



Symposium Article

# Analogues of the fundamental and secondary theorems of selection, assuming a log-normal distribution of expected fitness

Michael B. Morrissey<sup>1</sup> and Timothée Bonnet<sup>2,\*</sup>

<sup>1</sup>Dyers Brae House, School of Biology, University of St Andrews, St Andrews, Fife, United Kingdom, KY16 9TH

<sup>2</sup>Research School of Biology, 46 Sullivans Creek Road, The Australian National University, Acton ACT 2601, Australia

E-mail: [michael.morrissey@st-andrews.ac.uk](mailto:michael.morrissey@st-andrews.ac.uk) and [timotheebonnetc@gmail.com](mailto:timotheebonnetc@gmail.com)

Received July 26, 2018; First decision November 8, 2019; Accepted May 10, 2019.

Corresponding Editor: Anne Bronikowski

## Abstract

It is increasingly common for studies of evolution in natural populations to infer the quantitative genetic basis of fitness (e.g., the additive genetic variance for relative fitness), and of relationships between traits and fitness (e.g., the additive genetic covariance of traits with relative fitness). There is a certain amount of tension between the theory that justifies estimating these quantities, and methodological considerations relevant to their empirical estimation. In particular, the additive genetic variances and covariances involving relative fitness are justified by the fundamental and secondary theorems of selection, which pertain to relative fitness on the scale that it is expressed. However, naturally-occurring fitness distributions lend themselves to analysis with generalized linear mixed models (GLMMs), which conduct analysis on a different scale, typically on the scale of the logarithm of expected values, from which fitness is expressed. This note presents relations between evolutionary change in traits, and the rate of adaptation in fitness, and log quantitative genetic parameters of fitness, potentially reducing the discord between theoretical and methodological considerations to the operationalization of the secondary and fundamental theorems of selection.

**Subject area:** Quantitative genetics and Mendelian inheritance

**Key words:** fitness, fundamental theorem of selection, generalised linear mixed model, genetic variation, natural selection, secondary theorem of selection

## Introduction

The fundamental and secondary theorems of selection (Fisher 1930; Robertson 1966) define how mean additive genetic values for fitness and traits and change, within generations, as a result of selection. The fundamental theorem of natural selection (FTNS) states that

$$\Delta\bar{a}_w = \sigma_{a_w}^2, \quad (1)$$

where  $\Delta\bar{a}_w$  is the within-generation change in the genetic value for relative fitness, and  $\sigma_{a_w}^2$  is the variance of genetic values for relative fitness. In the absence of any change in the environment, the change

in mean fitness from one generation to the next would be  $\Delta\bar{a}_w$ . However, the environment generally does change, and in the case of relative fitness, it necessarily changes. The definition of a change in the environment intended by Fisher is extremely broad, and allele frequency – i.e., evolutionary change itself – can constitute a change in the environment (Ewens 1989; Price 1972; see Walsh and Morrissey 2019 for an overview of this concept). Genetic values,  $a$ , as opposed to genotypic values, for any quantity (e.g., phenotype, fitness), are defined to be the predictions from the multiple regression of phenotype on allelic dose at all loci. Genetic values thus refer to the total effects of alleles, not genotypes, on a character

(in this case fitness), as obtained from predictions of the multiple regression of the trait on genotype. The thought experiment in this definition of genetic values is one where genotypes are expressed numerically as the number of each type of allele present at each locus in each individual (Falconer (1960), chapter 7), and fitness is simultaneously regressed on the allelic compositions of individuals at every locus. Genetic effects as defined here are often called “additive” genetic effects or “breeding values”. The secondary theorem of selection (STS) states that

$$\Delta \bar{a}_z = \sigma_{a_z, w}, \quad (2)$$

where  $\Delta \bar{a}_z$  is the within generation change in genetic values for trait  $z$ , i.e., of  $a_z$ , and  $\sigma_{a_z, w}$  is the covariance of genetic values for the trait with relative fitness.

A brief aside about different versions of the STS may prevent ambiguity later. Alan Robertson originally (Robertson 1966) gave the secondary theorem as it is presented in equation 2, but subsequently (Robertson 1968) referred to the theorem as pertaining to the additive genetic covariance of phenotype and fitness, i.e.,  $\Delta \bar{a}_z = \sigma_{a_z, a_w}$ , i.e., the genetic covariance of genetic values for trait and fitness, rather than the covariance of genetic values for the trait with (phenotypic) fitness. The former expression, given in equation 2 is the true theorem; Walsh and Lynch (2018), chapter 6, describes this distinction with care. The two expressions are equivalent if environmental effects on the trait and fitness are independent of genetic effects. This may not be the case in practice, but assuming this independence may be useful in both theoretical and statistical models. Independence of genetic and environmental effects, and thus that  $\sigma_{a_z, w} = \sigma_{a_z, a_w}$ , will be assumed in this paper. The covariance of genetic values, or the additive genetic covariance, will generally be expressed with the conventional notation of  $\sigma_a(z, w)$ .

There is a certain amount of interest in estimating the quantities appearing in equations 1 and 2. These are fundamental quantities determining the rate of adaptation and phenotypic evolution. Also, it is sometimes useful to directly estimate  $\sigma_{a_w}^2$  and  $\sigma_{a_z, w}$  for the purpose of comparing these results to evolutionary predictions based on empirical estimates of phenotypic selection (Rausher 1992; Stinchcombe et al. 2002; Kruuk et al. 2003; Morrissey et al. 2010, 2012). The latter may be misleading if unmeasured quantities (either traits or environmental variables that may influence both trait values and individual fitness) cause covariance between traits and fitness. Mis-matches between direct application of the fundamental theorems of selection and predictions derived from separate inferences of phenotypic selection and genetic variation, i.e., using the breeder's or Lande equations (Lande 1979; Lush 1937), could be indicative of the presence of unmeasured quantities interfering with phenotypic selection analyses.

Whether for inferences of the genetic variance of relative fitness, and the genetic covariance of relative fitness with phenotype in their own right, or their inference for the purpose of comparisons with predictions of evolution made in other ways, empirical inference of  $\sigma_{a_w}^2$  and  $\sigma_{a_z, w}$  will be extremely challenging. It is likely that robust statistical models of fitness will require elaborations on basic mixed modelling techniques that are widely used for inference of quantitative genetic parameters in nature (Kruuk 2004; Wilson et al. 2010). In particular, genetic and environmental influences on any quantity that is constrained to take non-negative values, particularly fitness, are likely to be multiplicative, and so understanding how effects that may be additive on a logarithmic scale relate to evolutionary quantitative genetic theory would be very useful. Furthermore, statistical modelling techniques are now available (Generalized Linear Mixed

Model (GLMM); Bolker et al. 2009; Hadfield 2010) that allow inference to be conducted, not only on a logarithmic underlying scale, but also to apply distributions for variability that conform naturally to fitness data, for example, to directly accommodate fitness data in the form of counts. The purpose of this article is to determine what relations there might be between inferences of the genetic variance of fitness, and its genetic covariance with traits, on the log scale, with the rate of adaptation and phenotypic evolution.

## An analogue of the Fundamental Theorem of Natural Selection, for log-normal expected fitness

Throughout, we will mostly be concerned with properties of expected fitness. This notion may be most expediently clarified in relation to an empirical analysis. Consider variation among individuals in realized fitness, for example, of many fitness values such as  $W_i=0$  or  $W_i=10$  for hypothetical individuals that produced zero or ten offspring during their lives. Such variability might be described by a model of the form

$$x_i = \alpha + a_{x,i} + e_{x,i}, \quad (3a)$$

$$W_i \sim \mathcal{V}(\mu = e^{x_i}, \theta), \quad (3b)$$

where additive genetic values for fitness on the log scale,  $\alpha_x$ , might be assumed to be normally distributed according to  $\mathbf{a}_x \sim N(\mathbf{0}, \mathbf{A}\sigma_x^2)$ , where  $\mathbf{A}$  is the genetic relatedness matrix among individuals with breeding values in  $\mathbf{a}_x$ , and residuals are independently distributed according to  $e_{x,i} \sim N(0, \sigma_{e_x}^2)$ . This definition of additive genetic values corresponds to covariances among kin arising from additive genetic effects (Fisher 1918; Falconer 1960).  $\mathcal{V}()$  represents an arbitrary distribution with mean  $\mu$ , and possibly with additional parameters  $\theta$ ; for example,  $\mathcal{V}()$  might represent a Poisson distribution for realisations of fitness around each individual's expected fitness,  $e^{x_i}$ , considering both its genetic composition ( $a_{x,i}$ ) and any additional sources of variation that cause differences among individuals in their expected fitness. Expected fitness can be thought of as a latent quantity, that is essentially unobservable on individuals, but where it may nonetheless be possible to model heterogeneity among individuals. This model (equation 3), upon which results in this paper are based, coincides with a generalized linear animal model analysis (Hadfield 2010; Wilson et al. 2010).

Let  $x$  represent the logarithm of expected fitness as in equation 3. Mean expected absolute fitness, assuming expected fitness to be log-normally distributed, i.e., of  $E[W]_i = e^{x_i}$  when  $x \sim N(\mu_x, \sigma_x^2)$ , is

$$\bar{W} = \bar{E}[W] = \int e^x N(x, \mu_x, \sigma_x^2) dx = e^{\mu_x + \frac{\sigma_x^2}{2}}, \quad (4)$$

where  $N(x, \mu_x, \sigma_x^2)$  denotes the density of a normal distribution with mean  $\mu_x$  and variance  $\sigma_x^2$ , evaluated at  $x$ . A detailed derivation is given in the appendix, but one may note that this is merely the standard expression for the mean of a log-normal distribution (Aitchison and Brown 1957). Assume that an individual's value on the log-normal scale,  $x_p$ , is determined by additive genetic and other non-genetic factors, such that for individual  $i$ ,  $x_i = \mu_x + a_{x,i} + e_{x,i}$  (as in equation 3a), and that genetic effects are independent of other effects such that  $\sigma_x^2 = \sigma_{a_x}^2 + \sigma_{e_x}^2$ . Expected relative fitness, i.e.,  $\frac{E[W]_i}{\bar{W}}$  of an individual with genetic value  $a_x$  is thus

$$E[w|a_x] = \frac{e^{\mu_x + a_x + \frac{\sigma_x^2 - \sigma_{a_x}^2}{2}}}{e^{\mu_x + \frac{\sigma_x^2}{2}}} = e^{\mu_x + a_x + \frac{\sigma_x^2}{2} - \frac{\sigma_{a_x}^2}{2} - \mu_x - \frac{\sigma_x^2}{2}} = e^{a_x - \frac{\sigma_{a_x}^2}{2}}, \quad (5)$$

and so the covariance of  $a_x$  with relative fitness is

$$\sigma_{a_x, w} = \int a_x e^{a_x - \frac{\sigma_{a_x}^2}{2}} N(a_x, 0, \sigma_{a_x}^2) da_x = \sigma_{a_x}^2. \quad (6)$$

A detailed derivation is given in the appendix. This result is perfectly natural, as taking the natural logarithm of a quantity is closely analogous to re-expressing that value in units of its own mean. Thus, the within-generation change in genetic values for log expected fitness is the genetic variance for log fitness.

Equation 6 gives a direct analogue of the FTNS. However, our assumption of a log-linear relationship between  $x$  and fitness introduces a consistent sort of distortion of inheritance of fitness, such that even in a constant external environment (e.g., effects of population density, climate, etc.), the change in allele frequencies underlying fitness will themselves constitute a change in the environment (or rather of the environment, in the special way that it was conceived by Fisher), such that the actual change in mean fitness will be distorted. On the assumption that genetic values of log fitness, but not fitness on its original scale, are faithfully transmitted to the next generation (i.e., the genetic environment is constant on the log scale), then from the secondary theorem of selection, the change in mean expected log fitness between generations is simply  $\Delta \bar{x} = \Delta \bar{a}_x = \sigma_{a_x, w} = \sigma_{a_x}^2$ .

Under the assumptions already introduced, we can calculate the distortion resulting from the non-linear (logarithmic) relationship between  $x$  and fitness. The function  $x_i = \log(E[W]_i)$  can be thought of as a non-linear developmental function, and mechanics for predicting evolution in such a system can be used (Morrissey 2015). From equation 4, the mean expected value for fitness in the next generation is then

$$\bar{W}' = \bar{E}[W]' = e^{\mu_x + \sigma_{a_x}^2 + \frac{\sigma_x^2}{2}}.$$

From this, we can obtain an expression for the change in relative fitness from one generation to the next due to selection, expressed in relation to relative fitness in the first generation. Assuming that the distribution of effects on phenotype arising in the external environment (such as effects of weather and competitors) are unchanged, the change relative to fitness in the first generation is

$$\Delta \bar{w} = \frac{\bar{W}'}{\bar{W}} - \frac{\bar{W}}{\bar{W}} = \frac{e^{\mu_x + \sigma_{a_x}^2 + \frac{\sigma_x^2}{2}}}{e^{\mu_x + \frac{\sigma_x^2}{2}}} - 1 = e^{\sigma_{a_x}^2} - 1. \quad (7)$$

That the between-generation change in relative fitness due to the response to selection under a log-normal model of expected fitness is given by  $\Delta \bar{w} = e^{\sigma_{a_x}^2} - 1$  should not be regarded as a theorem. In other words, this relation should not be regarded as a very general result with virtually no assumptions as is the case for how the FTNS and STS predict the within-generation change in genetic value (Price 1972; Ewens 1989). Rather,  $\Delta \bar{w} = e^{\sigma_{a_x}^2} - 1$  is a theoretical result given some assumptions that may or may not be useful. Specifically, the result  $\Delta \bar{w} = e^{\sigma_{a_x}^2} - 1$  applies to the expected change in relative fitness due to evolution alone (i.e., changes in the non-genetic environment due, for example, to quantities such as weather and the

density of competitors, may generate a different change in phenotype), assuming a constant genetic environment (i.e., that the slope of the regression of phenotype – fitness in this case – on genotype does not change) on the scale of log fitness. This expression assumes that fitness is log-normal, and that genetic variation is additive on the log-normal scale, and genetic values on the log-normal scale are normally distributed. Furthermore, it applies either to a population under equilibrium between the effects of selection and recombination on  $\sigma_a^2$ , or it applies to the permanent change in  $\bar{W}$ , i.e., that which would occur with the restoration of linkage equilibrium under random mating. The expression  $\Delta \bar{a}_w = e^{\sigma_{a_x}^2} - 1$  makes no assumptions about the distribution of errors around  $E[W]$ , other than that they are independent of  $x$  and  $a_x$ .

Several miscellaneous comments on equations 6 and 7 may be useful. First, it is important to note that *realized* fitness is not assumed to be log-normal. Rather this assumption applies to expected fitness given a value on the log-normal scale. Note also that the total variance in expected fitness, and therefore non-additive components of variation in expected fitness, do not appear in equation 7; these are not assumed to be absent, rather, they can have any arbitrary (positive) variance. Finally, one should note while the change in relative fitness that follows from equation 7 is constant, if  $\sigma_{a_x}^2$  is constant, the change in absolute fitness is an accelerating function. Such an evolutionary dynamic is clearly unlikely to hold over more than a very few generations. This does not necessarily detract from the utility of this result for studying the dynamics of selection within generations, and of evolution from one generation to that which follows immediately.

Expected change in relative fitness, and the *genotypic* (as opposed to genetic) variance in relative fitness are, intriguingly, equal. The genotypic variance of fitness is the variance of expected fitness, or of expected fitness averaging over all possible environmental effects, given genetic value; for example, equation 5 gives the genotypic value for relative fitness, given genetic value. To show the relationship between evolution of fitness and the genotypic variance of relative fitness, it is simplest to begin with the genotypic value for absolute fitness. Individuals with additive genetic value for  $x$  of  $a_x$  will have expected fitness that is lognormal with (log) mean  $\mu_x + a_x$  and variance  $\sigma_{a_x}^2$ . Expected absolute fitness on the data scale is  $E[W|a_x] = e^{\mu_x + a_x + \frac{\sigma_x^2}{2}}$ , hence the distribution of genotypic values for expected fitness is also log normal, with mean  $\mu_x + \frac{\sigma_x^2}{2}$  and variance  $\frac{\sigma_x^2}{2}$ . The variance of a lognormal distribution, of an arbitrary random variable  $y$ , where  $\log(y) \sim (\mu_{\log(y)}, \sigma_{\log(y)}^2)$  is  $\sigma_y^2 = e^{2\mu_{\log(y)} + \sigma_{\log(y)}^2} (e^{\sigma_{\log(y)}^2} - 1)$  (Aitchison and Brown 1957), so the variance of genotypic values for expected absolute fitness is

$$\sigma_{E[W|a_x]}^2 = e^{2\left(\mu_x + \frac{\sigma_x^2}{2}\right) + \sigma_{a_x}^2} (e^{\sigma_{a_x}^2} - 1), \quad (8)$$

$$\sigma_{E[W|a_x]}^2 = \bar{W}^2 (e^{\sigma_{a_x}^2} - 1),$$

where the simplification uses the expression for mean fitness given in equation 4. Relative fitness is the linear transformation of relative fitness according to  $w = \frac{1}{\bar{W}} W$ , and so the genotypic variance of relative fitness is

$$\sigma_{E[w|a_x]}^2 = \left(\frac{1}{\bar{W}}\right)^2 \bar{W}^2 (e^{\sigma_{a_x}^2} - 1) = e^{\sigma_{a_x}^2} - 1, \quad (9)$$

which coincides with equation 7 for the total between-generation change in relative fitness due to the response to selection and associated segregation distortion.

The genotypic variances in expected fitness referred to in equations 8 and 9 refer only to that component of the genotypic variance for fitness on the expected scale, that arises from the strictly additive component of variance on the log scale, due to the nonlinear exponential relationship between the two scales. If there was non-additive genetic variance on the log scale, it would manifest as non-additive genetic variance (and potentially as additive genetic variance as well), on the expected fitness scale. Any such components of fitness would cause the genotypic variance of fitness to be some value other than that given by equation 9. Thus, it seems unlikely that this relation should be interpreted as giving the genotypic variance of fitness any simple and specific meaning with regards to the generation-to-generation change in mean fitness due to natural selection. Rather, it should probably be interpreted as a coincidental relation that is unique to a model of log-normal and additive fitness.

### An analogue of the Secondary Theorem of Selection, for a normal trait and log-normal expected fitness

Assume that a trait,  $z$  is influenced by additive genetic and environmental effects according to  $z_i = \mu_z + a_{z,i} + e_{z,i}$ . This model of phenotype is similar to that for log expected fitness in the previous section (equation 3a), and corresponds to a simple linear animal model analysis (Kruuk 2004; Wilson et al. 2010). Given the additive genetic (co)variance matrix of log expected fitness and a trait,

$$\Sigma_{\mathbf{a}} = \begin{bmatrix} \sigma_{a_x}^2 & \sigma_{a_x, a_z} \\ \sigma_{a_x, a_z} & \sigma_{a_z}^2 \end{bmatrix},$$

the expected genetic value of log fitness ( $a_x$ ) given genetic value for the trait ( $a_z$ ) is

$$E[a_x|a_z] = a_z \left( \frac{\sigma_{a_x, a_z}}{\sigma_{a_z}^2} \right),$$

where  $\frac{\sigma_{a_x, a_z}}{\sigma_{a_z}^2}$  is the genetic regression of log fitness on phenotype. The variance of genetic values for  $x$ , given genetic value for  $z$  is the conditional genetic variance of log fitness, given phenotype (see Walsh and Blows 2009 for a review of the mathematical concept of conditional genetic variance),

$$\sigma^2(a_x|a_z) = \sigma_{a_x}^2 - \sigma_{a_z}^2 \left( \frac{\sigma_{a_x, a_z}}{\sigma_{a_z}^2} \right)^2 = \sigma_{a_x}^2 - \frac{(\sigma_{a_x, a_z})^2}{\sigma_{a_z}^2}.$$

Relative fitness given  $a_z$  is then given by the expression for the mean of a log normal distribution, with location parameter  $\mu_x + E[a_x|a_z]$  and dispersion parameter representing all log-scale phenotypic variance that is independent of genetic variance in the trait, so

$$E[w|a_z] = \frac{E[W|a_z]}{\bar{W}} = \frac{e^{\mu_x + a_z \frac{\sigma_{a_x, a_z}}{\sigma_{a_z}^2} + \frac{\sigma_x^2 - \frac{(\sigma_{a_x, a_z})^2}{\sigma_{a_z}^2}}{2}}}{e^{\mu_x + \frac{\sigma_x^2}{2}}} = e^{\frac{2a_z \sigma_{a_x, a_z} - (\sigma_{a_x, a_z})^2}{2\sigma_{a_z}^2}}.$$

The evolutionary change in the mean of  $z$  can then be obtained from the secondary theorem of selection, because the assumption of additivity in  $z$  guarantees that the genetic environment is constant for  $z$ . Thus, assuming that environmental effects are independent of genetic effects on  $z$  and  $x$ ,

$$\begin{aligned} \sigma_{a_z, w} &= E[a_z E[w|a_z]] - E[a_z] E[w] = E[a_z E[w|a_z]] \\ \sigma_{a_z, w} &= \int a_z e^{\frac{2a_z \sigma_{a_x, a_z} - (\sigma_{a_x, a_z})^2}{2\sigma_{a_z}^2}} N(a_z, \mu_x, \sigma_x^2) da_z \\ \sigma_{a_z, w} &= \sigma_{a_x, a_z}, \end{aligned} \tag{10}$$

noting that  $\bar{a}_z = 0$  by construction, and from equation 2, it follows that

$$\Delta \bar{a}_z = \sigma_{a_x, a_z}. \tag{11}$$

So, the additive genetic covariance with relative fitness, and the additive genetic covariance with log expected absolute fitness are the same thing when expected fitness is log normal. The expression  $\Delta \bar{z} = \sigma_{a_x, a_z}$  requires that genetic values for the trait and for log expected fitness are bivariate normal, but does not require equilibrium between the effects of selection and recombination on the genetic variance of  $z$ .

The analogue of the STS given in equation 11 is not subject to any segregation distortion under the assumptions used here, because we assume a linear relationship between  $a_z$  and  $z$ . Of course, a net phenotypic change given by this expression may be not realized even when the genetic assumptions used here hold, if the external environment changes.

### An analogue of the Secondary Theorem of Selection, for a log-normal trait and log-normal expected fitness

Many traits are necessarily positive (e.g., morphology, expected rates of physiological processes or rates or expected counts of behavioural events, etc.), and consequently they should probably most pragmatically be modeled as additive on a logarithmic scale. The within-generation changes in such traits on the log scale can be modeled using equation 10. However, the realized change will be subject to distortion. Given the covariance of the logarithms of a trait and expected fitness

$$\Sigma_{\mathbf{a}} = \begin{bmatrix} \sigma_{a_x}^2 & \sigma_{a_x, a_{\log(z)}} \\ \sigma_{a_x, a_{\log(z)}} & \sigma_{a_{\log(z)}}^2 \end{bmatrix},$$

then the evolution of  $z$  between generations is simply given by

$$\begin{aligned} \Delta E[e^{\log(z)}] &= e^{\mu_{\log(z)} + \sigma_{a_x, a_{\log(z)}} + \frac{\sigma_{\log(z)}^2}{2}} - e^{\mu_{\log(z)} + \frac{\sigma_{\log(z)}^2}{2}}, \\ \Delta E[z] &= e^{\mu_{\log(z)} + \frac{\sigma_{\log(z)}^2}{2}} (e^{\sigma_{a_x, a_{\log(z)}}} - 1). \end{aligned} \tag{12}$$

So, the analogue to the FTNS given in equation 7 acts as a scaling factor for the proportional change in any log-normal traits that may covary genetically with fitness on a log-log scale. This result applies at equilibrium between the effects of selection and recombination on the genetic variance of the log-normal trait; changes in the linkage disequilibrium contribution to  $\sigma_{a_{\log(z)}}$  due



to selection will generate additional, transient, responses to selection.

### Consistency metrics under the analogues of the secondary theorem

The course of adaptive evolution is not normally predicted using the FTNS and/or the STS. Empirical operationalization of the ideas underlying these theorems requires estimation of some of the most challenging parameters in evolutionary quantitative genetics, genetic variances, and covariances pertaining to fitness. Estimates of these quantities are notoriously imprecise. Furthermore, measures of how phenotypes affect fitness – which are not components of the STS or FTNS, as they say nothing about how genetic variances and covariances pertaining to fitness arise – can be much more informative about the ecology of natural selection. The main mechanisms for relating more ecologically informative measures of phenotypic selection to genetic variation to predict adaptive evolution is the breeder's equation,

$$\Delta\bar{z} = \mathbf{GP}^{-1}\mathbf{S}, \quad (13a)$$

$$\Delta\bar{z} = \mathbf{G}\beta, \quad (13b)$$

these are thoroughly reviewed in chapter 13 of Walsh and Lynch (2018). The first expression is typically referred to as the multivariate breeder's equation, and the second is often referred to as the Lande equation (that formulation arose in Lande 1979).  $\mathbf{G}$  and  $\mathbf{P}$  are matrices of (additive) genetic and phenotypic variances and correlations.  $\mathbf{S}$  in equation 13a is the phenotypic selection differential, which may be defined equivalently as the change in the mean phenotype vector, within generations (Lush 1937), due to selection, or by the covariance of traits with relative fitness (see Robertson 1966; Price 1970; Walsh and Morrissey 2019) attempts to explain the equivalence to the former definition as simply as possible).  $\beta$ , termed the selection gradient vector, gives the partial regressions of relative fitness on phenotypes, which is equivalent to the derivative of relative fitness with respect to phenotype, averaged over the distribution of phenotype.

The equivalence of equation 13 parts a and b follows from the general relation for multiple regression,  $\beta = \Sigma_x^{-1}\Sigma_{xy}$ . Where  $\beta$  are the partial regressions of  $y$  on the vector of predictor variables  $\mathbf{x}$ ,  $\Sigma_x$  is the covariance matrix of  $\mathbf{x}$ , and  $\Sigma_{xy}$  is the vector of covariances of  $\mathbf{x}$  and  $y$ . This same relation, applied to genetic components of phenotypic variance and covariance with fitness, is the basis of testing for consistency of the breeder's equation with the STS. Application of the theorems of selection does not require referencers to phenotypic values of traits, and so any traits, or quantities in the environment that may vary among individuals and cause trait-fitness covariance, that are not measured will not bias predictions of evolution. However, application of the theorems of selection provides no information about the ecology of why a given evolutionary trajectory does (or does not) occur. Therefore, a empirical motivation for using the theorems of selection, particularly the STS, is to check that predictions based on the more informative breeder's equation (equation 13, or further elaborations thereupon, e.g., Lande 1980, 1982; Morrissey 2014, 2015) do not diverge from predictions of the STS, which is robust to missing traits.

Several measures of consistency between the breeder's equation and the STS have been proposed (Rauscher 1992; Morrissey et al. 2010) to directly use parameters related to the STS to test of the signatures of missing variables in phenotypic selection analysis (reviewed in Walsh and Lynch 2018). This section considers consistency metrics as they relate to log-normal fitness distributions, and the STS, as in the previous two sections.

The consistency metric given in Morrissey et al. (2010) is based on the phenotypic, genetic, and environmental (residual, in a statistical model) regressions of fitness on phenotype. The breeder's equation is consistent with the STS when these regressions are equal. The necessary relations for application of this metric can be recovered for an analysis using log fitness by setting the STS (for multiple traits, i.e.,  $\Delta\bar{\mathbf{a}}_z = \sigma_{z,w}$ ) and the multivariate breeder's equation  $\Delta\bar{\mathbf{z}} = \mathbf{GP}^{-1}\mathbf{S}$  to be equal. Subsequent manipulations then serve to illuminate the conditions under which they are in fact equal, or in other words, the conditions under which the predictions of the breeder's equation are valid:

$$\sigma_a(\mathbf{z}, w) = \mathbf{GP}^{-1}\mathbf{S}, \quad (14a)$$

$$\mathbf{G}^{-1}\sigma_a(\mathbf{z}, w) = \mathbf{P}^{-1}\mathbf{S}, \quad (14b)$$

$$\beta_a = \beta_p. \quad (14c)$$

In equation 14a,  $\beta_a$  and  $\beta_p$  are the additive genetic and phenotypic partial regressions of relative fitness on the traits.  $\beta_a$  cannot be estimated directly from data, since the predictor variables – genetic values for the traits – cannot be directly observed, but mixed model analysis can be employed to estimate these (Morrissey et al. 2010, 2012). From equation 11 it is apparent that the consistency metrics for a normal trait and log-normal fitness are the same expressions, but with  $\log(E[W])$  substituted for  $w$ . Since evolution of a trait is given by  $\mathbf{G}\beta$  when all relevant variables are included in a selection analysis, consistency requires equality of the phenotypic regression of fitness on traits and the genetic regression of expected absolute fitness on traits. What may not be so immediately obvious is that the phenotypic regression may be expressed equivalently as either as the regression of relative fitness on traits, or as the regression of log absolute fitness on traits:

$$\begin{aligned} \beta_i &= \frac{\int \frac{\partial e^{a+\sum_k b_k z_k}}{\partial z_i} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}}{\int e^{a+\sum_k b_k z_k} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}} \\ \beta_i &= \frac{\int b_i e^{a+\sum_k b_k z_k} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}}{\int e^{a+\sum_k b_k z_k} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}} \\ \beta_i &= \frac{b_i \int e^{a+\sum_k b_k z_k} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}}{\int e^{a+\sum_k b_k z_k} N(\bar{\mathbf{z}}, \mathbf{P}) d\mathbf{z}} \\ \beta_i &= b_i \end{aligned} \quad (15)$$

see also (Morrissey and Goudie 2016) for more results pertaining to selection gradients under fitness functions with an exponential form.

The same consistency metric also applies to the analogue of the STS for a log-normal trait. The phenotypic selection analysis that corresponds to the second STS analogue is one applied to

log-transformed phenotypic data. As such, it is expected that the change in genetic values on the log scale is given by the Lande equation and the STS in either its original form, or its analogue for a normal trait and log-normal fitness. So, from eq. 12, but substituting **Gb** for  $\sigma_{a_z, a_x}$ , and setting the STS analogue and evolutionary predictions from the phenotypic analysis to be equal shows,

$$\begin{aligned} \Delta \bar{E}[e^z] &= e^{\mu_z + \frac{\sigma_z^2}{2}} (e^{G\beta} - 1) \\ e^{\mu_z + \frac{\sigma_z^2}{2}} (e^{\sigma_{a_z, a_x}} - 1) &= e^{\mu_z + \frac{\sigma_z^2}{2}} (e^{G\beta} - 1) \quad (16) \\ \sigma_{a_z, a_x} &= G\beta \\ G^{-1} \sigma_{a_z, a_x} &= \beta. \end{aligned}$$

So from equations 15 and 16, the consistency measure equivalent to equation 14a, but for an analysis involving genetic parameters of log fitness, and either a log or identity link function for traits, is that the genetic regression of log fitness on traits is the same as the phenotypic regression of log fitness on traits.

### Discussion

The relationships given here between the genetic variance of log fitness, and genetic covariances of traits with log fitness, are immediately amenable to statistical models that may be useful for empirical analyses of fitness in nature. Previous attempts to estimate quantitative genetic parameters relating to (relative) fitness have used models that assume Gaussian distributions for all characters, including fitness. This has been necessary, despite the fact that fitness residuals will almost always be extremely non-normal, because the biological questions at hand relate to relative fitness on the scale that it is observed. It has not been clear that conducting inferences on some other scale that might allow more sophisticated treatment of the distribution of fitness could still speak to the relevant biological questions. However, the expressions given in this note show that fitness can in fact be directly analysed on the log scale, and its genetic variance, and genetic covariance with traits, on that scale, has direct relationships to evolutionary theory.

Any GLMM that treats a measure of fitness with a log-link function can generate estimates of  $\sigma_a^2(x)$  and  $\sigma_a(z, x)$  (and for log-normal traits, of  $\sigma_a(\log(z), x)$ ). The key component of any such model is that a log-link function is used for fitness (and for a trait, if a log-normal model of the trait is desired). Any distribution of realized fitness around expected values could be useful. This will include the Poisson distribution and overdispersed Poisson distributions, both with additive and multiplicative (negative binomial) overdispersion. Other distributions that could potentially be accommodated with a log link function could include exponential and geometric distributions.

It will be very rare that a standard statistical distribution will very closely fit the distribution of errors of fitness - particularly of measures of lifetime fitness. It may be reasonable to hope that by using distributions for overdispersed count variables, such as the Poisson distribution with additive overdispersion, or the negative binomial distribution, will lead to more robust analyses than can be achieved with models that assume Gaussian distributions, but with parameters that directly reflect key quantities in the fundamental and secondary theorems of selection. It should be noted that there is no guarantee that GLMM-based analysis, since it will rarely closely reflect the processes that generate variation in fitness, will generate robust results. Similarly, it will not necessarily follow that any deficiencies of model fit will necessarily seriously hinder inference of quantities such as  $\sigma_a^2$  and  $\sigma_a(z, x)$ . Any empirical applications making use the formulae given here should be accompanied by checks of the

ability of the fitted model to recover basic parameters of the distribution of observed fitness.

### Acknowledgements

MBM was supported by a University Research Fellowship from the Royal Society (London). This note benefitted greatly from my discussions with Andy Gardner and Loeske Kruuk.

### 1. Appendix

Intermediate steps in obtaining equation 4 are

$$\begin{aligned} \bar{E}[W] &= \int e^x N(x, \mu_x, \sigma_x^2) dx \\ &= \int e^x \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} dx \\ &= \int \frac{e^{x - \frac{(x-\mu_x)^2}{2\sigma_x^2}}}{\sqrt{2\pi\sigma_x}} dx \\ &= \left[ -\frac{1}{2} e^{\mu_x + \frac{\sigma_x^2}{2}} \text{ERF} \left[ \frac{-x + \mu_x + \sigma_x^2}{\sqrt{2}\sigma_x} \right] + C \right]_{x=-\infty}^{x=\infty} \\ &= \left( -\frac{1}{2} e^{\mu_x + \frac{\sigma_x^2}{2}} (-1) \right) - \left( -\frac{1}{2} e^{\mu_x + \frac{\sigma_x^2}{2}} (+1) \right) \\ &= e^{\mu_x + \frac{\sigma_x^2}{2}}. \end{aligned}$$

The fourth step relies on the fact that  $\lim_{x \rightarrow \infty} \text{ERF} \left[ \frac{-x + \mu_x + \sigma_x^2}{\sqrt{2}\sigma_x} \right] = -1$  and  $\lim_{x \rightarrow -\infty} \text{ERF} \left[ \frac{-x + \mu_x + \sigma_x^2}{\sqrt{2}\sigma_x} \right] = +1$ , where *ERF* denotes the error function.

Similarly, intermediate steps in obtaining equation 6 are

$$\begin{aligned} \sigma_{a_z, w} &= \int a_x e^{a_x - \frac{\sigma_{a_x}^2}{2}} N(a_x, 0, \sigma_{a_x}^2) da_x \\ &= \int a_x e^{a_x - \frac{\sigma_{a_x}^2}{2}} \frac{1}{\sqrt{2\pi\sigma_{a_x}^2}} e^{-\frac{a_x^2}{2\sigma_{a_x}^2}} da_x \\ &= \int \frac{a_x e^{a_x - \frac{\sigma_{a_x}^2}{2} - \frac{a_x^2}{2\sigma_{a_x}^2}}}{\sqrt{2\pi\sigma_{a_x}^2}} da_x \\ &= \left[ \frac{1}{2} \sigma_{a_x}^2 \text{ERF} \left[ \frac{a_x - \sigma_{a_x}^2}{\sqrt{2}\sigma_{a_x}} \right] - \frac{e^{-\frac{(a_x - \sigma_{a_x}^2)^2}{2\sigma_{a_x}^2}}}{\sqrt{2\pi}} + C \right]_{a_x=-\infty}^{a_x=\infty} \\ &= \left( \frac{\sigma_{a_x}^2}{2} (+1) - \frac{e^{-\frac{(-\infty - \sigma_{a_x}^2)^2}{2\sigma_{a_x}^2}}}{\sqrt{2\pi}} \right) - \left( \frac{\sigma_{a_x}^2}{2} (-1) - \frac{e^{-\frac{(-\infty - \sigma_{a_x}^2)^2}{2\sigma_{a_x}^2}}}{\sqrt{2\pi}} \right) \\ &= \sigma_{a_x}^2. \end{aligned}$$

As for the detailed derivation of equation 4, the instances of the error function again take their maximum and minimum values of +1 and -1, respectively. Additionally, the final step relies on noting that  $\lim_{y \rightarrow \infty} e^{-\frac{(y - \sigma_{a_x}^2)^2}{2\sigma_{a_x}^2}} = 0$  and  $\lim_{y \rightarrow -\infty} e^{-\frac{(y - \sigma_{a_x}^2)^2}{2\sigma_{a_x}^2}} = 0$ . A detailed derivation of equation 11 is nearly identical.

### References

Aitchison J, Brown JAC. 1957. *The Lognormal Distribution, with special reference to it uses in economics*. Cambridge: Cambridge University Press.  
 Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MH, White JS. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol*. 24:127-135.

- Ewens WJ. 1989. An interpretation and proof of the fundamental theorem of natural selection. *Theor Popul Biol.* 36:167–180.
- Falconer DS. 1960. *Introduction to Quantitative Genetics*. Edinburgh: Oliver and Boyd.
- Fisher RA. 1918. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh.* 52:399–433.
- Fisher RA. 1930. *The genetical theory of natural Selection*. Oxford: Clarendon Press.
- Hadfield J. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software.* 33:1–22.
- Kruuk LE. 2004. Estimating genetic parameters in natural populations using the “animal model”. *Philos Trans R Soc Lond B Biol Sci.* 359:873–890.
- Kruuk LEB, Merilä J, Sheldon BC. 2003. When environmental variation short-circuits natural selection. *Trends Ecol Evol.* 18:207–209.
- Lande R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution.* 33:402–416.
- Lande R. 1980. Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution.* 34:292–305.
- Lande R. 1982. A quantitative genetic theory of life history evolution. *Ecology.* 63:607–615.
- Lush JL. 1937. *Animal breeding plans*. Ames (Iowa): Iowa State College Press.
- Morrissey MB. 2014. Selection and evolution of causally covarying traits. *Evolution.* 68:1748–1761.
- Morrissey MB. 2015. Evolutionary quantitative genetics of nonlinear developmental systems. *Evolution.* 69:2050–2066.
- Morrissey MB, Goudie I. 2016. Analytical results for directional and quadratic selection gradients for log-linear models of fitness functions. *bioRxiv.* 10.1101/040618.
- Morrissey MB, Kruuk LE, Wilson AJ. 2010. The danger of applying the breeder’s equation in observational studies of natural populations. *J Evol Biol.* 23:2277–2288.
- Morrissey MB, Parker DJ, Korsten P, Pemberton JM, Kruuk LE, Wilson AJ. 2012. The prediction of adaptive evolution: empirical application of the secondary theorem of selection and comparison to the breeder’s equation. *Evolution.* 66:2399–2410.
- Price GR. 1970. Selection and covariance. *Nature.* 227:520–521.
- Price GR. 1972. Fisher’s ‘fundamental theorem’ made clear. *Ann Hum Genet.* 36:129–140.
- Rausher MD. 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.
- Robertson A. 1968. *Population biology and evolution, chap. The spectrum of genetic variation*. New York: Syracuse University Press. Pp. 516.
- Stinchcombe JR, Rutter MT, Burdick DS, Tiffin P, Rausher MD, Mauricio R. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Am Nat.* 160:511–523.
- Walsh B, Blows MW. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annu Rev Ecol Evol Syst.* 40:41–59.
- Walsh B, Lynch M. 2018. *Evolution and selection of quantitative traits*. Oxford (UK): Oxford University Press.
- Walsh B, Morrissey MB. 2019. Evolutionary Quantitative genetics, pp. 421–454. In: *Handbook of Statistical Genetics*, Wiley, New York.
- Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LE, Nussey DH. 2010. An ecologist’s guide to the animal model. *J Anim Ecol.* 79:13–26.