 2013). The approach of Morrissey & Sakrejda (2013) can then be used to obtain selection gradients by averaging local gradient values over the approximated fitness surface.

Discussion

The DSD provides an integrated measure of selection on a trait, including both linear and nonlinear selection. It measures how much the trait distribution differs within a generation between all individuals and the parents of offspring. Although the DSD quantifies the overall strength of selection, it does not provide qualitative information on the shape of the fitness function or even the direction of mean trait change. In particular, many different selection regimes may generate the same value of the DSD. We consequently see the DSD as complementary to existing approaches that provide such qualitative information, including linear and quadratic selection differentials and gradients (Lande & Arnold 1983) and numerical approximation of fitness surfaces (Schluter 1988; Schluter & Nychka 1994; Shaw & Geyer 2010; Morrissey & Sakrejda 2013). The maximizer h associated with the covariance definition of the DSD also provides a visual summary of fitness functions (Figs 1c,d and 2).

The distributional approach is general and can be used whenever the total strength of selection must be quantified. It may be particularly useful in cases where traditional analyses provide limited insight. First, the DSD can be used to compare the strength of selection across traits, taxa or environments that differ qualitatively in their selection regime. For example, directional selection on one trait can be compared with stabilizing selection on another. By contrast, linear and quadratic selection coefficients are not measured in the same units, and so the existing framework does not allow for comparison of selection across modes.

Secondly, we have shown that the DSD can be partitioned into a directional component, representing the change in the trait mean, and a non-directional component, representing changes in the shape of the trait distribution after accounting for the change in the mean. The DSD can consequently be used to compare the relative importance of directional and non-directional (e.g. stabilizing or disruptive) selection on a given trait. We believe that this is more informative than the traditional distinction between linear and nonlinear selection. Biologists are generally more interested in the direction of selection than in the precise shape of the fitness surface (Mitchell-Olds & Shaw 1987). When fitness is an increasing (or decreasing) function of trait values, non-directional selection d_N in our framework will equal zero, even though curvature in the fitness surface may still lead to nonzero quadratic selection gradients (Schluter 1988). Our approach consequently allows biologists to assess whether selection acts consistently in one direction, or in opposing directions over different parts of the trait distribution.

Thirdly, the DSD can be used to quantify selection when the fitness surface is complex in shape. For instance, some species show bi- or multimodal trait distributions, where selection may (at least conceivably) be stabilizing around each peak but disruptive between them (e.g. M-shaped selection: Rueffler *et al.* 2006; Hendry *et al.* 2009). Potential examples include size dimorphism in social insect queens (Heinze & Tsuji 1995; Wolf & Seppä 2016), horn size in male beetles with alternative reproductive tactics (Eberhard & Gutierrez 1991; Moczek & Emlen 2000; Nijhout 2003; Buzatto, Tomkins & Simmons 2014) and adaptive divergence of bill size in bird species that straddle multiple ecological niches (Smith 1993; Hendry *et al.* 2009). The DSD can be used to quantify the total selection acting on such trait distributions, including both stabilizing and disruptive components.

Like traditional selection differentials and gradients, the DSD and distributional selection gradients are strictly descriptions of the relationship between phenotypes and fitness. Inferences about the resulting change across generations (i.e. the response to selection, whether genetic or phenotypic) are complicated by the complexity of the genotype–phenotype relationship, including the possibility of unmeasured traits or environmental factors that covary with both measured traits and fitness (Mitchell-Olds & Shaw 1987; Rausher 1992). Importantly, selection will only lead to cross-generational change if trait values and fitness are associated at the genetic level (Morrissey, Kruuk & Wilson 2010).

Although our current analysis is restricted to selection on phenotypes, it may be possible to derive a genetic analogue of the DSD using estimated breeding values of traits instead of phenotypic values. Any such approach would need to account for statistical weaknesses in the derivation of breeding values (Hadfield *et al.* 2010). An alternative approach is to incorporate distributional thinking directly into the construction of animal models (Kruuk 2004; Wilson *et al.* 2010). A 'genetic' DSD could then be constructed without relying on estimates of individual breeding values [cf. Stinchcombe, Simonsen & Blows (2014), where linear selection gradients are carried into a genetic setting using the secondary theorem of natural selection].

Authors' contributions

J.M.H. and Y.Z. conceived the ideas and designed the methodology; J.M.H. designed the simulations and produced the figures; Y.Z. wrote the online implementation in R; J.M.H. and Y.Z. wrote the manuscript and approved its final form.

Acknowledgements

J.M.H. would like to thank Michael Jennions for his continued support and for feedback on the manuscript. Valuable feedback and comments were provided by Karoline Fritzsche, Loeske Kruuk, Steve Ramm, Kiyoko Gotanda, and the laboratory group of Andrew Hendry. J.M.H. is supported by the Gwendolyn Woodroofe PhD Scholarship in the Biological Sciences. Y.Z. is supported by a European Research Council Starting Grant Award awarded to Victor M. Panaretos.

Data accessibility

This article does not use any data.

References

Arnold, S.J. & Wade, M.J. (1984) On the measurement of natural and sexual selection: theory. *Evolution*, **38**, 709–719.

Boyer, D.M., Lipman, Y., St Clair, E., Puente, J., Patel, B.A., Funkhouser, T., Jernvall, J. & Daubechies, I. (2011) Algorithms to automatically quantify the geometric similarity of anatomical surfaces. *Proceedings of the National Academy of Sciences, USA*, **108**, 18221–18226.

Brodie, E.D. III, Moore, A.J. & Janzen, F.J. (1995) Visualizing and quantifying natural selection. *Trends in Ecology & Evolution*, **10**, 313–318.

Buzatto, B.A., Tomkins, J.L. & Simmons, L.W. (2014) Alternative phenotypes within mating systems. *The Evolution of Insect Mating Systems* (eds D.M. Shuker & L.W. Simmons), pp. 106–128. Oxford University Press, Oxford, UK.

Cowell, F.A. (2011) *Measuring Inequality*, 3rd edn. Oxford University Press, Oxford, UK.

Crow, J.F. (1958) Some possibilities for measuring selection intensities in man. *Human Biology*, **30**, 1–13.

Damgaard, C. & Weiner, J. (2000) Describing inequality in plant size or fecundity. *Ecology*, **81**, 1139–1142.

Eberhard, W.G. & Gutierrez, E.E. (1991) Male dimorphisms in beetles and earwigs and the question of developmental constraints. *Evolution*, **45**, 18–28.

Frank, S.A. (2012) Natural selection. IV. The Price equation. *Journal of Evolutionary Biology*, **25**, 1002–1019.

Gibbs, A.L. & Su, F.E. (2002) On choosing and bounding probability metrics. *International Statistical Review*, **70**, 419–435.

Gini, C. (1912) Variabilità e Mutabilità. *Studi Economico-Giuridici delli Università di Cagliari*, **3**, 1–158.

Gregorius, H.-R., Gillet, E.M. & Ziehe, M. (2003) Measuring differences of trait distributions between populations. *Biometrical Journal*, **8**, 959–973.

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C. & Kruuk, L.E.B. (2010) The misuse of BLUP in ecology and evolution. *The American Naturalist*, **175**, 116–125.

Heinze, J. & Tsuji, K. (1995) Ant reproductive strategies. *Researches on Population Ecology*, **37**, 135–149.

Hendry, A.P., Huber, S.K., De León, L.F., Herrel, A. & Podos, J. (2009) Disruptive selection in a bimodal population of Darwin's finches. *Proceedings of the Royal Society B*, **276**, 753–759.

Henshaw, J.M., Kahn, A.T. & Fritzsche, K. (2016) A rigorous comparison of sexual selection indexes via simulations of diverse mating systems. *Proceedings of the National Academy of Sciences, USA*, **113**, E300–E308.

Jones, A.G. (2009) On the opportunity for sexual selection, the Bateman gradient and the maximum intensity of sexual selection. *Evolution*, **63**, 1673–1684.

Kämpke, T. & Radermacher, F.J. (2015) *Income Modeling and Balancing*. Springer, Cham, Switzerland.

Kranstauber, B., Smolla, M. & Safi, K. (2016) Similarity in spatial utilization distributions measured by the Earth Mover's distance. *Methods in Ecology and Evolution*, DOI: 10.1111/2041-210X.12649.

Kruuk, L.E.B. (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **359**, 873–890.

Lande, R. (1982) A quantitative genetic theory of life history evolution. *Ecology*, **63**, 607–615.

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

Levina, E. & Bickel, P. (2001) The Earth Mover's distance is the Mallows distance: some insights from statistics. *Proceedings of the 8th IEEE International Conference on Computer Vision*, **2**, 251–256.

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

Moczek, A.P. & Emlen, D.J. (2000) Male horn dimorphism in the scarab beetle, *Onthophagus taurus*: do alternative reproductive tactics favour alternative phenotypes? *Animal Behaviour*, **59**, 459–466.

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

Morrissey, M.B., Kruuk, L.E.B. & Wilson, A.J. (2010) The danger of applying the breeder's equation in observational studies of natural populations. *Journal of Evolutionary Biology*, **23**, 2277–2288.

Morrissey, M.B. & Sakrejda, K. (2013) Unification of regression-based methods for the analysis of natural selection. *Evolution*, **67**, 2094–2100.

Nijhout, H.F. (2003) Development and evolution of adaptive polyphenisms. *Evolution & Development*, **5**, 9–18.

Phillips, P.C. & Arnold, S.J. (1989) Visualizing multivariate selection. *Evolution*, **43**, 1209–1222.

Price, G.R. (1970) Selection and covariance. *Nature*, **227**, 520–521.

Rausher, M.D. (1992) The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution*, **46**, 616–626.

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

Rubner, Y., Tomasi, C. & Guibas, L.J. (1998) A metric for distributions with applications to image databases. *Proceedings of the 6th IEEE International Conference on Computer Vision*, 59–66.

Rubner, Y., Tomasi, C. & Guibas, L.J. (2000) The Earth Mover's Distance as a metric for image retrieval. *International Journal of Computer Vision*, **40**, 99–121.

Rudnicki, R. & Zwoleński, P. (2015) Model of phenotypic evolution in hermaphroditic populations. *Journal of Mathematical Biology*, **70**, 1295–1321.

Rueffler, C., Van Dooren, T.J.M., Leimar, O. & Abrams, P.A. (2006) Disruptive selection and then what? *Trends in Ecology & Evolution*, **21**, 238–245.

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

Schluter, D. & Nychka, D. (1994) Exploring fitness surfaces. *The American Naturalist*, **143**, 597–616.

Shamoun-Baranes, J., van Loon, E.E., Purves, R.S., Speckmann, B., Weiskopf, D. & Camphuysen, C.J. (2012) Analysis and visualization of animal movement. *Biology Letters*, **8**, 6–9.

Shaw, R.G. & Geyer, C.J. (2010) Inferring fitness landscapes. *Evolution*, **64**, 2510–2520.

Smith, T.B. (1993) Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes*. *Nature*, **363**, 618–620.

Stinchcombe, J.R., Simonsen, A.K. & Blows, M.W. (2014) Estimating uncertainty in multivariate responses to selection. *Evolution*, **68**, 1188–1196.

Villani, C. (2003) *Topics in Optimal Transportation*. American Mathematical Society, Providence, RI, USA.

Villani, C. (2009) *Optimal Transport: Old and New*. Springer, Berlin, Germany.

Wade, M.J. & Arnold, S.J. (1980) The intensity of sexual selection in relation to male sexual behaviour, female choice, and sperm precedence. *Animal Behaviour*, **28**, 446–461.

Walsh, B. & Lynch, M. (2014) Theorems of natural selection: results of Price, Fisher, and Robertson. *Evolution and Selection of Quantitative Traits*. Advanced copy: http://nitro.biosci.arizona.edu/zbook/NewVolume_2/newvol2.html.

Weiner, J. & Solbrig, O.T. (1984) The meaning and measurement of size hierarchies in plant populations. *Oecologia*, **61**, 334–336.

Wilson, A.J., Réale, D., Clements, M.N., Morrissey, M.M., Postma, E., Walling, C.A., Kruuk, L.E.B. & Nussey, D.H. (2010) An ecologist's guide to the animal model. *Journal of Animal Ecology*, **79**, 13–26.

Wolf, J.I. & Seppä, P. (2016) Queen size dimorphism in social insects. *Insectes Sociaux*, **63**, 25–38.

Zwoleński, P. (2015) Trait evolution in two–sex populations. *Mathematical Modelling of Natural Phenomena*, **10**, 163–181.

Received 14 July 2016; accepted 10 October 2016
Handling Editor: Holger Schielzeth

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Proofs of properties of the DSD and maximizers.