

# THE PREDICTION OF ADAPTIVE EVOLUTION: EMPIRICAL APPLICATION OF THE SECONDARY THEOREM OF SELECTION AND COMPARISON TO THE BREEDER'S EQUATION

Michael B. Morrissey,<sup>1,2</sup> Darren J. Parker,<sup>1,3</sup> Peter Korsten,<sup>1</sup> Josephine M. Pemberton,<sup>1</sup> Loeske E.B. Kruuk,<sup>1</sup> and Alastair J. Wilson<sup>1</sup>

<sup>1</sup>Institute of Evolutionary Biology, King's Buildings, University of Edinburgh, Edinburgh, Midlothian, UK EH9 2LD

<sup>2</sup>E-mail: michael.morrissey@st-andrews.ac.uk

<sup>3</sup>Current Address: School of Biology, University of St Andrews

Received March 30, 2011

Accepted January 6, 2012

Data Archived: Dryad doi:10.5061/dryad.td6tm4f6

Adaptive evolution occurs when fitness covaries with genetic merit for a trait (or traits). The breeder's equation (BE), in both its univariate and multivariate forms, allows us to predict this process by combining estimates of selection on phenotype with estimates of genetic (co)variation. However, predictions are only valid if all factors causal for trait-fitness covariance are measured. Although this requirement will rarely (if ever) be met in practice, it can be avoided by applying Robertson's secondary theorem of selection (STS). The STS predicts evolution by directly estimating the genetic basis of trait-fitness covariation without any explicit model of selection. Here we apply the BE and STS to four morphological traits measured in Soay sheep (*Ovis aries*) from St. Kilda. Despite apparently positive selection on heritable size traits, sheep are not getting larger. However, although the BE predicts increasing size, the STS does not, which is a discrepancy that suggests unmeasured factors are upwardly biasing our estimates of selection on phenotype. We suggest this is likely to be a general issue, and that wider application of the STS could offer at least a partial resolution to the common discrepancy between naive expectations and observed trait dynamics in natural populations.

**KEY WORDS:** Natural selection, fitness, microevolution, breeder's equation, secondary theorem of selection, Robertson–Price identity, genetic covariance, Soay sheep, *Ovis aries*.

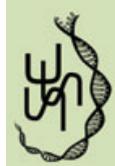
In the absence of genetic constraint, evolutionary change is the expected outcome of natural selection, which can itself be defined as occurring when phenotypic variation causes variation in fitness (Endler 1986). Thus, adaptive evolution will generally occur if heritable traits have causative effects on fitness, although for any single trait our expectations for change are also contingent on patterns of genetic covariance with other traits under selection. Conceptually, if a heritable trait causes fitness variation, then breeding values for that trait (i.e., the influence of an individual's genes relative to population trait mean; Falconer 1981) become breeding values for fitness. Consequently, the phenotypic distribution of the trait will change within a generation, and the

resultant change in the distribution of the breeding values will transmit phenotypic change to future generations.

Quantitative genetic theory provides us with two key models for predicting the evolution of phenotypic traits under selection. The first, and most commonly used, is the breeder's equation (BE) (Lush 1937), which in its univariate form (UVBE) is

$$\Delta \bar{z} = h^2 S = \frac{\sigma_a^2(z)}{\sigma_p^2(z)} \sigma_p(z, w), \quad (1)$$

where  $\Delta \bar{z}$  is the predicted change in trait mean phenotype after one generation,  $h^2$  is the heritability of the trait (defined as the ratio of additive genetic  $\sigma_a^2(z)$  to total phenotypic  $\sigma_p^2(z)$  variance), and  $S$



is the selection differential, defined equivalently as the change in mean phenotype within a generation or as the phenotypic covariance of the trait with relative fitness  $\sigma_p(z, w)$ . Less well known, at least to empiricists, is a second model known alternatively as the Robertson–Price identity or Robertson’s secondary theorem of selection (STS) (Robertson 1966, 1968; Price 1970). This states that the evolutionary change is equal to the genetic covariance of a trait with relative fitness such that:

$$\Delta \bar{z} = \sigma_G(z, w). \quad (2)$$

These two equations follow directly from the definition of adaptive evolution given above, and are equivalent when the causative effect of the trait on fitness is solely responsible for the covariance between the trait and fitness (Queller 1992; Morrissey et al. 2010). In other words this is when the statistical relationship between the trait and fitness, estimated as  $S$  in the BE, arises completely from the fact that the trait variation causes fitness variation. If this is not the case and additional factors influence the covariance between the trait and fitness, a multivariate version of the BE (MVBE; Lande 1979) is required in which:

$$\Delta \bar{z} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S}, \quad (3)$$

where  $\bar{z}$  is a vector of population mean phenotypic values,  $\mathbf{G}$  and  $\mathbf{P}$  are the genetic and phenotypic variance–covariance matrices, and  $\mathbf{S}$  is a vector of selection differentials. This multivariate form of the model is predictive, but requires the important assumption that all the factors (which may be additional traits or environmental effects such as food availability or temperature) contributing to the trait–fitness covariance are known, measured, and included (Morrissey et al. 2010). Thus, the key assumption of the BE—in both univariate and multivariate forms—is that the trait–fitness covariance is caused by the trait(s) in the model and that there are no “missing traits” (Queller 1992; Hadfield 2008; Morrissey et al. 2010). Under this condition the shift in breeding values, and consequently in population mean phenotypes must follow according only to patterns of genetic variation and covariation captured in  $\mathbf{G}$ .

The STS, on the other hand, requires no assumption about causation to be made (Morrissey et al. 2010). Instead trait evolution is predicted by directly estimating the genetic component of the association between a trait of interest and relative fitness. In fact, despite its name, the STS need not actually be about selection on the focal trait at all. For instance, covariance of focal trait breeding values and relative fitness could arise from selection on a genetically correlated trait (whether measured or not). As such the secondary theorem is arguably more about genetics than it is about selection, especially because the genetic covariance of a trait with fitness can actually be nonzero in the complete absence

of selection (e.g., in the case of evolutionary change caused by drift). The important point is that the secondary theorem provides a direct prediction of evolutionary change. However, this prediction, for better or worse, is not generated in conjunction with any insight as to the true form of selection.

We have recently proposed that the STS might provide a more robust predictor of evolutionary change than the BE in studies of natural populations (Morrissey et al. 2010). The core argument for this is that the genetics, physiology, behavior, and ecology of phenotypic variation and fitness variation are so complex that we are unlikely to identify, much less meaningfully measure and adequately model, all of the factors contributing to trait–fitness covariation. Consequently, our estimates of selection acting on individual traits are often likely to be biased (Rauscher 1992; Stinchcombe et al. 2002; Kruuk et al. 2003), in which case predictions from the STS should be more robust than those obtained from the BE. If this view is correct, application of the STS may offer a resolution to the problem that while phenotypic traits frequently appear to be selected (Kingsolver et al. 2001; Kingsolver and Pfennig 2004) and heritable (Roff and Mousseau 1987), there is a paucity of evidence for evolutionary responses in intensively studied populations (Merilä et al. 2001), and indeed phenotypic trends in trait values are often counter to expectations based on the direction of selection (Gienapp et al. 2008).

Here we test this proposal with an analysis of morphological trait variation in Soay sheep, *Ovis aries*, from the island of Hirta, St. Kilda, Scotland. Many morphological traits in this population have a partial genetic basis of variation (e.g., Robinson et al. 2006; Wilson et al. 2006), and also covary phenotypically with fitness such that they are subject to apparent natural selection (e.g., Preston et al. 2003; Milner et al. 2004). Body size in particular is heritable (Milner et al. 2000; Wilson et al. 2007), and is positively related to survival (Coltman et al. 1999b; Milner et al. 2004) and reproductive success (Coltman et al. 1999a; Preston et al. 2003). However, the body size of Soay sheep has actually declined during the course of the intensive ongoing individual-based study (Wilson et al. 2007; Ozgul et al. 2009). This decline has occurred in part because of within-generation changes in phenotypic distributions of body size, possibly influenced by a partial relaxation of viability selection (Ozgul et al. 2009). At the genetic level, it has proven very difficult to determine whether mean genetic merit of the population has also changed. Wilson et al. (2007) reported a small upward trend in breeding values, but see Hadfield et al. (2010) for a criticism of the methods by which statistical uncertainty was evaluated (the trend is nonsignificant and small in magnitude). Consequently, this system provides an ideal opportunity to test whether empirical application of the STS will (1) produce different predictions of adaptive phenotypic evolution than the BE and (2) whether predictions of the STS are more consistent with the general pattern of stasis in selected phenotypic

traits that predominates in data from long-term individual-based field studies (Merilä et al. 2001). We apply the animal model (Henderson 1973; Kruuk 2004) to estimate genetic parameters associated with both the BE and the STS. Specifically, this allows direct estimation of parameters such as heritabilities, covariances of traits with relative fitness, that is, selection differentials, and the genetic covariances of traits with relative fitness. Consequently, we can evaluate both BE-based predictions of evolutionary change and predictions based on the STS in a common framework. This allows unbiased estimation of key parameters in equations (1), (2), and (3), including and in addition to implementation of statistical tests advocated by Rausher (1992).

## Methods

We applied linear mixed effect models, specifically various implementations of the animal model (Henderson 1973; Kruuk 2004), to estimate the parameters of the UVBE, MVBE, and the STS for four morphological traits in Soay sheep. All models were fitted by restricted maximum likelihood using ASREML (Gilmour et al. 2002). We first summarize the methodology relating to data collection and the estimation of pedigree before describing the quantitative genetic modeling in full detail.

### DATA COLLECTION AND ESTIMATION OF PEDIGREE AND FITNESS

The Soay sheep population inhabiting Village Bay on the island of Hirta, St. Kilda, has been the subject of intensive, individual-based study since 1985. Each year, extensive censusing and field work is conducted during which the majority of the lambs born in the study area are caught, individually tagged, and tissue samples are obtained to allow the determination of paternity by molecular methods (described below). Each August, a large proportion of the study population is captured and phenotyped for multiple traits, so that multiple measurements may be available for an individual across different years of its life. Here we focus our analysis on weight (in kg, number of observations,  $n_o = 4337$ , number of individuals,  $n_i = 1829$ ); length of the hind leg, measured as metatarsus length (in mm,  $n_o = 4618$ ,  $n_i = 2206$ ); horn length (in mm,  $n_o = 1496$ ,  $n_i = 878$ ); and scrotal circumference (in mm,  $n_o = 322$ ,  $n_i = 114$ ). Note that while many females have horns we treat the latter two traits as sex-limited (i.e., consider horns in males only) and furthermore limit our analysis of horn-length to those males with the “normal” morph (i.e., large, strong, curled horns that do not break during fights; Clutton-Brock et al. 1997).

The pedigree information required to parameterize the quantitative genetic models was obtained through a combination of observation (4373 maternities assigned), and molecular paternity

assignment using microsatellite and allozyme marker data analyzed with the R package MASTERBAYES (Hadfield et al. 2006). Horn type and linear and quadratic terms of age were included as additional predictors of paternity in an approach similar to that recently described by Walling et al. (2010). Briefly, we simultaneously analyzed paternity of all lambs born between 1985 and 2009, using the molecular data as previously described (e.g., Overall et al. 2005), but with recent cohorts having been genotyped at a core set of 18 microsatellite loci, and also using unique vectors of candidate sires constructed for each year based on the total set of males known to be alive. MASTERBAYES estimates the pedigree and the effect of the phenotypic predictors of paternity jointly, providing posterior distributions for both. As expected from previous reports (Pemberton et al. 1999; Robinson et al. 2006), age (both linear and quadratic terms) and horn type are highly significant predictors of paternity (i.e., the posterior distributions of associated coefficients do not overlap zero). We categorically assigned paternity of all lambs for which a particular candidate sire was specified in at least 80% of the samples of the posterior distribution of the pedigree. Our pedigree analysis thus generated 2253 assignments of paternity. The mean individual-level posterior support for these assignments is approximately 98% (see Walling et al. 2010 for detailed discussion of the important distinction between individual- and pedigree-wide level statistical confidence in parentage assignments); this statistic is comparable to the assignment thresholds implemented in the program CERVUS (Marshall et al. 1998; Kalinowski et al. 2007), which has previously been used for pedigree estimation in this and many other such study systems.

We calculated lifetime breeding success (LBS) as our metric of fitness for both males and females. This quantity was defined as the number of newborn lambs attributed to an individual, as determined either by observation for females or by paternity assignment (as described) for males. LBS was only calculated for individuals born in the study area between 1985 and 2002, but with LBS calculated from lambs born up until and including 2010. This was to reduce bias in estimated fitness of individuals that only visit the study area occasionally, and to reduce censoring bias in those cohorts from which many individuals are still alive. Relative fitness  $w$  was calculated by dividing LBS for each individual by the appropriate sex-specific mean. Note that the true mean LBS must be equal in the two sexes and is known to be slightly greater than 2 because the population has increased in size over the course of the study. However, as life-history data are not complete for all individuals, especially males, this treatment should provide the closest approximation of true relative fitness. In total, estimates of LBS (and thus  $w$ ) are available for 1107 females and 1007 males, including 392 females and 114 males with nonzero estimated LBS.

### UNIVARIATE ANALYSES OF GENETIC VARIATION IN FITNESS AND PHENOTYPE

We first tested for genetic variation in relative fitness, because in the absence of genetic variance for fitness, genetic covariances of traits with fitness are undefined. We fit the univariate animal model

$$\mathbf{w} = \boldsymbol{\mu} + \mathbf{Z}_1\mathbf{n} + \mathbf{Z}_2\mathbf{a} + \mathbf{Z}_3\mathbf{m} + \mathbf{Ie}, \quad (4)$$

where  $\boldsymbol{\mu}$ , the population mean, is the only fixed effect, and where  $\mathbf{Z}_1$ ,  $\mathbf{Z}_2$ , and  $\mathbf{Z}_3$  are design matrices relating observations to the random effects of natal year  $\mathbf{n}$ , breeding value  $\mathbf{a}$ , maternal identity  $\mathbf{m}$ , respectively.  $\mathbf{I}$  is an identity matrix and  $\mathbf{e}$  is a vector of residual errors. All random effects (and residuals) are assumed to be normal. Elements of  $\mathbf{a}$  are further assumed to be drawn from  $\mathbf{a} \sim N(\mathbf{0}, \mathbf{G} \otimes \mathbf{A})$ , where  $\mathbf{G}$  is the additive genetic variance covariance matrix (or rather the additive genetic variance in the univariate analyses) and  $\mathbf{A}$  is the pedigree-derived additive genetic relatedness matrix. For brevity, we refer to  $\mathbf{G}$  and its elements as “genetic,” and other components of variation as “environmental,” although strictly,  $\mathbf{G}$  is expected to represent additive genetic variation, and nonheritable genetic variation, such as dominance and epistatic (co)variances are included in other modeled terms, such as permanent environment effects (see below) and residual terms. We tested the significance of each random effect by comparing the likelihood of this model to one in which the appropriate random effect was omitted. This was done using likelihood ratio tests assuming that the test statistic  $2\Delta \ln L$  is  $\chi^2$  distributed with one degree of freedom.

We also applied univariate animal models to partition phenotypic variation in each of the four morphometric traits among potential causal sources of variation, and particularly, to estimate additive genetic variances. For each adult (age one or older) morphometric trait, we then fitted the model

$$\mathbf{z} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{n} + \mathbf{Z}_2\mathbf{y} + \mathbf{Z}_3\mathbf{pe} + \mathbf{Z}_4\mathbf{a} + \mathbf{Z}_5\mathbf{m} + \mathbf{Ie}, \quad (5)$$

where  $\mathbf{X}$  is a design matrix relating individual observations to the fixed effects  $\mathbf{b}$ , which consisted of the population means and multilevel effects of age. Fixed effects included the population mean and age as a multilevel factor (a separate level for each age class in years). For weight and hind leg length only, additional fixed effects of sex and a sex-by-age interaction were included. Random effects were as for equation (4), but with the addition of measurement year  $\mathbf{y}$  and a permanent environment effect  $\mathbf{pe}$  to account for the nongenetic component of individual-level repeatability. Statistical significance of random effects was assessed by likelihood ratio tests as described above.

### ANALYSES OF SELECTION AND PREDICTIONS FROM THE UVBE AND STS

We used bivariate animal models to estimate the key parameters for both the UVBE (the selection differentials, or phenotypic covariances between focal traits and relative fitness) and the STS (the genetic covariances between trait and fitness). In each of these models, relative fitness and one of the phenotypic traits were treated together as dependent variables

$$[\mathbf{w}, \mathbf{z}] = \mathbf{Xb} + \mathbf{Z}_1\mathbf{n} + \mathbf{Z}_2\mathbf{y} + \mathbf{Z}_3\mathbf{pe} + \mathbf{Z}_4\mathbf{a} + \mathbf{Z}_5\mathbf{m} + \mathbf{Ie}. \quad (6)$$

This model is very similar in construction to that specified by equation (5), but the vectors of random effects are replaced by matrices describing variation in both dependent variables. Other than the mean, no fixed effects were fit for  $w$ , whereas the year and residual variance components for  $w$  were constrained to be zero. Consequently, residual variance in relative fitness is represented in  $\mathbf{pe}$ , allowing estimation of the covariance between relative fitness and the nongenetic component of individual-level repeatability (i.e., permanent environment component) of  $\mathbf{z}$ . This is not to imply that we can somehow calculate a metric of the repeatable component of variation for a trait, relative fitness, that is measured only once. Rather, the biologically interesting, nongenetic aspect of the relationship between phenotype and relative fitness is between the residuals of fitness and the repeatable component of trait variation, and this constraint in the mixed model renders this relationship estimable. Critically, in this bivariate animal model with fitness and phenotypic traits treated as dependent variables, the additive genetic effects are assumed to be drawn from the distribution

$$[a_i(w), a_i(z)] \sim N \left( 0, \begin{bmatrix} \sigma_a^2(w)\mathbf{A}, & \sigma_a(w, z)\mathbf{A} \\ \sigma_a(w, z)\mathbf{A}, & \sigma_a^2(z)\mathbf{A} \end{bmatrix} \right). \quad (7)$$

This definition of  $\mathbf{a}$  is simply an explicit bivariate description of the general definition of  $\mathbf{a}$  following equation (4), above. We provide this second definition to highlight how the genetic covariance of a trait and relative fitness (i.e.,  $\sigma_a(w, z)$ ) can be obtained directly from the solution of the mixed model equations for an animal model.

We then calculated the phenotypic variance of each trait based only on those modeled effects that are associated with the individual, that is,  $\sigma_p^2(z) = \sigma_{pe}^2(z) + \sigma_a^2(z) + \sigma_r^2(z)$ . We calculated the heritability of each trait as the quotient of  $\sigma_a^2(z)$  and this metric of the phenotypic variance. The corresponding selection differential for each trait was then estimated as the phenotypic covariance between trait and fitness,  $\sigma_p(w, z) = \sigma_{pe}(w, z) + \sigma_a(w, z)$ , and predicted responses from the UVBE determined as  $h^2(z)\sigma_p(w, z)$ . The procedure by which standard errors of these predictions were generated (and standard errors of the MVBE, see below) is provided in the appendix.

To test whether the conditions under which the UVBE is predictive hold, we applied the test recently advocated by Morrissey et al. (2010; based on equations presented in Hadfield 2008 and Queller 1992). Specifically we compared, for each trait, the likelihood of the model specified by equation (6) to one in which we constrained the model such that  $b_a(w, z) = b_{pe}(w, z)$ , where  $b$  are regression coefficients of  $w$  on  $z$  (assuming the test statistic  $2\Delta \ln L$ , is  $\chi^2$  distributed with one degree of freedom). When the assumptions of the BE are met, that is, the trait in question is the sole cause of its covariance with relative fitness, the genetic and phenotypic regressions of relative fitness on phenotype will be equal. Thus, this test treats the case where the conditions under which the BE is predictive of evolutionary change as the null hypothesis. This is a quantitative implementation of the type of analysis suggested by Rausher (1992), directly testing for differences between the relationship between phenotype and fitness at the phenotypic and genetic levels. This test is unbiased by assumptions made when working with predicted breeding values (Postma 2006; Hadfield et al. 2010), which have been problematic in the tests reported to date. This constraint can be imposed using ASREML by specifying the variance structures with the CHOL command (ASREML user guide release 2, p. 124), in conjunction with cross-covariance structure constraints (ASREML user guide release 2, p. 137). Hence, if the unconstrained model is justifiably better, the BE is shown not to be predictive. We similarly fitted constrained models where either  $\sigma_a(z, w)$  alone, or both  $\sigma_a(z, w)$  and  $\sigma_{pe}(z, w)$ , were constrained to values of zero to assess the statistical significance of the genetic and phenotypic (or among-individual) covariances of each trait with relative fitness. The former is a direct test of the predicted evolutionary change under the STS (against a null hypothesis of zero change), whereas the latter tests the significance of the selection differential used to parameterize the UVBE.

### MULTIVARIATE ANALYSIS OF SELECTION AND PREDICTION OF EVOLUTION BY THE MVBE

To apply the MVBE, we used a multivariate animal model to determine the additive genetic variance–covariance structure among all four morphometric traits ( $\mathbf{G}$ ). We fitted a model structure equivalent to that specified in equation (5), but where  $\mathbf{z}$  is a matrix containing all records from all years for all individuals for all traits. To obtain estimates of the form of multivariate selection on the morphometric traits for which we have repeated measures, we also fitted a mixed model equivalent to that specified by equation (6), where  $\mathbf{z}$  again contains records of all traits, in addition to records of estimated relative fitness. However, in this case we did not include the additive genetic random effect because our goal was to estimate the vector of phenotypic selection differentials (to parameterize the MVBE); because the “permanent environment” effect for each individual therefore represents all differences, ge-

**Table 1.** Univariate animal model-based genetic analysis of relative fitness in Soay sheep (*Ovis aries*) on St. Kilda. The maternal effect was bounded at zero, and therefore we do not report its parameters; maternal effects on relative fitness were excluded from subsequent analyses.  $\sigma^2$  is the variance of each component, and  $\frac{\sigma^2}{\sigma_p^2}$  is the proportion of the total phenotypic variance attributable to each component. Statistical significances were not calculated for the proportional metrics (i.e., heritability, etc.) nor for the residual or total phenotypic variance.

Component	$\sigma^2 \pm 1 \text{ SE}$	$\frac{\sigma^2}{\sigma_p^2} \pm 1 \text{ SE}$
Birth year	3.27 ± 1.32***	0.0549 ± 0.0211
Genetic	1.55 ± 0.868**	0.0259 ± 0.0145
Maternal	–	–
Residual	54.9 ± 1.89	0.919 ± 0.0255
total phenotypic variance	59.7 ± 2.17	–

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

netic or otherwise, between individuals, we refer to it as  $\mathbf{id}$  in this model, to distinguish it from  $\mathbf{pe}$ , above. We defined the entries in  $\mathbf{id}$  relating to covariances between relative fitness and the repeatable component of phenotypic variation in each of the traits as  $\mathbf{S}$  (the vector of selection differentials). We defined the repeatable individual (i.e., arising from genetic and permanent environment effects combined) and residual variance–covariance structures for the morphometric traits as  $\mathbf{id}_{\text{morph}}$  and  $\mathbf{e}_{\text{morph}}$ , respectively. We were then able to estimate the vector of selection gradients as

$$\beta = (\mathbf{id}_{\text{morph}} + \mathbf{e}_{\text{morph}})^{-1} \mathbf{S}, \quad (8)$$

which is conceptually equivalent to a standard multiple regression analysis, but appropriate when repeated measures are available for the traits of interest. We include  $\mathbf{e}_{\text{morph}}$  in equation (8) for comparability to the situation in which a single measurement of the morphometric traits had been made, in which case the phenotypic variance would be the combination of the within- and among-individual variances, which we are able to separate in the current context. Having thus estimated  $\mathbf{G}$  and  $\beta$ , we obtained multivariate BE-based predictions,  $\Delta \bar{\mathbf{z}}_{MVBE}$ , of evolutionary change for the morphological traits by applying the MVBE (in a version called the “Lande equation”; Lande 1979, 1982; Lande and Arnold 1983)

$$\Delta \bar{\mathbf{z}} = \mathbf{G}\beta. \quad (9)$$

## Results

We detected significant (additive) genetic variance for LBS (scaled in the analysis as relative fitness), our measure of individual fitness in the Soay sheep (Table 1). Birth year, fitted as a random effect, is also significant source of variance, and was therefore included for all subsequent analyses in which fitness was modeled. We did not detect a significant maternal effect on

**Table 2.** Univariate animal model-based genetic analysis of adult morphometric traits in Soay sheep (*Ovis aries*) on St. Kilda. Variance component ( $\sigma^2$ ) subscripts are as follows: *P*, phenotypic variance conditional on the fixed effects; *id*, the total variance associated with individuals; *pe*, the nongenetic component of individual variance; *a*, the genetic component of the individual variance; *by*, birth year; *y*, measurement year; *m*, mother, and *r*, the residual variance. Statistical significances of the residual variance, and the compound variance terms (i.e., *P* and *id*) were not calculated.

Trait	$\sigma_p^2 \pm 1$ SE	$\sigma_{id}^2 \pm 1$ SE	$\sigma_{pe}^2 \pm 1$ SE	$\sigma_a^2 \pm 1$ SE
Weight	29.9 $\pm$ 2.00	9.65 $\pm$ 0.651	8.10 $\pm$ 0.7***	1.50 $\pm$ 0.70**
Leg length	1.3 $\cdot$ 10 <sup>2</sup> $\pm$ 10.6	67.9 $\pm$ 3.2	49.3 $\pm$ 3.7***	18.60 $\pm$ 4.2***
Horn length	4.6 $\cdot$ 10 <sup>3</sup> $\pm$ 5.2 $\cdot$ 10 <sup>3</sup>	2.1 $\cdot$ 10 <sup>3</sup> $\pm$ 1.8 $\cdot$ 10 <sup>2</sup>	1.4 $\cdot$ 10 <sup>3</sup> $\pm$ 2.2 $\cdot$ 10 <sup>2</sup> ***	6.6 $\cdot$ 10 <sup>2</sup> $\pm$ 2.6 $\cdot$ 10 <sup>2</sup> **
Scrotal circ	6.7 $\cdot$ 10 <sup>2</sup> $\pm$ 64.8	1.7 $\cdot$ 10 <sup>2</sup> $\pm$ 31.8	2.3 $\cdot$ 10 <sup>-4</sup> $\pm$ 1.5 $\cdot$ 10 <sup>-5</sup>	1.7 $\cdot$ 10 <sup>2</sup> $\pm$ 31.8***
Trait	$\sigma_{bv}^2 \pm 1$ SE	$\sigma_v^2 \pm 1$ SE	$\sigma_m^2 \pm 1$ SE	$\sigma_r^2 \pm 1$ SE
Weight	4.0 $\pm$ 1.3***	4.6 $\pm$ 1.5***	1.5 $\pm$ 0.4***	10.2 $\pm$ 0.3
Leg length	32.0 $\pm$ 10.2***	0.8 $\pm$ 0.3***	8.4 $\pm$ 2.1***	19.9 $\pm$ 0.6
Horn length	1.4 $\cdot$ 10 <sup>3</sup> $\pm$ 5.0 $\cdot$ 10 <sup>2</sup> ***	2.1 $\cdot$ 10 <sup>2</sup> $\pm$ 78.0***	3.4 $\cdot$ 10 <sup>2</sup> $\pm$ 1.4 $\cdot$ 10 <sup>2</sup> **	5.9 $\cdot$ 10 <sup>2</sup> $\pm$ 33.9
Scrotal circ	21.6 $\pm$ 17.5*	1.3 $\cdot$ 10 <sup>2</sup> $\pm$ 57.1***	3.8 $\cdot$ 10 <sup>-5</sup> $\pm$ 2.6 $\cdot$ 10 <sup>-6</sup>	3.5 $\cdot$ 10 <sup>2</sup> $\pm$ 24.0

\* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ .

LBS, and indeed the maternal variance estimate was constrained at the lower limit (i.e., very close to zero) when estimates were constrained to positive parameter space. On this basis we chose to simplify model fitting by excluding maternal effects on fitness in subsequent analyses. We also detected significant additive genetic variance for the four morphological traits tested: adult weight, leg length, horn length, and scrotal circumference (Table 2). The other fitted random effects were statistically significant sources of variance for all traits, except for scrotal circumference, where both the nongenetic component of individual repeatability (permanent environment effect) and the maternal effects were nonsignificant and estimated to be very small.

For all four adult morphometric traits, we found a positive association between the repeatable component of phenotype and fitness (Table 3). This is shown in the estimated selection differentials which are highly significant for weight, hind leg length, and horn length. Note that the nongenetic component of repeatability was very small for scrotal circumference (see above), and so the estimate of selection on this trait arises almost entirely from the genetic component of trait-fitness covariance which was found to be positive, but not statistically significant (Table 3). Consequently, the UVBE predicts positive evolutionary change in all traits (Figure 1), although the predictions of change in leg length and scrotal circumference are very small (<2 mm/generation), and the estimate for scrotal circumference has a very large sampling error relative to the magnitude of the estimate. Estimates of per-generation change  $\pm$  1 SE are as follows: weight, +0.57  $\pm$  0.25 kg; hind leg length, +1.58  $\pm$  0.53 mm; horn length, +12.22  $\pm$  5.33 mm; scrotal circumference, +1.59  $\pm$  1.33 mm.

The multivariate genetic variance-covariance structure is predominantly characterized by positive relationships among the four adult sheep traits (Table 4). The only exception is a modest

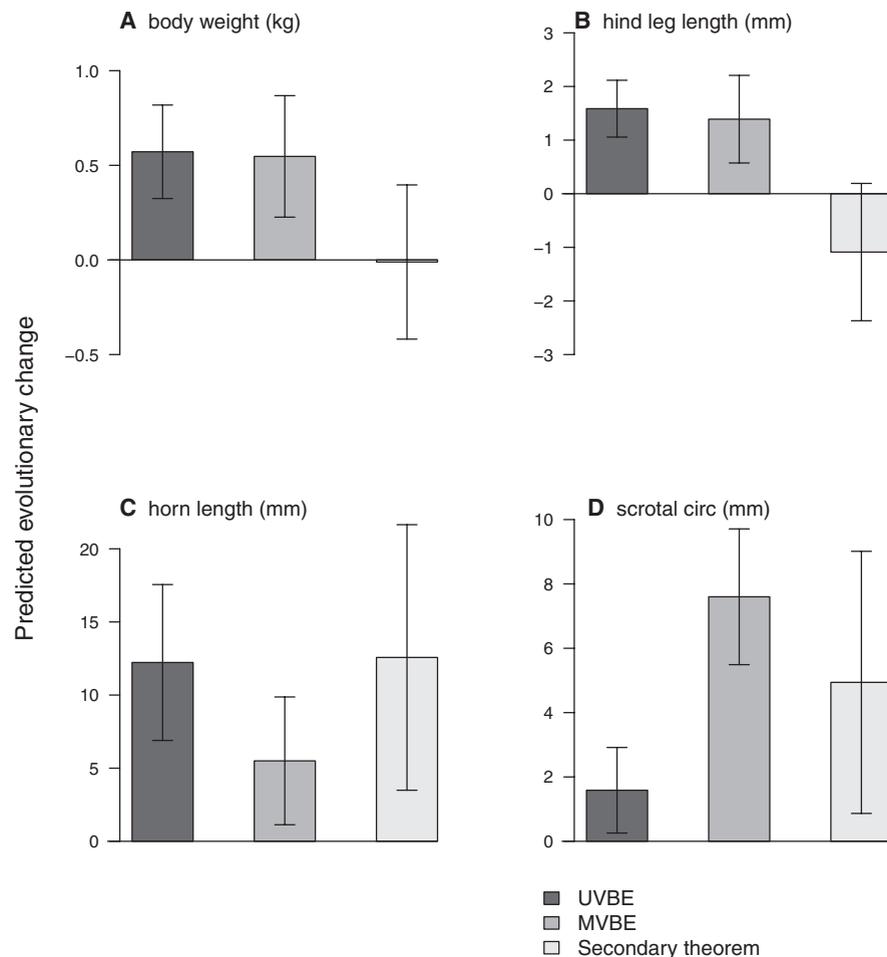
negative genetic correlation between leg length and horn length, which is not statistically significant as inferred from the overlap of its standard error (i.e., approximate 50% confidence interval) with zero. Although nearly all phenotypic and genetic correlations among traits are positive, the relationships are not so strong that we should reasonably consider the traits, as measured, to be biologically equivalent aspects of the phenotype. For example, the phenotypic and genetic correlations between weight and leg length are only 0.53 and 0.50, respectively, suggesting that these two measures of size capture at least partially distinct aspects of individual phenotype, and should be considered as genetically distinct traits (i.e.,  $r_G$  does not equal +1). An exception to this lies in the estimated genetic correlation between weight and scrotal circumference, which is actually estimated at slightly greater than the upper limit of +1. (Note that while correlations in excess of one are not biologically interpretable as such, convergence of the multivariate animal model was only possible with **G** modeled as a completely unstructured and unconstrained matrix.)

The vector  $\beta$  contains the partial regression coefficients, or selection gradients, for the (phenotypic) regression of relative fitness on the repeatable component of individual phenotype, which are positive for all traits except scrotal circumference. Based on approximately doubling the standard errors, these selection gradients are statistically significant for weight and horn length only (Table 5). The MVBE-based predictions of evolutionary change are positive for all traits (Table 5, Fig. 1). Despite the negative selection gradient for scrotal circumference, positive evolutionary change is expected for this trait, in large part as a consequence of its strong genetic covariance with body weight (Table 4), which is apparently subject to strong positive selection as estimated here.

In contrast to the predictions of both the univariate and multivariate applications of the BEs, the predictions based on the

**Table 3.** Bivariate animal model-based estimates of individual-level covariation between morphometric traits and relative fitness in Soay sheep (*Ovis aries*) on St. Kilda. The phenotypic covariance between each trait and relative fitness,  $\sigma_a(w, z) + \sigma_{pe}(w, z)$ , is the selection differential, and the additive genetic covariance of each trait with relative fitness,  $\sigma_a(w, z)$ , is the prediction of evolutionary change based on the secondary theorem of selection. The two-tailed probability values associated with the test of the null hypothesis of the equivalence of the regressions  $b$  of relative fitness of the genetic and nongenetic but repeatable components of phenotypic variance,  $\sigma_a(w, z)$  and  $\sigma_{pe}(w, z)$  are presented in the last column. This test pertains to whether the predictions of the univariate breeder's equation and the secondary theorem of selection are consistent with one another. The nongenetic component of the repeatability of scrotal circumference is fixed at the boundary (zero), and so we do not report it or  $P$  values from statistical tests that would involve this parameter.

Trait	$\sigma_a(w, z) + \sigma_{pe}(w, z) \pm 1 \text{ SE}$	$P(\sigma_a(w, z) + \sigma_{pe}(w, z) = 0)$	
Weight	7.52 ± 0.73	$< 10^{-7}$	
Hind leg length	7.48 ± 1.94	$3.40 \cdot 10^{-7}$	
Horn length	49.5 ± 11.4	$1.09 \cdot 10^{-7}$	
Scrotal circ	–	–	
Trait	$\sigma_a(w, z) \pm 1 \text{ S.E.}$	$P(\sigma_a(w, z) = 0)$	$P(b_a(w, z) = b_{pe}(w, z))$
Weight	–0.011 ± 0.407	0.157	0.0483
Hind leg length	–1.09 ± 1.28	0.291	$6.393 \cdot 10^{-4}$
Horn length	12.57 ± 9.08	0.074	1.000
scrotal circ	4.94 ± 4.07	0.121	–



**Figure 1.** Predictions of phenotypic evolution of adult morphometric traits in Soay sheep (*Ovis aries*) on St. Kilda using the univariate breeder's equation (UVBE), the multivariate breeder's equation (MVBE), and the empirical application of the secondary theorem of natural selection. Error bars show standard errors.

**Table 4.** Multivariate animal model-based genetic analysis of individual-associated variance components for adult morphometric traits in Soay sheep (*Ovis aries*) on St. Kilda. Values on the diagonal are variances, covariances are reported below the diagonals and correlations are reported above the diagonals. All values are reported  $\pm 1$  SE. Traits are (left to right and top to bottom) weight, hind leg length, (male) length of normal morph horns, and (male) scrotal circumference.

(a) Phenotypic, conditional on all nonindividual associated effects			
17.3 $\pm$ 0.5	0.53 $\pm$ 0.016	0.50 $\pm$ 0.027	0.56 $\pm$ 0.030
20.3 $\pm$ 1.0	85.0 $\pm$ 2.99	0.35 $\pm$ 0.033	0.29 $\pm$ 0.042
100.3 $\pm$ 7.2	158.4 $\pm$ 16.89	2350.9 $\pm$ 141.2	0.2828 $\pm$ 0.043
52.2 $\pm$ 4.0	61.4 $\pm$ 9.64	310.4 $\pm$ 50.81	512.5 $\pm$ 31.26
(b) Permanent environment			
7.7 $\pm$ 0.6	0.86 $\pm$ 0.03	0.70 $\pm$ 0.07	0.90 $\pm$ 0.83
16.1 $\pm$ 1.2	45.8 $\pm$ 3.48	0.64 $\pm$ 0.07	0.69 $\pm$ 0.71
67.6 $\pm$ 8.1	151.8 $\pm$ 19.9	1226.8 $\pm$ 174.2	0.19 $\pm$ 0.42
10.8 $\pm$ 4.2	20.2 $\pm$ 10.5	29.0 $\pm$ 62.8	18.7 $\pm$ 41.7
(c) Additive genetic			
1.2 $\pm$ 0.5	0.50 $\pm$ 0.15	0.33 $\pm$ 0.24	1.13 $\pm$ 0.19
2.4 $\pm$ 1.2	19.1 $\pm$ 3.9	-0.18 $\pm$ 0.20	0.54 $\pm$ 0.15
9.2 $\pm$ 7.6	-19.0 $\pm$ 20.5	612.8 $\pm$ 202.4	0.54 $\pm$ 0.20
16.2 $\pm$ 4.0	30.7 $\pm$ 10.4	174.2 $\pm$ 69.4	167.3 $\pm$ 49.0
(d) Residual			
8.4 $\pm$ 0.2	0.14 $\pm$ 0.02	0.36 $\pm$ 0.03	0.48 $\pm$ 0.04
1.8 $\pm$ 0.3	20.1 $\pm$ 0.6	0.25 $\pm$ 0.03	0.13 $\pm$ 0.05
23.5 $\pm$ 2.2	25.5 $\pm$ 3.8	511.3 $\pm$ 28.4	0.26 $\pm$ 0.05
25.2 $\pm$ 2.3	10.6 $\pm$ 4.2	107.2 $\pm$ 21.2	326.5 $\pm$ 21.8

**Table 5.** Selection gradients ( $\beta$ ) and multivariate breeder's equation-based predictions of the evolution ( $\Delta\bar{z}$ ) of adult morphometric traits in Soay sheep (*Ovis aries*) on St. Kilda.

Trait	Units	$\beta$ (units/ $\sigma^2$ ) $\pm 1$ SE	$\Delta\bar{z}$ (units) $\pm 1$ SE
Weight	kg	0.436 $\pm$ 0.060	0.547 $\pm$ 0.321
Hind leg length	mm	0.018 $\pm$ 0.022	1.39 $\pm$ 0.817
Horn length	mm	8.70.10 <sup>-3</sup> $\pm$ 2.90.10 <sup>-3</sup>	5.50 $\pm$ 4.37
Scrotal circ	mm	-9.55.10 <sup>-3</sup> $\pm$ 5.88.10 <sup>-3</sup>	7.60 $\pm$ 2.11

application of the secondary theorem of natural selection are actually negative for weight and for hind leg length, although the estimate is very small in magnitude for weight. Furthermore, the genetic and nongenetic regressions of relative fitness on the repeatable components of phenotypic variation for weight and hind leg length are significantly different (Table 3), with the genetic

regressions being smaller than the nongenetic regressions. This result indicates that, for these two traits, the conditions under which the UVBE should be predictive (and equivalent to the STS) do not hold.

## Discussion

We detected covariance of morphometric traits with fitness, additive genetic variance for those traits, and also additive genetic variance for fitness in Soay sheep on St. Kilda. Thus, a number of parameters have values that are consistent with potential for evolutionary change. However, the point estimates of evolutionary change as predicted by the BE, in both its univariate and multivariate forms, and by the STS, are qualitatively different for body weight and leg length. Furthermore, for these two body size traits, the genetic and nongenetic regressions of fitness on phenotype are statistically significantly different, and as such we can demonstrate that the UVBE is not quantitatively predictive. This is because these regressions will differ when the assumption of causality that is required in the UVBE does not hold. In contrast, predictions of evolutionary change based on the alternate approaches are similar for horn size and scrotal circumference. Thus, we have provided an empirical demonstration of the potential for predictions of the BE to differ from those of the STS.

With respect to the evolution of body size, the qualitative contrast between evolutionary predictions based on the BE and the STS is quite stark. The former predicts that sheep should increase in size, while stasis, or indeed evolution of smaller size, is predicted by the latter approach (Fig. 1A, B). This result suggests that the positive (phenotypic) associations between body size and fitness do not arise, at least in totality, from a causal effect of size on fitness. Consequently it seems likely that estimates of positive selection on body size are upwardly biased. This conclusion may provide some resolution to the observation that stasis or counterintuitive phenotypic trends are often found for body size traits (Merilä et al. 2001), in spite of apparently widespread directional selection and heritable variation (e.g., Réale et al. 2003; Kingsolver and Pfennig 2004).

Importantly, although the application of the STS indicates that we should not expect evolution of size, it does not tell us why. Specifically, it provides no answers to the questions of (1) what the true form of natural selection is (with respect to size and/or other traits of interest)? or (2) what genetic constraints may be acting to generate the lack of predicted evolution? It may be that body size is not selected, but rather that an unmeasured trait or aspect of the environment is a causative agent of both fitness variation and of trait (size) variation, but that this is independent of any effect of size on fitness. In this scenario, size is heritable and covaries with fitness, but there is in fact no selection (the scenario primarily advocated by Rausher 1992 and Kruuk

et al. 2003). Alternatively, size may be positively selected, that is, have a causative effect on fitness, but genetically correlated traits may be antagonistically selected, a situation consistent with the notion that different aspects of the way body size is related to fitness variation should generate trade-offs (Blackenhorn 2000). In both scenarios, which are not mutually exclusive, the model of selection, genetics, and ultimately evolutionary change provided in the BE is inadequate because of unmeasured quantities (traits, or environmental variation at the level of different experiences of individuals).

The evolution of horn length and scrotal circumference have not been studied in detail in Soay sheep. Here, the predictions of evolutionary change are more consistently positive across the three models, although we note all predictions are associated with substantial statistical uncertainty. Detailed further analyses, similar to those that have been conducted for body size (Wilson et al. 2007; Ozgul et al. 2009), will have to be conducted for these traits before we can determine whether these evolutionary predictions are consistent with the temporal dynamics of these characters. The current analysis indicates, as is the case for these two traits, that the evolutionary predictions of the UVBE and the STS can be consistent.

The multivariate formulation of the BE is generally presented as an improvement on the univariate form in the sense that it is biologically unrealistic to assume that natural selection acts on single traits in isolation. Although the “missing trait” problem still persists (i.e., one can always posit the existence of an additional factor that should have been included), it seems intuitive that multivariate models must be preferable to univariate models in the sense that at least there will be fewer missing factors. However, there is also a hidden danger in the application of the MVBE. The danger arises from the fact that an inadequate model of selection for one trait will not only result in erroneous evolutionary predictions for that trait (as it would in the application of the UVBE), but that error may also be propagated to predictions for genetically correlated traits included in the analysis. Thus, it is entirely possible that UVBE-based evolutionary predictions may be more accurate than those of the MVBE. Although of course we do not know the “truth” here, it is notable that the predictions of evolutionary change for horn length are similar from UVBE and STS but less so from MVBE. This discrepancy arises from the negative genetic correlation of horn length with leg length, and our model of selection of leg length that we suspect may be inadequate, as discussed above. Similarly, the apparent over-prediction of evolutionary change of scrotal circumference by the MVBE, relative to the prediction based on the STE, may arise from an inadequate model of selection of body mass, combined with the strong genetic correlation of body mass and this trait.

It is of course somewhat unsatisfying to argue that (1) the predictions of the BE and the STS are qualitatively different, and

(2) that the STS does not predict evolution of body size, when there are clearly very substantial uncertainties associated with all quantitative predictions (Fig. 1). Nevertheless, further dissection of the heritable and nonheritable components of the relationships between the traits and fitness appears useful in supporting our contention. The test for the inequality of the genetic and non-genetic regressions of fitness on the phenotypic traits provides us with more statistically justifiable grounds for the argument that the UVBE and the STS predictions differ for weight and leg length (Table 3). We hope that this test will prove a particularly useful addition to the set of tools available for studying the genetics, selection, and evolution of traits in natural populations.

### ANALYTICAL CONSIDERATIONS

Empirical application of the STS is in its infancy. The application of mixed models where both relative fitness and phenotypic traits are treated as dependent variables is a fairly novel methodological approach in this respect (but see Etterson and Shaw 2001; Kruuk et al. 2002; and Morrissey and Ferguson 2011), and to our knowledge, this is the first such application to calculate the phenotypic covariance of a repeated measures trait with relative fitness to apply the BE. A number of reports exist from studies where fitness has been treated as a dependent variable in a mixed model to estimate the genetic variance of fitness (e.g., Price and Schluter 1991; Merilä and Sheldon 2000; Coltman et al. 2005; Teplitsky et al. 2009). Thus, this methodology provides a pragmatic way forward, but several considerations should be noted, and additionally, a number of avenues for refinement of the methods are worth highlighting.

First, some discussion of previous similar analyses is worthwhile. Etterson and Shaw (2001) estimated genetic covariances between traits and relative fitness as predictors of evolution in data from a field experiment, and their approach was directly motivated by Robertson (1966) and Price (1970). In discussing their estimates, however, they suggested that their results might be different had other, unmeasured, traits been included in their analysis. Aside from the general tendency of models with different data to give numerically different results, we feel that their estimates should hold, because the genetic covariances of traits with relative fitness should hold as evolutionary predictors, regardless of whether they are generated by direct or indirect selection. A number of estimates of the genetic basis of trait-fitness relationships have been based on the regression of fitness on predicted breeding values (e.g., Kruuk et al. 2001, 2002). These have been informative in that they have demonstrated the potential for trait-fitness relationships to differ at the phenotypic and genetic levels. However, this approach is biased, potentially severely, toward the form of the *environmental* relationship between traits and fitness, and thus can be biased both in magnitude and direction (Postma 2006; Hadfield et al. 2010). This bias arises from

the fact that errors in the prediction of breeding values represent a portion of the environmental variance. Finally, other investigations of the genetic basis of phenotype–fitness relationships have been less quantitative, at least with respect to the key parameter of the STS; Robinson et al. (2006), Morrissey and Ferguson (2011), and Kruuk et al. (2002) made genetic inferences, with respect to selection and evolutionary change, of the genetic basis of phenotype–fitness relationships based on correlations of traits with absolute, rather than relative, fitness. These tests were thus not quantitatively predictive of evolutionary trajectories, but were unbiased.

The fitting of mixed models of the type used here relies on an assumption of normality of residuals. Because it is not always clear exactly what the consequences of violating this assumption will be, it is best to design studies in such a way that comparisons are made among quantities that are estimated in similar ways. The key comparisons that we present are designed this way: the covariances of traits with fitness that are at the heart of the BE (both univariate and multivariate) and the STS, that is, the phenotypic and genetic covariances of the traits with relative fitness, were all estimated in models with the same assumptions about the distribution of fitness residuals. Similarly, the comparisons of the genetic and nongenetic regressions of fitness on the traits were conducted in such a comparable manner.

However, it is possible that quantitative inferences of the sort we report here could be obtained from models that allow more sophisticated treatment of fitness distributions. Multivariate generalized linear mixed models (Bolker et al. 2009) should provide a means of testing relationships among nonnormal dependent variables, and allow estimation of covariance components between normally and nonnormally distributed traits. This is possible by assuming normality on an imaginary underlying scale, and linking observed data to this scale with a convenient “link” function. Sources of covariance among traits, or in the current context, among traits and fitness can then be estimated on this imaginary (but very useful) scale. Of course natural selection is probably best thought of as acting on the observed phenotype, rather than on some statistically convenient link scale, although this problem is not insurmountable. Both exact (Rice 2004) and numerical approximation techniques could be devised to recover observed-scale parameters from generalized linear mixed models. Recovering estimates of (observed) trait–relative fitness covariances may not be a particularly complex task, although evaluating differences between the genetic and nongenetic regressions of fitness on traits may be much more difficult. Additionally, because fitness is commonly measured as a count (e.g., LBS), it will not be normally distributed (and neither will be its residuals from any plausible model we can readily conceive of). However, it may not readily conform to a Poisson distribution either, but rather may often require a distribution that allows overdispersion (e.g.,

overdispersed Poisson). Again this is possible in a generalized linear mixed model framework, but derivation of the critical parameters on the observed scale will be more difficult still. Currently the most sophisticated available software for fitting generalized linear mixed models, the R package MCMCGLMM (Hadfield 2010) cannot readily accommodate multivariate models where some traits have repeated records whereas others (e.g., fitness) do not. We also note that the numerical problems associated with these kinds of models are enormous in the most flexible framework in which fitting these generalized linear mixed models might be conducted, that is, the BUGS (Lunn et al. 2000) or JAGS (Plummer 2010) Bayesian programming environments.

Finally, it is of note that there is little tradition in evolutionary quantitative genetics of reporting uncertainties associated with evolutionary predictions, whether obtained by application of the BE, or otherwise. Our attempts to calculate standard errors around predictions of evolutionary change should be considered approximations for a number of reasons. First, our estimated standard errors are all based on normal approximations of the sampling error. Second, our estimates for the UVBE assume independence between the estimates of heritability and selection (and estimates of the MVBE assumed independence between  $\mathbf{G}$  and  $\beta$ ). Third, and perhaps most importantly, all estimates are derived from mixed models that assume normality of residuals, as discussed above. Consequently, the standard errors we report here should not be used for hypothesis testing. However, we have included them to provide some feel of the relative uncertainties associated with the different estimated parameters and predictions that we present. In that context, it is certainly of note that the uncertainties in the evolutionary predictions based on the STS are not vastly greater than those based on the BE.

#### ACKNOWLEDGMENTS

We thank the National Trust for Scotland and Scottish Natural Heritage for permission to work on St. Kilda, and the Royal Artillery Range (Hebrides) and QinetiQ for logistic support. We thank J. Hadfield and I. White for useful comments and discussions. The Soay sheep data were collected primarily by field assistants J. Pilkington, A. MacColl, T. Robertson, J. Kinsley, and many volunteers. The long-term data collection on St. Kilda has been supported by the Natural Environment Research Council, the Wellcome Trust, the Biotechnology and Biological Sciences Research Council, and the Royal Society. MBM is supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada, LEBK was supported by a University Research Fellowship from the Royal Society, AJW is supported by a David Phillips Research Fellowship from the BBSRC, and PK is supported by a Marie Curie International Training Fellowship.

#### LITERATURE CITED

- Blackenhorn, W. U. 2000. The evolution of body size: what keeps organisms small? *Q. Rev. Biol.* 75:385–407.
- Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulson, M. H. H. Stevens, and J. S. S. White. 2009. Generalized linear mixed models:

- a practical guide for ecology and evolution. *Trends Ecol. Evol.* 24:127–135.
- Clutton-Brock, T. H., K. Wilson, and I. R. Stevenson. 1997. Density-dependent selection on horn phenotype in Soay sheep. *Philos. Trans. R Soc, B: Biol Sci* 352:839–850.
- Coltman, D. W., D. R. Bancroft, A. Robertson, J. A. Smith, T. H. Clutton-Brock, and J. M. Pemberton. 1999a. Male reproductive success in a promiscuous mammal: behavioural estimates compared with genetic paternity. *Mol Ecol.* 8:1199–1209.
- Coltman, D. W., J. A. Smith, D. R. Bancroft, J. Pilkington, A. D. C. MacColl, T. H. Clutton-Brock, and J. M. Pemberton. 1999b. Density-dependent variation in lifetime breeding success and natural sexual selection in Soay rams. *Am. Nat.* 154:730–746.
- Coltman, D. W., P. O. Donoghue, J. T. Hogg, and M. Festa-Bianchet. 2005. Selection and genetic (co)variance in bighorn sheep. *Evolution* 59:1372–1382.
- Ender, J. A. 1986. *Natural selection in the wild*. Princeton Univ. Press, Princeton, NJ.
- Etterson, J. R., and R. G. Shaw. 2001. Constraint to adaptive evolution in response to global warming. *Science* 294:151–154.
- Falconer, D. S. 1981. *Introduction to quantitative genetics*. Oliver and Boyd, New York, NY.
- Gienapp, P., C. Teplitsky, J. S. Alho, A. Millsa, and J. Merilä. 2008. Climate change and evolution: disentangling environmental and genetic responses. *Mol. Ecol.* 17:167–178.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, S. J. Welham, and R. Thompson. 2002. ASReml user guide release 1.0. VSN International Ltd, Hemel Hempstead, United Kingdom.
- Hadfield, J. 2008. Estimating evolutionary parameters when viability selection is operating. *Proc. R. Soc. Lond.* 275:723–734.
- . 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Software* 33:1–22.
- Hadfield, J. D., D. S. Richardson, and T. Burke. 2006. Towards unbiased parentage assignment: combining genetic, behavioural and spatial data in a Bayesian framework. *Mol. Ecol.* 15:3715–3731.
- Hadfield, J. D., A. J. Wilson, D. Garant, B. C. Sheldon, and L. E. B. Kruuk. 2010. The misuse of BLUP in ecology and evolution. *Am. Nat.* 175:116–125.
- Henderson, C. R. 1973. Sire evaluation and genetic trends. Pp. 10–43 in *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush*. Published by the American Society of Animal Science and the American Dairy Science Association. Champaign, IL.
- Kalinowski, S. T., M. L. Taper, and T. C. Marchall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases confidence in paternity. *Mol. Ecol.* 19:1099–1106.
- Kingsolver, J. G., and D. W. Pfennig. 2004. Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution* 58:1608–1612.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, C. Vignieri, D. Berrigan, E. Hill, A. Hoang, P. Gilbert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–261.
- Kruuk, L. E. B. 2004. Estimating genetic parameters in natural populations using the 'animal model'. *Philos. Trans. R. Soc. Lond. B* 359:873–890.
- Kruuk, L. E. B., J. Merilä, and B. C. Sheldon. 2001. Phenotypic selection on a heritable size trait revisited. *Am. Nat.* 158:557–571.
- . 2003. When environmental variation short-circuits natural selection. *Trends Ecol. Evol.* 18:207–209.
- Kruuk, L. E. B., J. Slate, J. M. Pemberton, S. Brotherstone, F. Guinness, and T. Clutton-Brock. 2002. Antler size in red deer: heritability and selection but no evolution. *Evolution* 56:1683–1695.
- Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* 33:402–416.
- . 1982. A quantitative genetic theory of life history evolution. *Ecology* 63:607–615.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS — A Bayesian modelling framework: concepts, structure and extensibility. *Stat Comput.* 10:325–337.
- Lush, J. L. 1937. *Animal breeding plans*. Iowa State College Press, Ames, Iowa.
- Lynch, M., and B. Walsh. 1998. *Genetics and analysis of quantitative traits*. Sinauer, Sunderland, MA.
- Marshall, T. C., J. Slate, L. E. B. Kruuk, and J. M. Pemberton. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7:639–655.
- Merilä, J., and B. C. Sheldon. 2000. Lifetime reproductive success and heritability in nature. *Am. Nat.* 155:301–310.
- Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica* 122–222:112–113.
- Milner, J. M., J. M. Pemberton, S. Brotherstone, and S. D. Albon. 2000. Estimating variance components and heritabilities on the wild: a case study using the animal model approach. *J. Evol. Biol.* 13:804–813.
- Milner, J. M., S. D. Albon, L. E. B. Kruuk, and J. M. Pemberton. 2004. Soay sheep: dynamics and selection in an island population. Pp. 190–216. *in* J. M. Pemberton and T. H. Clutton-Brock, eds. *Selection on phenotype*, Cambridge Univ. Press.
- Morrissey, M. B., and M. M. Ferguson. 2011. A test for the genetic basis of natural selection: an individual-based longitudinal study in a stream-dwelling fish, currently available online early. *Evolution* 65:1037–1047.
- Morrissey, M. B., L. E. B. Kruuk, and A. J. Wilson. 2010. The danger of applying the breeder's equation in observational studies of natural populations. *J. Evol. Biol.* 23:2277–2288.
- Overall, A. D. J., K. A. Byrne, J. G. Pilkington, and J. M. Pemberton. 2005. Heterozygosity, inbreeding and neonatal traits in soay sheep on St. Kilda. *Mol. Ecol.* 14:3383–3393.
- Ozgul, A., S. Tuljapurkar, T. G. Benton, J. M. Pemberton, T. H. Clutton-Brock, and T. Coulson. 2009. The dynamics of phenotypic change and the shrinking sheep of St. Kilda. *Science* 325:464–467.
- Pemberton, J. M., D. W. Coltman, J. A. Smith, and J. Pilkington. 1999. Molecular analysis of a promiscuous, fluctuating mating system. *Biol. J. Linnean Soc.* 68:289–301.
- Plummer, M. 2010. JAGS version 2.0 Manual. International Agency for Research on Cancer. Available at: <http://mcmc-jags.sourceforge.net/>
- Postma, E. 2006. Implications of the difference between true and predicted breeding values for the study of selection and microevolution. *J. Evol. Biol.* 19:309–320.
- Preston, B. T., I. R. Stevenson, J. M. Pemberton, D. W. Coltman, and K. Wilson. 2003. Overt and covert competition in a promiscuous mammal: the importance of weaponry and testes size to male reproductive success. *Proc. R. Soc. B* 270:633–640.
- Price, G. R. 1970. Selection and covariance. *Nature* 227:520–521.
- Price, T., and D. Schluter. 1991. On the low heritability of life-history traits. *Evolution* 45:853–861.
- Queller, D. C. 1992. Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* 139:540–558.
- Rausher, M. D. 1992. The measurement of selection on quantitative traits: biases due to environmental covariances between traits and fitness. *Evolution* 46:616–626.
- Réale, D., A. G. Berteaux, A. G. McAdam, and S. Boutin. 2003. Lifetime selection on heritable life history traits in north american red squirrels. *Evolution* 57:2416–2423.

Rice, S. H. 2004. Developmental associations between traits: covariance and beyond. *Genetics* 166:513–526.

Robertson, A. 1966. A mathematical model of the culling process in dairy cattle. *Anim Prod.* 8:95–108.

———. 1968. Population biology and evolution. Pp. 5–16. The spectrum of genetic variation. Syracuse Univ. Press, New York.

Robinson, M. R., J. G. Pilkington, T. H. Clutton-Brock, J. M. Pemberton, and L. E. B. Kruuk. 2006. Live fast, die young: a balance of natural, sexual and antagonistic selection maintains phenotypic polymorphism of weaponry in male and female soay sheep. *Evolution* 60:2168–2181.

Roff, D. A., and T. A. Mousseau. 1987. Quantitative genetics and fitness: lessons from *Drosophila*. *Heredity* 58:103–118.

Stinchcombe, J., M. Rutter, D. Burdick, P. Tiffin, M. Rausher, and R. Maruicio. 2002. Testing for environmentally induced bias in phenotypic estimates of natural selection: theory and practice. *Am. Nat.* 160:511–523.

Teplitsky, C., J. A. Mills, J. W. Yarrall, and J. Merilä. 2009. Heritability of fitness components in a wild bird population. *Evolution* 63:716–726.

Walling, C. A., J. M. Pemberton, J. D. Hadfield, and L. E. B. Kruuk. 2010. Comparing parentage inference software: re-analysis of a red deer pedigree. *Mol. Ecol.* 19:1914–1928.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, D. W. Coltman, D. M. Mifsud, T. H. Clutton-Brock, and L. E. B. Kruuk. 2006. Environmental coupling of selection and heritability limits phenotypic evolution. *PLOS Biol.* 4:1270–1275.

Wilson, A. J., J. M. Pemberton, J. G. Pilkington, T. H. Clutton-Brock, D. W. Coltman, and L. E. B. Kruuk. 2007. Quantitative genetics of growth and cryptic evolution of body weight in an island population. *Evol. Ecol.* 21:337–356.

Associate Editor: M. Reuter

## Appendix

### STANDARD ERRORS

Unless specifically noted here, all standard errors were obtained directly from ASREML, which uses the delta method (Lynch and Walsh 1998). For example, the standard errors of  $\sigma_p^2(z)$ ,  $\sigma_p(w, z)$ , and  $h^2(z)$  were obtained this way, fully accounting for sampling variances of each parameter and the sampling covariances among the component parameters. No methods are implemented in ASREML to obtain the standard errors of the product of the heritability and the selection gradient, and so we obtained approximate standard errors (indeed they are all approximate) assuming that squared relative standard errors are additive, that is,

$$\sigma[\Delta\bar{z}] = h^2(z)\sigma_p(w, z)\sqrt{\left(\frac{\sigma[h^2(z)]}{h^2(z)}\right)^2 + \left(\frac{\sigma[\sigma_p(w, z)]}{\sigma_p(w, z)}\right)^2}, \quad (\text{A1})$$

where we use  $\sigma$  followed by square brackets to denote sampling error. This approximation of the standard error of  $\Delta\bar{z}$  assumes that the sampling covariance between  $h^2(z)$  and  $\sigma_p(w, z)$  is zero. Although this may not be the case, we do expect it to be small in magnitude relative to the sampling variances of the terms in the summations (i.e., sampling variances of components of  $\sigma_a^2(z)$  and  $\sigma_p(w, z)$ ). It is therefore our expectation that this should be a useful approximation for the standard error on the predicted selection response.

Similarly, no automated methods are available for calculation of standard errors for predictions of the MVBE, so we adopted a Monte Carlo simulation approach. We approximated the standard errors for the estimates of  $\beta$  and  $\Delta\bar{z}_{MVBE}$  by simulating 1000 replicate multivariate normal matrices each of  $\mathbf{id}$  and  $\mathbf{G}$ , using the maximum likelihood estimates of these matrices as the mean vectors, and the variance–covariance matrices of their component parameters as the variance. Specifically, we sampled

$$\begin{bmatrix} \mathbf{id}_{\text{morph}} \\ \mathbf{e}_{\text{morph}} \\ \bar{\mathbf{S}} \end{bmatrix}' \sim N \left( \begin{bmatrix} \mathbf{id}_{\text{morph}} \\ \mathbf{e}_{\text{morph}} \\ \bar{\mathbf{S}} \end{bmatrix}, \sigma^2 \begin{bmatrix} \mathbf{id}_{\text{morph}} & \mathbf{id}_{\text{morph}} \cdot \mathbf{e}_{\text{morph}} & \mathbf{id}_{\text{morph}} \cdot \bar{\mathbf{S}} \\ \mathbf{id}_{\text{morph}} \cdot \mathbf{e}_{\text{morph}} & \mathbf{e}_{\text{morph}} & \mathbf{e}_{\text{morph}} \cdot \bar{\mathbf{S}} \\ \mathbf{id}_{\text{morph}} \cdot \bar{\mathbf{S}} & \mathbf{e}_{\text{morph}} \cdot \bar{\mathbf{S}} & \bar{\mathbf{S}} \end{bmatrix} \right),$$

$$\bar{\mathbf{G}}' \sim N(\bar{\mathbf{G}}, \sigma^2[\bar{\mathbf{G}}]),$$

where the vector symbol ( $\rightarrow$ ) denotes a vector of the elements of a variance–covariance matrix, where  $'$  indicates a parametric bootstrap replicate sample of the vector of parameters, where the bar indicates the mean estimates, and where  $\sigma[\ ]$  indicates the sampling variances and covariances of the parameters. We then applied equations (8) and (9) to each of the 1000 sets of bootstrap matrices and took the standard deviations of the resulting 1000 estimates of  $\beta$  and  $\Delta\bar{z}_{MVBE}$  to be the standard errors.