Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. Journal of Evolutionary Biology 21:1201-1219.

This document contains a PDF file of the published version of this paper, plus a series of corrections for errors in the two Appendices. The errors in Appendix 1 affected the approximations for mean conditional evolvability, $\bar{c}$ (Eq. 5), mean respondability, $\bar{r}$ (p. 1206), mean autonomy, $\bar{a}$ (Eq. 6), and mean response difference, $\bar{d}$ (Eq. 8) The corrected versions are

$$\bar{c} \approx \mathrm{H}[\lambda]\left(1+\frac{2\,\mathrm{I}[1/\lambda]}{k+2}\right),$$

$$\bar{r} \approx \sqrt{\mathrm{E}[\lambda^2]}\left(1-\frac{\mathrm{I}[\lambda^2]}{4(k+2)}\right),$$

$$\bar{a} \approx \frac{\mathrm{H}[\lambda]}{\mathrm{E}[\lambda]}\left(1+2\frac{\mathrm{I}[\lambda]+\mathrm{I}[1/\lambda]-1+\mathrm{H}[\lambda]/\mathrm{E}[\lambda]+2\,\mathrm{I}[\lambda]\,\mathrm{I}[1/\lambda]/(k+2)}{k+2}\right),$$

$$\bar{d} \approx \sqrt{\mathrm{E}[\delta^2]}\left(1-\frac{\mathrm{I}[\delta^2]}{4(k+2)}\right),$$

where the meaning of the symbols and functions are given in the paper. These errors also affected results in Figures 2, 4 and 5. New versions of these are also included here, although the differences from the original figures are very small.

doi:10.1111/j.1420-9101.2008.01573.x

Measuring and comparing evolvability and constraint in multivariate characters

T. F. HANSEN*† & D. HOULE†

*Department of Biology, Center for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway
†Department of Biological Science, Florida State University, Tallahassee, FL, USA

Keywords:

autonomy;
evolvability;
G matrix;
integration;
measurement theory;
modularity;
quantitative genetics;
scaling.

Abstract

The Lande equation forms the basis for our understanding of the short-term evolution of quantitative traits in a multivariate context. It predicts the response to selection as the product of an additive genetic variance matrix and a selection gradient. The selection gradient approximates the force and direction of selection, and the genetic variance matrix quantifies the role of the genetic system in evolution. Attempts to understand the evolutionary significance of the genetic variance matrix are hampered by the fact that the majority of the methods used to characterize and compare variance matrices have not been derived in an explicit theoretical context. We use the Lande equation to derive new measures of the ability of a variance matrix to allow or constrain evolution in any direction in phenotype space. Evolvability captures the ability of a population to evolve in the direction of selection when stabilizing selection is absent. Conditional evolvability captures the ability of a population to respond to directional selection in the presence of stabilizing selection on other trait combinations. We then derive measures of character autonomy and integration from these evolvabilities. We study the properties of these measures and show how they can be used to interpret and compare variance matrices. As an illustration, we show that divergence of wing shape in the dipteran family Drosophilidae has proceeded in directions that have relatively high evolvabilities.

Introduction

The **G** matrix summarizes the nature of additive genetic variance and covariance around a multivariate population mean. Its most familiar use is to calculate how a pattern of selection on a multivariate character, described by a directional selection gradient, $\boldsymbol{\beta}$, is transformed into a response to selection, $\Delta\bar{\mathbf{z}}$, through the 'Lande equation', $\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}$ (Lande, 1979). This equation, together with a method for estimating the selection gradient in natural populations (Lande & Arnold, 1983), has spurred tremendous interest in the study of multivariate selection and its relationship to selection responses. The result has been a small boom in empirical estimates of **G** matrices (Steppan *et al.*, 2002) and of selection gradients (Kingsolver *et al.*, 2001; Hereford *et al.*, 2004), although rarely of both in the same systems (for exceptions see, e.g. Blows *et al.*, 2004; Coltman *et al.*, 2005; Foerster *et al.*, 2007).

Although the importance of predictions based on the Lande equation is appreciated, there has been little work on how to compare responses as a function of **G**. Even when selection gradients are estimated, biologists suspect that gradients will vary with the environment and thus not necessarily be typical of natural selection over longer time periods. Consequently, most work has interpreted the structure of **G** matrices with no reference to particular selection gradients. The challenge with any general analysis is that the **G** matrix is a very complex entity, consisting of $k(k + 1)/2$ parameters when k traits are studied. This task is relatively simple when $k = 2$, in

Correspondence: David Houle, Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA.
Tel.: +1 850 645 0388; fax: +1 850 644 9829;
e-mail: dhoule@bio.fsu.edu

which case the **G** matrix consists of two additive genetic variances and an additive genetic covariance. Simulation studies often restrict themselves to two-character cases for this reason (e.g. Jones *et al.*, 2003, 2004, 2007).

Starting with Olson & Miller (1958), many proposals have been made for summary measurements and decompositions that aid the interpretation of variance matrices (Van Valen, 1978, 2005; Cheverud *et al.*, 1983, 1989; Flury, 1984; Wagner, 1984; Airoldi & Flury, 1988; Zhivotovsky, 1988; Schluter, 1996; Chernoff & Magwene, 1999; Phillips & Arnold, 1999; Klingenberg & Leamy, 2001; Magwene, 2001; Blows & Higgie, 2003; Hansen *et al.*, 2003a; Blows *et al.*, 2004; Blows, 2007; Cheverud & Marroig, 2007; Mitteroecker & Bookstein, 2007; Kirkpatrick, 2008). Some of these measures have an interpretation that connects them to evolutionary theory, such as Schluter's (1996) genetic lines of least resistance, but, unfortunately, most do not. For example, element-wise matrix correlations are often used to compare the similarity of variance matrices, including **G** matrices. The correlation of matrix elements is a measure of some sort of similarity, but there is no clear relationship between this measure and evolution. This criticism can also be raised against principal components analysis. Although it makes statistical sense to decompose variance matrices into independent axes of information, the biological and evolutionary meanings of those axes are usually unclear (Houle *et al.*, 2002; Mezey & Houle, 2003).

Schluter's (1996) contribution was to provide an evolutionary interpretation of the first principal component of the **G** matrix as the direction in phenotype space with the highest evolvability. This 'genetic line of least resistance' thus answers one evolutionary question we may ask about the **G** matrix. There are, however, many other evolutionary questions that can be asked, and we need to develop measures to answer them.

In this paper, we develop measures of evolvability based on the **G** matrix that are derived from simple evolutionary models. The goal is to present a toolbox for analysing the multivariate evolvability of natural populations. We illustrate the methods with an analysis of the relationship between evolvability and among-species divergence in the wings of drosophilid flies.

Prelude: the meaning of measurement

To understand our approach, it is helpful to discuss some concepts and ideas from formal measurement theory (e.g. Hand, 2004). Measurement theory concerns the relationship between attributes of reality and the measurements we make to represent those attributes. The fundamental idea is that measurements are intended to capture specific relationships among the attributes we measure; therefore, any manipulation of the measurements after they are obtained should preserve the relationships of interest. An explicit measurement-theoretical perspective helps to ensure that measurements and statistical procedures are made in a manner consistent with the theoretical context that motivated the measurements. All too often, data are collected with only a vague sense that they are correlated with something useful to know, and manipulations are chosen to conform to particular statistical models, regardless of whether the theoretical relevance of the measurements is preserved. This can remove the results of an analysis from the theoretical context that inspired an experiment in the first place.

Here, the attributes of 'reality' that concern us are those that capture the ability of a population to evolve. The Lande equation makes clear that the **G** matrix tells us what evolutionary response to expect when the genetic system is exposed to a given pattern of directional selection, as described by a selection gradient. We are concerned primarily with measuring and comparing the evolvabilities of traits within a population and the evolvabilities of different populations. To make such comparisons, we need to pay close attention to the scales of measurement employed. The claim that trait A is more evolvable in centimetres than trait B is in grams is not meaningful. Most researchers therefore put traits on a common scale, for example by dividing measurements by the mean or by the standard deviation. These scales have different properties and different interpretations (see, e.g. Hereford *et al.*, 2004).

Formal measurement theory recognizes a number of different scale types defined by the information we want to preserve about the measured attributes (Stevens, 1946, 1959, 1968; Sarle, 1997). Consider a set of measures such that x_A denotes a measurement taken on object A, x_B denotes a measurement taken on object B, etc. If values only reflect the order of the attributes measured, e.g. $x_A > x_B > x_C > x_D$, we have an ordinal scale. If values also convey information about the sizes of differences, such as $x_A - x_B > x_C - x_D$, we have an interval scale. If values convey information about ratios, e.g. $x_A/x_B > x_C/x_D$, but differences are not meaningful, we have a log-interval scale. If values convey information about the sizes of both ratios and differences, we have a ratio scale. Finally, if the individual measurements themselves have a natural meaning, so any alteration of values alters their meaning, we have an absolute scale. Important examples of an absolute scale are absolute fitness and probabilities.

With this in mind, it is clear that many transformations of data will alter the relationships implied by the scale type. Transformations that do preserve the scale-specific information are called permissible (Stevens, 1946). For example, on a ratio scale only transformations, $f(\cdot)$, that ensure both that if $x_A - x_B > x_C - x_D$ then $f(x_A) - f(x_B) > f(x_C) - f(x_D)$ and that if $x_A/x_B > x_C/x_D$ then $f(x_A)/f(x_B) > f(x_C)/f(x_D)$ are permissible. The only such transformation is multiplication by a positive constant. The log-interval scale permits more general

power transformations, $f(x) = ax^b$. An interval scale permits any linear transformation, that is, both addition and multiplication. On an ordinal scale, any monotonic transformation is permissible.

Quantitative genetics is usually concerned with characters that are on absolute, ratio, log-interval or interval scales. For example, measurements of relative fitness are on a ratio scale, because both ratios and differences are meaningful. Many measures of fitness components, size or amount are either on ratio or on log-interval scales, depending on how they are interpreted. For example, a set of measurements of lengths can all be treated as being on a ratio scale, but if those lengths are treated as a measure of size, which can also be measured by area or weight, then each measure is related to size and to other measures of size by a power function (Lande, 1977; Houle, 1992) and are therefore on a log-interval scale. Traits such as the seasonal timing of germination or break from diapause are most usefully treated as interval, as the origin for the timing of these events is rarely clear.

Log transformations are widely used, but measurement theory suggests that they may be overused. A log-interval scale can be transformed to an interval scale by a log transformation, allowing the same inferences to be made about differences that were reflected in ratios of the original measurements, but log transformation also changes the relative magnitudes of differences, and therefore of statistics like variances. Therefore, the use of a log transformation implies the strong assumption that differences are not meaningful and thus that the measurements are not on a ratio scale. Note also that logarithms do not have units; so, further standardization is usually not sensible. Other common transformations used to solve statistical problems are seldom permissible from a measurement-theoretical perspective. For example, transformations to normality, which are by necessity nonlinear, fundamentally change the scale and therefore the meaning of the measurements. In response to this, many defend nonlinear transformations by pointing out that back transformation is always possible. Although this claim is true for the data, it is not true for the parameter estimates derived from the data, which are nearly always the only thing that is published. This does not preclude use of such transformations, but it means that the resulting estimates lose some of their connection with the theoretical context.

Measurement theory also helps us think about which transformations are meaningful, as opposed to what is permissible. Division by mean or standard deviation is permissible on ratio, log-interval or interval scales, but it will aid interpretation only when the mean or standard deviation is itself meaningful. For example, scaling in relation to the mean is both permissible and meaningful on a ratio or log-interval scale because it gives readily interpretable meaning to the value 1. On an interval scale, scaling by the mean does not add meaning, because the mean is arbitrary, but it is permissible because inferences about differences are not altered. A variance requires preservation of relations among differences and is thus meaningful on interval and ratio scales, but not on a log-interval scale. For example, variances of size do not have the same relationship when length, area or volume is used to measure size (Lande, 1977; Houle, 1992). Neither scaling is meaningful on an ordinal scale.

Traits on different scales and scale types are often combined in the same study and lead to selection gradients and **G** matrices in which the elements have different units, raising interpretational problems for summary statistics computed on any part of the model. If the **G** matrix is studied in isolation from a selection gradient or a selection response, then some traits or directions in phenotype space may numerically dominate the rest in a meaningless manner. Comparable scaling of the traits is therefore desirable but may be difficult to achieve if we combine traits on different scale types. Below, we suggest some general approaches to multivariate scaling. In most cases, we find that mean scaling is most informative (Houle, 1992; Hansen *et al.*, 2003b; Hereford *et al.*, 2004), but we warn that this rule of thumb should not to be elevated to a general principle. There is no substitute for careful attention to the properties and interpretations of measurements in each case.

Standardization of evolvability and selection strength

The theoretical context that we shall use to develop our measures is the standard quantitative genetic account of the response to selection as embodied in Lande's (1979) multivariate equation $\Delta\bar{\mathbf{z}} = \mathbf{G}\boldsymbol{\beta}$, where $\Delta\bar{\mathbf{z}}$ is the response to selection (i.e. the change in the mean of the trait vector from one generation to the next), **G** is the additive genetic variance matrix and $\boldsymbol{\beta}$ is the selection gradient (a vector of partial regression coefficients of relative fitness on the traits in the trait vector). This simple set-up shows that any evolutionary measurement of **G** must be consistent with measurements of $\bar{\mathbf{z}}$ and $\boldsymbol{\beta}$. A variance-scaled selection gradient cannot be meaningfully combined with a mean-scaled evolvability (Hereford *et al.*, 2004).

Univariate standardizations

On a ratio or log-interval scale, we can obtain dimension-free and internally consistent measurements of response to selection, evolvability and strength of selection by standardizing with the mean:

$$\frac{\Delta\bar{z}}{\bar{z}} = \left(\frac{G}{\bar{z}^2}\right)(\beta\bar{z}) \equiv I_A\beta_\mu,$$

where G is the univariate additive genetic variance and I_A is the mean-scaled additive variance (Houle, 1992). Alternatively, values on both ratio and interval scales can

be standardized with the phenotypic standard deviation, σ_z, as

$$\frac{\Delta\bar{z}}{\sigma_z} = \left(\frac{G}{\sigma_z^2}\right)(\beta\sigma_z) \equiv h^2\beta_\sigma,$$

where h^2 is the heritability, which is thus a measure of evolutionary potential for traits on an interval scale.

Multivariate standardizations

In the multivariate case, the selection gradient and response become column vectors. For traits on ratio and log-interval scales, the elements of the response vector and the selection gradient can each be standardized by the corresponding trait means, yielding the mean-standardized selection gradient $\boldsymbol{\beta}_{\boldsymbol{\mu}} \equiv \bar{\mathbf{z}} \odot \boldsymbol{\beta}$, where $\odot$ denotes element-wise multiplication, also known as a Hadamard product. The additive genetic variances and covariances must therefore also be mean standardized by dividing with the products of the trait means to yield the mean-standardized $\mathbf{G}$,

$$\mathbf{G}_{\boldsymbol{\mu}} \equiv \mathbf{G} \oslash (\bar{\mathbf{z}}\bar{\mathbf{z}}'),$$

where $\oslash$ denotes element-wise division. Thus, the ijth element of the $\mathbf{G}$ matrix is standardized by division by the product of the means of traits i and j. The resulting matrix has I_A values on the diagonal. The mean-standardized vector of selection responses is then

$$\Delta\bar{\mathbf{z}} \oslash \bar{\mathbf{z}} = \mathbf{G}_{\boldsymbol{\mu}}\boldsymbol{\beta}_{\boldsymbol{\mu}}.$$

To make meaningful inferences about differences, two different standardizations yield sensible results. First, we can standardize each trait and each element in the selection gradient with the squares or cross-products of phenotypic standard deviations to obtain the variance-standardized $\mathbf{G}$,

$$\mathbf{G}_{\boldsymbol{\sigma}} \equiv \mathbf{G} \oslash (\boldsymbol{\sigma}\boldsymbol{\sigma}'),$$

where $\boldsymbol{\sigma}$ is a column vector of phenotypic standard deviations. The resulting matrix has heritabilities on the diagonal. The variance-standardized vector of selection responses is

$$\Delta\bar{\mathbf{z}} \oslash \boldsymbol{\sigma} = (\mathbf{G} \oslash (\boldsymbol{\sigma}\boldsymbol{\sigma}'))(\boldsymbol{\beta} \odot \boldsymbol{\sigma}) \equiv \mathbf{G}_{\boldsymbol{\sigma}}\boldsymbol{\beta}_{\boldsymbol{\sigma}},$$

where each element of the response vector is in standard deviation units. Second, we can use a full multivariate standardization,

$$\mathbf{P}^{-1/2}\Delta\bar{\mathbf{z}} = (\mathbf{P}^{-1/2}\mathbf{G}\mathbf{P}^{-1/2})(\mathbf{P}^{1/2}\boldsymbol{\beta}) \equiv \mathbf{G}_{\mathbf{P}}\boldsymbol{\beta}_{\mathbf{P}},$$

where $\mathbf{P}$ is the phenotypic variance matrix, and the matrix $\mathbf{G}_{\mathbf{P}} \equiv \mathbf{P}^{-1/2}\mathbf{G}\mathbf{P}^{-1/2}$ is a multivariate generalization of the heritability. Under this scaling, the lengths of the response and gradient vectors are in units of the phenotypic standard deviation in that direction. In the following, we assume that all traits are mean standardized, unless we state otherwise.

Multivariate evolvability: theoretical considerations

Figure 1 shows the response to a directional selection gradient, $\boldsymbol{\beta}$, in a two-dimensional phenotypic space. Our measures of evolvability follow from this geometrical representation of selection response.

A proposal for measuring multivariate evolvability along a selection gradient

Our first goal is to derive simple scalar measures of unconditional evolvability for a multivariate character subject only to linear directional selection. The initial rate of response is the vector norm

$$|\Delta\bar{\mathbf{z}}| = |\mathbf{G}\boldsymbol{\beta}| = \sqrt{\boldsymbol{\beta}'\mathbf{G}^2\boldsymbol{\beta}},$$

which can be calculated from any consistently standardized $\mathbf{G}$ and $\boldsymbol{\beta}$. We do not directly use this as a measure of evolvability, because doing so would give the same measure of evolvability for a population with a response in the direction of the selection gradient as one whose response is in a different direction. Instead, we base our measures of evolvability on the length of the projection of the evolutionary response vector on the selection gradient

$$|\Delta\bar{\mathbf{z}}|\cos[\theta] = \frac{\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}}{|\boldsymbol{\beta}|},$$

where θ is the angle between the selection gradient and the response vector. The projection measures the

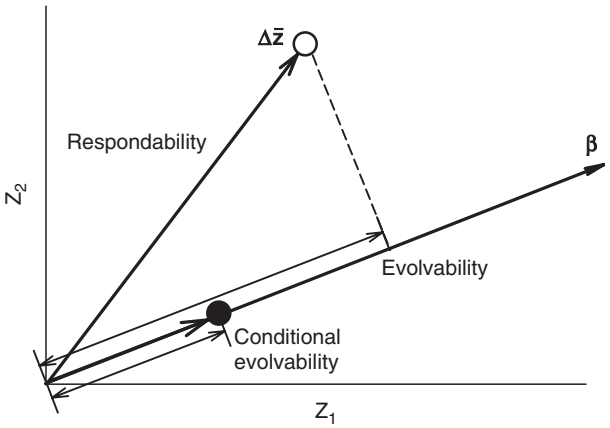


Fig. 1 Measures of selection response. Circles show the response to selection in two traits, z_1 and z_2, in response to two scenarios. The open circle shows the response to pure standardized linear directional selection gradient $\boldsymbol{\beta}$ of length 1. Respondability is the length of the predicted response to selection. Evolvability is the length of response in the direction of $\boldsymbol{\beta}$ and corresponds to the length of the projection of $\Delta\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$. In our second scenario, we assume that, in addition to directional selection, there is stabilizing selection around the direction of $\boldsymbol{\beta}$, such that the population cannot deviate from $\boldsymbol{\beta}$. The closed circle shows selection response under this scenario; the length of this constrained response is conditional evolvability.

response in the selected direction, ignoring responses in every other direction.

To turn this projection into a measure of evolvability, we normalize by the strength of selection, the length of the selection gradient. Thus, we define evolvability in the direction of an arbitrary-length $\boldsymbol{\beta}$ as

$$e(\boldsymbol{\beta}) \equiv \frac{\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}}{|\boldsymbol{\beta}|^2}. \quad (1)$$

Although $e(\boldsymbol{\beta})$ can be calculated on any of the scales discussed, the mean-standardized $e_\mu(\boldsymbol{\beta})$ is most readily interpreted. As discussed by Hansen *et al.* (2003b), the mean-standardized strength of selection on relative fitness is 1. Mean-standardized evolvability gives the predicted proportional change in the mean-standardized trait index when selection is as strong as that on fitness. For a single selected trait, it reduces to the mean-scaled additive genetic variance, I_A. On a variance-standardized scale, $e_\sigma(\boldsymbol{\beta})$ gives the predicted change in the standard-deviation-standardized trait index when relative fitness changes by a value of 1 over a unit change in the standardized trait index. On the **P**-standardized scale, $e_P(\boldsymbol{\beta})$ gives the number of standard deviations that the selected index will evolve when relative fitness changes by a value of 1 over 1 standard deviation of the selected trait index. Although $e(\boldsymbol{\beta})$ can be calculated on the raw scale, the result is in a mixture of the units of the traits in direction $\boldsymbol{\beta}$. This will only be sensible when all traits are measured on strictly comparable scales.

Note that $e(\boldsymbol{\beta})$ on any scale is strictly a measure of evolvability in one direction in phenotype space and is therefore likely to be different for every choice of $\boldsymbol{\beta}$. Note also that, if $\boldsymbol{\beta}$ points in the direction of an eigenvector of **G**, the multivariate evolvability reduces to the corresponding eigenvalue of **G**. Schluter's genetic line of least resistance is the direction in phenotype space in which $e(\boldsymbol{\beta})$ is maximized.

Respondability

Although the norm of the response vector is not a good measure of evolvability in the sense of ability to evolve along a given selection gradient, it is a useful measure of the respondability of a population, as it measures how rapidly the population will respond when under directional selection. To formalize this, we suggest the following measure of respondability to a selection gradient, $\boldsymbol{\beta}$,

$$r(\boldsymbol{\beta}) \equiv \frac{|\boldsymbol{\Delta}\bar{\mathbf{z}}|}{|\boldsymbol{\beta}|} = \frac{\sqrt{\boldsymbol{\beta}'\mathbf{G}^2\boldsymbol{\beta}}}{|\boldsymbol{\beta}|}. \quad (2)$$

The ratio between the evolvability and the respondability equals the cosine of the angle between the selection gradient and the response. Respondability is, thus, always greater than or equal to the evolvability.

Measuring constraints: conditional evolvability

The conditional evolvability of a trait y with respect to a set of constraining traits $\mathbf{x}$ is defined as the response in y to a unit directional selection if $\mathbf{x}$ is not allowed to change (Hansen *et al.*, 2003a). This is equivalent to the expected response in y when the strength of stabilizing selection on $\mathbf{x}$ has come to an equilibrium with directional selection on y. The conditional evolvability is equal to the conditional genetic variance of y given $\mathbf{x}$, or

$$c(y|\mathbf{x}) \equiv G_y - \mathbf{G_{yx}}\mathbf{G_x}^{-1}\mathbf{G_{xy}},$$

where G_y is the genetic variance in y, $\mathbf{G_{yx}}$ and $\mathbf{G_{xy}}$ are row and column vectors of covariances between y and the traits in $\mathbf{x}$, and $\mathbf{G_x}$ is the variance matrix of $\mathbf{x}$ (Hansen *et al.*, 2003a; Hansen, 2003). To better understand how the conditional evolvability measures constraints, we can rewrite the equation as

$$c(y|\mathbf{x}) = G_y(1 - G_y^{-1}\mathbf{G_{yx}}\mathbf{G_x}^{-1}\mathbf{G_{xy}}) = G_y(1 - i(y|\mathbf{x})) \equiv G_y a(y|\mathbf{x}),$$

where $i(y|\mathbf{x})$ is the square of the multiple correlation coefficient between y and $\mathbf{x}$ (Anderson, 1984), and we define $a(y|\mathbf{x}) \equiv 1 - i(y|\mathbf{x})$ as the autonomy of y with respect to $\mathbf{x}$. The conditional evolvability of a character is therefore equal to its evolvability multiplied by its autonomy. Autonomy is the fraction of genetic variation that is independent of potentially constraining characters.

The total potential for constraint on a trait can be measured through its conditional evolvability with respect to all other measured traits. This total conditional evolvability is easily computed from the inverse of the **G** matrix, as it is equal to the inverse of the corresponding diagonal element of $\mathbf{G}^{-1}$ (i.e. $c(z_i)=1/[\mathbf{G}^{-1}]_{ii}$ for the ith trait z_i, where the notation $[\mathbf{A}]_{ij}$ signifies the ijth element of the matrix **A**).

Motivated by the asymptotic invariance of the conditional evolvability with respect to the strength and pattern of multivariate stabilizing selection on the set of constraining characters (Hansen, 2003), we define the conditional evolvability along a selection gradient to be the response along this gradient when no response is allowed in any other direction of phenotype space.

Let $y = \boldsymbol{\beta}'\mathbf{z}$ and $\mathbf{x} = \mathbf{A}'\mathbf{z}$ be an orthogonal transformation of the coordinate system to a new set of traits, such that y is an index trait pointing in the direction of the selection gradient. Assume **G** is positive definite. When $|\boldsymbol{\beta}| = 1$, the response to selection in y when $\mathbf{x}$ is held constant is

$$c(\boldsymbol{\beta}) = e(\boldsymbol{\beta})a(\boldsymbol{\beta}) = (\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}. \quad (3)$$

A proof is given in Result 1 of Appendix 1. Here and below we omit $\mathbf{x}$ from the notation, with the understanding that all measured constraining characters are included in the conditioning. Note that the conditional evolvability along a selection gradient equals the

unconditional evolvability multiplied by the autonomy in the direction $\boldsymbol{\beta}$, $a(\boldsymbol{\beta})$.

Evolvability and respondability in response to random selection gradients

Evolvabilities will not be the same in different directions in phenotype space. In the absence of knowledge about the gradient that a population will actually experience, we can assess the evolutionary potential of a **G** matrix by computing its average evolvability over random selection gradients. If the vector $\boldsymbol{\beta}$ is symmetrically distributed in k-space with $\mathrm{E}[|\boldsymbol{\beta}|] = 1$, then

$$\bar{e} \equiv \mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] = \frac{\sum_i \lambda_i}{k} \equiv \mathrm{E}[\lambda], \qquad (4)$$

where λ_i are eigenvalues of **G** and the summation is over all k eigenvalues. A proof is given in Result 2 of Appendix 1. The average eigenvalue or, equivalently, the average trait additive variance, is therefore a measure of the evolutionary potential inherent in a **G** matrix. Note that the average unconditional evolvability is unaffected by covariances between traits, because unconditional evolvabilities are, by assumption, free of selective constraints.

To get a general measure of genetic constraints, we can calculate the average conditional evolvability over all directions in phenotype space. We have not been able to obtain a general analytical solution for this, but an approximate solution is the following. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and **G** is of full rank, then

$$\bar{c} \equiv \mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}] \approx \mathrm{H}[\lambda]\left(1 + \frac{\mathrm{I}[1/\lambda]}{k+1}\right), \qquad (5)$$

where $\mathrm{H}[x] \equiv 1/\mathrm{E}[1/x]$ denotes the harmonic mean and $\mathrm{I}[x] \equiv \mathrm{var}[x]/\mathrm{E}[x]^2$ denotes the mean-standardized variance. A proof is given in Result 3 of Appendix 1.

The approximation is best when the number of dimensions is large and the eigenvalues are similar. For two characters, we can show by direct integration that the average conditional evolvability is exactly the geometric mean of the two eigenvalues, $\bar{c} = \sqrt{\lambda_1\lambda_2}$, but this is not the case for higher dimensional phenotypes. Numerical simulations indicate that the geometric mean is an upper bound, and it can be shown that the harmonic mean is a lower bound. In fact, as the number of traits becomes large, the average conditional evolvability converges on the harmonic mean of the eigenvalues. We could also compute the value of $\bar{c}$ numerically by sampling a large number of selection gradients from the uniform distribution (or any distribution of interest). Doing so should not be necessary, however, as the approximation given in eqn 5 appears extremely accurate in all cases we have considered (Fig. 2).

The average respondability over random unit selection gradients can be taken as a general measure of respondability. From Result 5 in Appendix 1, we get the following approximation

$$\bar{r} \equiv \mathrm{E}[r(\boldsymbol{\beta})] = \mathrm{E}\left[\sqrt{\boldsymbol{\beta}'\mathbf{G}^2\boldsymbol{\beta}}\right] \approx \sqrt{\mathrm{E}[\lambda^2]}\left(1 - \frac{\mathrm{I}(\lambda^2)}{8(k+1)}\right),$$

where λ are the eigenvalues of **G** and k is the number of traits measured.

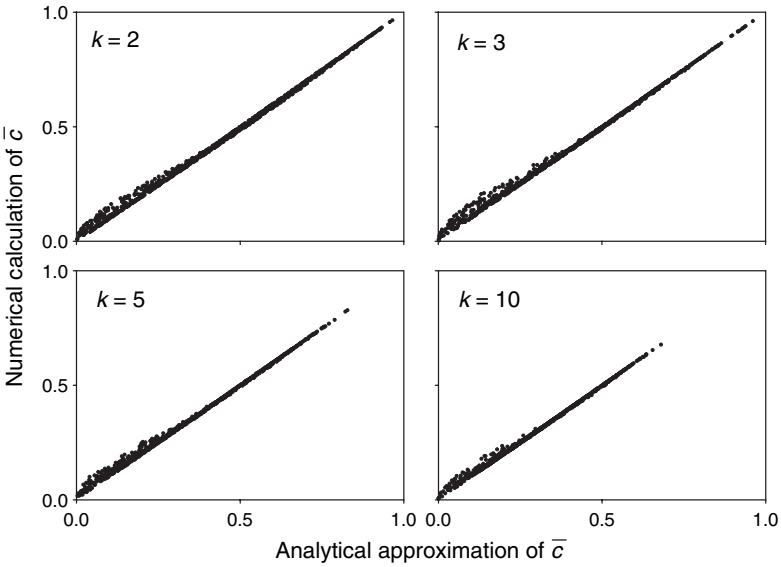


Fig. 2 Approximation of mean conditional evolvability, $\bar{c}$. The plots show numerically computed $\bar{c}$ plotted against the analytical approximation in Result 3, Appendix 1, for 1000 random **G** matrices of various dimensionalities (k). In all cases, the matrices have random diagonal entries drawn from a uniform [0,1] distribution and zero off-diagonal elements. This is justified as the symmetry of the random selection gradients implies that the results are unaffected by diagonalization. The numerical mean is computed over 10 000 random unit selection gradients.

Measuring integration and autonomy

The relative degree to which evolvability is reduced by conditioning on traits under stabilizing selection can be a measure of how integrated the selected trait combination is with the rest of the measured phenotype. This integration is captured by i, the relative reduction in evolvability due to correlated traits. We can also think in terms of autonomy, $a = 1 - i$, the proportion of evolvability that remains after conditioning on other traits. From the definitions of $c(\boldsymbol{\beta})$, $e(\boldsymbol{\beta})$ and $i(\boldsymbol{\beta})$ we have

$$a(\boldsymbol{\beta}) \equiv 1 - i(\boldsymbol{\beta}) = \frac{c(\boldsymbol{\beta})}{e(\boldsymbol{\beta})}.$$

If variation along $\boldsymbol{\beta}$ is independent of variation along other directions, $i = 0$ and $a = 1$. If variation along $\boldsymbol{\beta}$ is completely correlated with variation along other directions, $i = 1$ and $a = 0$. If we are interested in the autonomy of a specific character with respect to the rest, we take $\boldsymbol{\beta}$ to be the vector with a coefficient of 1 for this character and zero for the others. For the jth character the result is

$$a(z_j) = \left([\mathbf{G}^{-1}]_{jj}[\mathbf{G}]_{jj}\right)^{-1}.$$

We can also ask about the degree of autonomy averaged over all directions in phenotype space. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and $\mathbf{G}$ is of full rank, then

$$\begin{aligned}\bar{a} &\equiv \mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}] \\ &\approx \frac{\mathrm{H}[\lambda]}{\mathrm{E}[\lambda]}\left(1 + \frac{\mathrm{I}[\lambda] + \mathrm{I}[1/\lambda] + 1 - \mathrm{H}[\lambda]/\mathrm{E}[\lambda]}{k+1}\right). \qquad (6)\end{aligned}$$

A proof is given in Result 4 of Appendix 1. Note that $\bar{i} = 1 - \bar{a}$. Average autonomy will have a value of one if and only if all characters are uncorrelated and have the same variance. This situation is the only one in which the conditional and unconditional evolvabilities are the same in all directions. Average autonomy will decrease with increasing variation among the eigenvalues and approach a minimum of zero when some eigenvalues approach zero.

Comparing G matrices

The $\mathbf{G}$ matrix describes the mapping of a selection gradient to a selection response. We can therefore measure the difference between two $\mathbf{G}$ matrices in terms of the difference in the responses they generate to the same selection gradient. Doing so reduces the problem from comparing two matrices to comparing two vectors. These response vectors can be compared in several potentially meaningful ways. Figure 3 shows two response vectors in two space, along with parameters useful for comparing them. Angles between the response vectors, θ_d, capture differences in the direction of response (Cheverud *et al.*, 1983). The relative ability to respond in the direction of selection is the ratio of the projection of the responses onto $\boldsymbol{\beta}$. We caution against comparing functions of the length of the response vectors because these will usually be incommensurate whenever the response vectors point in different directions.

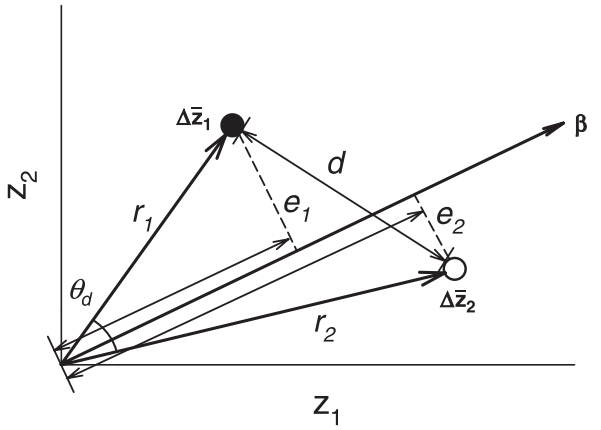


Fig. 3 Responses, $\Delta\bar{\mathbf{z}}$, of two populations with the same starting mean but different $\mathbf{G}$ matrices to the same selection gradient, $\boldsymbol{\beta}$, of unit length. The length of the response to selection is respondability, $r(\boldsymbol{\beta})$. The length of the projection of the responses on $\boldsymbol{\beta}$ are the evolvabilities of each population along $\boldsymbol{\beta}$, $e(\boldsymbol{\beta})$. The length of the difference between populations under the assumption that they start with the same mean is the response difference $d(\boldsymbol{\beta})$. In the figure, we suppress the use of the parenthetical $\boldsymbol{\beta}$ in these measures for clarity. The angle between the selection responses is θ_d.

The distance between the endpoints of the vectors captures the extent to which uniform linear selection would cause divergence between populations, if they had the same starting mean. Let $\mathbf{G}_1$ and $\mathbf{G}_2$ be the $\mathbf{G}$ matrices we want to compare. The norm of the difference between the responses they generate is

$$d(\boldsymbol{\beta}) = |\Delta\bar{\mathbf{z}}_1 - \Delta\bar{\mathbf{z}}_2| = |(\mathbf{G}_1 - \mathbf{G}_2)\boldsymbol{\beta}| = \sqrt{\boldsymbol{\beta}'(\mathbf{G}_1 - \mathbf{G}_2)^2\boldsymbol{\beta}}. \qquad (7)$$

We call this the 'response difference' between $\mathbf{G}_1$ and $\mathbf{G}_2$.

The expectation of $d(\boldsymbol{\beta})$ over random selection gradients is a useful general measure of the potential for divergence engendered by selection. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and $\mathbf{G}_1$ and $\mathbf{G}_2$ are of full rank, then the expected response difference is

$$\begin{aligned}\bar{d} &\equiv \mathrm{E}[\Delta\bar{\mathbf{z}}_1 - \Delta\bar{\mathbf{z}}_2] = \mathrm{E}\left[\sqrt{\boldsymbol{\beta}'(\mathbf{G}_1 - \mathbf{G}_2)^2\boldsymbol{\beta}}\right] \\ &\approx \sqrt{\mathrm{E}[\delta^2]}\left(1 - \frac{\mathrm{I}(\delta^2)}{8(k+1)}\right), \qquad (8)\end{aligned}$$

where δ are the eigenvalues of the matrix $\mathbf{G}_1 - \mathbf{G}_2$. The approximation follows from Result 5 of Appendix 1 if $\mathbf{M}$ is set equal to $(\mathbf{G}_1 - \mathbf{G}_2)^2$.

Note that $d(\boldsymbol{\beta})$ and $\bar{d}$ are informative only if the two $\mathbf{G}$ matrices are on the same scale. In general, the

populations compared will have different means and variance matrices and therefore naturally have different standardizations. Two sensible standardizations are to use the mean of the two population mean vectors, or the square root of the average of the phenotypic variance vectors. With mean standardization, $\bar{d}$ is interpretable as the per cent difference of the responses relative to the chosen trait mean.

Cheverud *et al.* (1983) proposed assessing differences among matrices by use of the distribution of angles between responses to random selection gradients. They termed this method random skewers (Cheverud, 1996; Cheverud & Marroig, 2007).

The ratios $\bar{e}_1/\bar{e}_2$ and $\bar{c}_1/\bar{c}_2$ can also be used to assess relative evolvabilities. These are dimensionless and so can be interpreted either under a common standardization or when each population is standardized with its own vector, although the two results differ slightly in meaning. For example, when each population is standardized by its own mean vector, $\bar{e}_1/\bar{e}_2$ compares the relative ability of each population's mean vector to evolve. When each is standardized by the same mean vector, their ratio compares the relative ability to evolve toward or away from the chosen mean.

Example: evolvability of wing shape

We present two examples of the use of our measures of evolvability. Appendix 2 is a worked example of comparisons between hypothetical populations where the **G** matrix of each is known. In this section, we use data on wing shape in drosophilid flies to compare among-species divergence to the evolvability predicted from the **G** matrix of a single species.

A persistent question in evolutionary quantitative genetics is whether the **G** matrix constrains divergence among populations or species or, conversely, whether the pattern of evolution that leads to species differences shapes the **G** matrix. Schluter (1996) found that among-species and among-population variations tended to lie close to the direction in phenotype space with the highest evolvability, $\mathbf{g}_{\max}$. Calculating evolvability and conditional evolvability along the vectors that distinguish species' means allows us to examine more general versions of Schluter's hypothesis. If **G** shapes among-species differences (or is shaped by them), then the differences among species should be in those aspects of variation that have the highest evolvabilities, even if those are very different from $\mathbf{g}_{\max}$.

We took *Drosophila melanogaster* as our focal species and used the divergence between *D. melanogaster* and other representative drosophilid species that span the traditional genus *Drosophila* to define interesting directions in which to assess evolvability. Our estimates of a **G** matrix are taken from a large study of a wild-collected population of *D. melanogaster* from Wabasso Florida (Mezey & Houle, 2005). Species stocks were obtained by collection (*D. simulans*, *D. pseudoobscura*, *D. ananassae*, *D. willistoni* and *Scaptodrosophila latifasciaeformis*) or from the Drosophila Species Stock Center (*D. virilis* and *D. grimshawi*). The phylogenetic distance ranges from *D. simulans*, the sister taxon of *D. melanogaster*, to *S. latifasciaeformis*, a representative of the genus basal to the paraphyletic genus *Drosophila* (van der Linde & Houle, 2008). The mean of each species is based on approximately 200 images. All data were obtained from laboratory-reared flies.

What we are estimating is how readily *D. melanogaster* can be selected to have the wing shape of other species. The vector differences between species are not the paths along which these species diverged from their common ancestor. Although estimating evolvability along the path from the common ancestor would also be interesting, we have not done so, as it would introduce two additional sources of uncertainty. First, the mean of the common ancestor would have to be estimated on the basis of a specific evolutionary model. Second, and more important, our estimate of the **G** matrix is for *D. melanogaster*, and we have no other information with which to estimate the **G** matrix of the common ancestor.

The data for this analysis were the x, y coordinates of 12 vein intersections measured with WINGMACHINE, a semi-automated system that records scale information and detects vein positions from digital wing images (Houle *et al.*, 2003). The 24 coordinates obtained from each wing are registered using Procrustes least-squares superimposition, which removes centroid size as a scaling factor. Although the superimposed data are still in the form of 12 pairs of coordinates, four degrees of freedom are used for registration; so, the resulting **G** matrix has a maximum dimensionality of 20. Mezey & Houle (2005) estimated separate **G** matrices for each gender and found that each matrix had the maximum possible dimensionality of 20, that is, 20 eigenvalues were greater than zero. We averaged male and female **G** matrices to obtain an estimate relevant to among-species evolution. Because $\bar{c}$ is 0 for a singular **G** matrix, we first projected the **G** matrix and the species mean data into the subspace defined by the first 20 eigenvectors of the full **G** matrix.

The resulting data are coordinates with units of centroid size; the (0, 0) point corresponds to the centre of the shape. The coordinates themselves are therefore on an interval scale, as ratios of the coordinates are not meaningful; any other point could have been chosen as the origin. By definition, shape is what is preserved when translation (spatial position), orientation and size are removed from the coordinates. This means that aspects of shape are on a ratio scale, and that use of centroid-standardized shape data to parameterize **G** and **β** is similar to mean standardization.

In Table 1, we show the phenotypic distance between the mean of *D. melanogaster* and that of each of the other species. *Drosophila simulans* wing shape is so similar to that of *D. melanogaster* that individuals of the two species

Table 1 Phenotypic distances to *Drosophila melanogaster* from other members of its genus and from *Scaptodrosophila latifasicaeformis*, evolvability statistics in the direction of species divergences, and generations to evolve from one species mean to another.

Species	Distance to *D. melanogaster* (cs)	$e(\boldsymbol{\beta})$ (cs × 10^6)	$c(\boldsymbol{\beta})$ (cs × 10^6)	$r(\boldsymbol{\beta})$ (cs × 10^6)	$a(\boldsymbol{\beta})$	Generations*
D. simulans	0.011	34.4	2.7	44.2	0.079	4125
D. ananassae	0.087	66.7	13.7	70.5	0.205	6348
D. pseudoobscura	0.043	64.9	12.7	67.7	0.196	3362
D. willistoni	0.061	55.1	10.7	60.1	0.193	5681
D. virilis	0.058	46.6	10.5	52.8	0.225	5514
D. grimshawi	0.173	55.2	17.4	57.8	0.315	9939
S. latifasicaeformis	0.117	56.9	24.9	60.7	0.438	4688

'cs' is centroid size.
*Phenotypic distance divided by $c(\boldsymbol{\beta})$.

cannot be reliably distinguished by a discriminant function (Houle *et al*., 2003), unlike individuals of any of the other species pairs. To calculate the number of generations necessary to travel the distance between the species means when the population is constrained to follow the direct path to the other species mean and the length of $\boldsymbol{\beta}$ is 1, we divide the phenotypic distance by the conditional evolvability.

Figure 4 shows the range of evolvabilities possible, from the maximum, e_{max}, to the minimum, e_{min}, as defined by the range of the eigenvalues of the 20-dimensional **G** matrix. The maximum evolvability, e_{max}, is that along $\mathbf{g}_{max}$, the first eigenvector of the matrix (Schluter, 1996). Also shown are the average unconditional, $\bar{e}$, and conditional, $\bar{c}$, evolvabilities over the entire space spanned by **G**. The conditional evolvability also has a maximum of e_{max} and minimum of e_{min}.

The evolvabilities along the vectors that define the differences between *D. melanogaster* and each species are

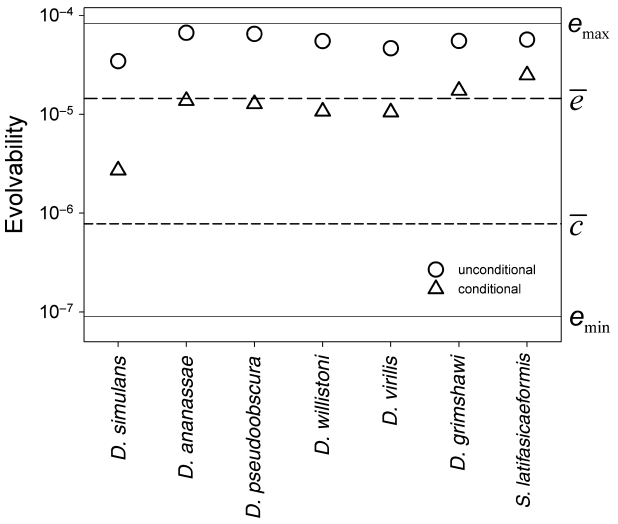


Fig. 4 Unconditional and conditional evolvabilities along the vector of differences in species means for wing shape between *Drosophila melanogaster* and other drosophilid species. The mean conditional and unconditional evolvabilities are shown as dashed horizontal lines. The evolvabilities are in units of centroid size.

mostly rather close to the maximum evolvability in the entire data set and well above $\bar{e}$. The conditional evolvabilities are all far above $\bar{c}$, and in fact most are near $\bar{e}$. The similarity of evolvabilities does not arise because all species have evolved in the same directions, as the median angle between divergence vectors is 67°, only slightly less than the median angle between random vectors in 20 space of 81.4° (determined from 1000 angles between pairs of random vectors). The smallest evolvability is along the vector between the sister taxa *D. simulans* and *D. melanogaster*, and it is still 42% of e_{max} and 236% of $\bar{e}$. Given the small phenotypic distance between *D. melanogaster* and *D. simulans*, the lower evolvability between these species may be due to error in the estimation of the species means, which would cast relatively more of the difference into dimensions of phenotype space with lower evolvabilities. Other than the small difference and relatively low evolvability between these two species, none of the measures show an obvious tendency to change with phylogenetic distance.

The median angle between $\mathbf{g}_{max}$ and the seven divergence vectors is 65° (the minimum is 38°). The directions of divergence are therefore quite different from $\mathbf{g}_{max}$, but still fall within a part of phenotype space with relatively high evolvabilities. Simply relying on comparisons with $\mathbf{g}_{max}$ would erroneously suggest that the mutual dependence of species divergence and **G** is not strong.

In summary, the characteristics of genetic variation along the vectors that distinguish *D. melanogaster* from other drosophilid species are quantitatively similar regardless of the timescale of divergence. These directions show far more than the average amount of variation, suggesting some causal connection between the forces that shape standing genetic variation and those that cause evolutionary divergence. This result suggests that broad characteristics of the **G** matrix are preserved over the more than 50 million years of evolution captured in this clade of flies. We can think of many hypothetical scenarios that would decouple long-term evolution from our short-term measures of evolvability; that they do not seem to occur is therefore quite striking (cf. Schluter, 1996).

Discussion

Variation is the basis for selection and is therefore central to evolutionary biology (Hallgrímsson & Hall, 2005). This should place the study of genetic variation of quantitative traits, summed up by the **G** matrix, at the centre of the study of evolution. Despite the clear theoretical interest of such estimates, studies of the **G** matrix for more than a small number of traits are still rare. This is usually ascribed, with good reason, to the fact that the empirical estimation of **G** matrices is laborious and difficult. We are concerned with a second, conceptual limitation; most methods for analysing **G** matrices are only tenuously related to evolutionary theory.

A great diversity of methods for measuring and describing variation of quantitative traits have been proposed. The justification and motivation for these methods may be as diverse as the methods themselves, but, with few exceptions, the role of theory in their formulation has been vague and qualitative, rather than precise and quantitative. For example, almost all the proposed methods for describing and quantifying modularity and integration among characters are based on purely statistical or intuitive considerations. Few attempts have been made to formalize what attributes of reality are measured by these techniques or at elucidating how statistical 'modularity' relates to attributes embedded in evolutionary theory. Quantitative measurements of genetic correlations are usually motivated by a desire to understand constraints on evolutionary change, but these have not generally been translated into quantitative statements about constraint.

We have developed measures of evolvability with an explicit link to the Lande model. Despite the fact that it makes a number of idealizing assumptions (Lande, 1979; Lande & Arnold, 1983), this model has motivated most of the empirical research on multivariate selection and constraints since 1980. The Lande model is therefore the appropriate context for developing measures of evolvability that take advantage of the existing data in evolutionary quantitative genetics.

In most usages of the term, evolvability refers to an ability to respond to a potential selective challenge. At least two different levels of evolvability are commonly recognized. Often it is defined as a property of the genotype–phenotype map and involves structural characteristics such as mutability, variability, autonomy, modularity, coordination and continuity (Wagner & Altenberg, 1996; Hansen, 2006). Alternatively, but not inconsistently, evolvability can be viewed in a quantitative genetics context as a short-term ability to respond to directional selection (Houle, 1992), which depends on standing additive genetic variation in the traits under selection and on the autonomy of this genetic variation from other traits under stabilizing selection (Hansen *et al.*, 2003a, b; Hansen, 2003; Hansen & Houle, 2004). The measurements we propose are derived under the short-term perspective. Nevertheless, our results from drosophilid wings, as well as the result of others (Cheetham *et al.*, 1993; Schluter, 1996, 2000; Blows & Higgie, 2003; Marroig & Cheverud, 2005; Estes & Arnold, 2007; Hunt, 2007), suggest that this short-term perspective still provides information about the effects of genetic constraints on macroevolutionary patterns. Why it does so is not entirely clear, given the many ways that long-term evolution can be decoupled from within-population variation. One potential explanation is that both short-term evolvability and long-term evolution depend on a conservative pattern of mutation. This dependence suggests that our measures of evolvability should be applied to the mutational variance matrix in addition to the **G** matrix.

Conditional and unconditional evolvabilities offer different perspectives on evolvability. Unconditional evolvability, e, is a measure of evolvability along a linear fitness landscape. Conditional evolvability, c, describes evolvability with respect to shifting adaptive peaks in concave fitness landscapes, and is typically much smaller than the unconditional evolvability (Hansen *et al.*, 2003a; Jensen *et al.*, 2003; Coltman *et al.*, 2005; Rolff *et al.*, 2005; Rønning *et al.*, 2007). Conditional evolvabilities reflect potential selective constraints due to the traits that are included in the estimates and are therefore particularly sensitive to the choice of traits. As we have suggested before (Hansen, 2003; Hansen *et al.*, 2003a; Hansen & Houle, 2004), if a good measure of general fitness were available, a conditional evolvability of a trait, or along a phenotypic direction, could be calculated relative to a measure of background fitness and would then constitute a test of the 'quality' of the genetic variation. Hansen & Houle (2004) pointed out that the high evolvability generally implied by high levels of additive genetic variation and high mutation rates may be illusory, as much of the genetic variation may be 'junk' caused by alleles with deleterious pleiotropic effects. Such variation may, particularly in the benign setting of artificial selection, allow a response to selection, but the genetic basis of this response will carry a fitness cost and is unlikely to form the basis of a permanent adaptation in the wild. This hypothesis remains essentially untested, but see Galis *et al.* (2006).

A key advantage of the measures of evolvability we have presented is that they can be computed along any direction in phenotype space. This allows for more precise ways of studying constraints on micro- or macroevolutionary transitions. For example, we can compute the evolvability of a specific transition from one species to another and quantify how constrained it would be. This result can then form a basis for testing for a relationship between directions of divergence and evolvability. In Fig. 4, we show that species differences in wing shape of drosophilid flies fall along directions where evolvabilities are much higher than average. Most of the conditional evolvabilities are more than an order

of magnitude above the average conditional evolvability. Schluter (1996) pioneered this approach by comparing the angle between divergence and the direction of maximum evolvability. Direct computation of evolvability for the divergence in question provides for a more general and detailed way of testing the same hypotheses. The interpretation of such similarities can be ambiguous, as neutral divergence is also expected to follow directions with large amounts of genetic variance (McGuigan *et al.*, 2005).

Our new measures of evolvability also provide new ways of quantifying and characterizing modularity and character integration in the context of evolvability. The autonomy, a, is the fraction of variation that is evolutionarily independent of variation in other directions of phenotype space. It measures the degree to which evolution in this direction can be unconstrained by potential stabilizing selection in other directions. Modular variation along a direction would thus have an autonomy of one; if the variation along a direction is completely integrated with variation in other directions, it will have autonomy of zero.

Other measures of autonomy and integration have been proposed. Cheverud *et al.* (1983) and Wagner (1984) proposed a coefficient of integration, $I = 1 - \mathrm{Ge}[\lambda]$, where $\mathrm{Ge}[\lambda]$ is the geometric mean of the eigenvalues of the phenotypic (or genetic) correlation matrix. Wagner (1984) proposed using the variance of the eigenvalues of the correlation matrix as a measure of integration. These are intuitively sensible measures, but both are calculated from a correlation matrix, which does not directly appear in any standard model of evolutionary dynamics. In contrast, our measure has a precise interpretation as the degree to which selection on traits other than that defined by the selection gradient reduce evolvability. The geometric mean of the eigenvalues of a variance matrix, the generalized variance (Anderson, 1984), has also been proposed as a general measure of variation (Zhivotovsky, 1988) and would, if applied to the **G** matrix, be a candidate measure of evolvability. As shown in Fig. 5, this intuition has partial justification in the similarity of the generalized variance to $\bar{c}$ for low-dimensional traits, although this relationship breaks down for high-dimensional traits. Kirkpatrick (2008) recently suggested using the ratio of the sum of the eigenvalues of a mean-standardized **G** matrix to the largest eigenvalue of that matrix as a measure of the 'effective number of dimensions', n_{D}. In terms of quantities we have discussed, $n_{\mathrm{D}} = k\bar{e}/e_{\max}$, where k is the number of traits studied. This method adopts $e_{\max}$, the variance along Schluter's line of least resistance, as a standard to which overall evolvability is compared. All of these measures clearly capture something about integration or constraint, but exactly what has not been explicitly determined. The advantage of our measures is that they have a precise quantitative meaning.

'Random skewers' (Cheverud *et al.*, 1983; Cheverud, 1996; Cheverud & Marroig, 2007) is a method for comparing **G** matrices that does have a theoretical interpretation in line with our philosophy. The basis for random skewers is the distribution of angles between the response vectors that two **G** matrices generate over a set of random selection gradients, and is thus related to our $\bar{d}$. A 'skewer' can, of course, also be computed for a single selection gradient. Random skewers does not consider

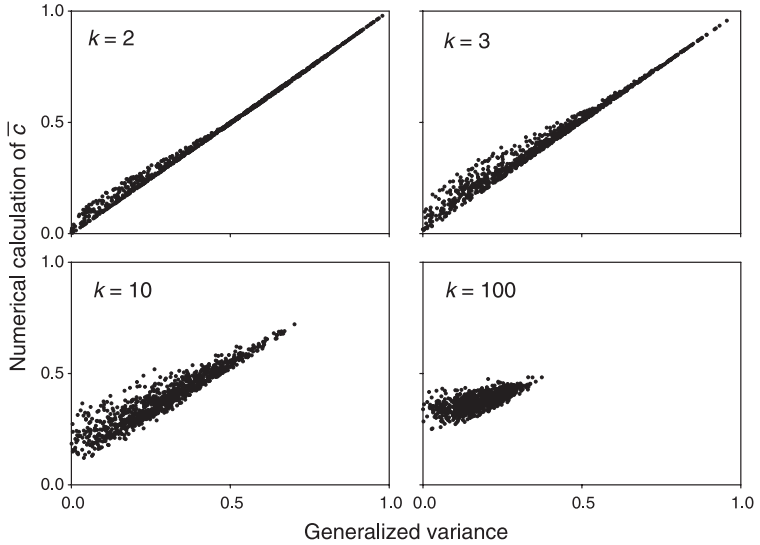


Fig. 5 Mean conditional evolvability, $\bar{c}$, vs. generalized variance for 1000 random **G** matrices of various dimensionalities (k). Calculation of $\bar{c}$ is based on the analytical approximation in Result 3 (Appendix 1). For the two-dimensional **G** matrices ($k = 2$), the generalized variance is exactly equal to $\bar{c}$ and the error is due to our approximation. See the legend of Fig. 2 for an explanation of how the random matrices were generated.

the length of the response vectors, and may therefore judge matrices of very different size to be similar. This is appropriate if matrices maintain proportionality, but matrix size is variable over time. Our proposed alternative, response difference, $\bar{d}$, measures the total difference between the selection responses of the two matrices, and is therefore sensitive to both differences in matrix size and structure.

The question of whether the **G** matrix is stable enough to structure macroevolution is central to evolutionary quantitative genetics. Lande (1976, 1980) developed models that suggested that such stability is plausible, whereas Turelli (1984, 1985, 1988b) suggested that some of Lande's assumptions are incorrect and that **G** matrices would be unlikely to remain stable over evolutionary time. Clearly, the stability of **G** is an important empirical question (Lande, 1988; Turelli, 1988a), and this has led to attempts at comparing **G** matrices in related species (see Steppan *et al.*, 2002, for review). These efforts have, however, often not been built on a principled theory for what constitutes a valid comparison. For example, methods such as matrix correlations are often used. We submit that these methods tell us next to nothing about whether matrices differ in a meaningful manner and should simply not be used. Methods such as factor analysis (Zelditch, 1987; Mitteroecker & Bookstein, 2007) and common principal components analysis (Flury, 1988; Phillips & Arnold, 1999) are easier to interpret geometrically and should be included in most comparisons of **G** matrices, but they still lack a foundation in biological theory and do not provide direct comparison of evolvabilities.

A crucial decision in any multivariate analysis involving traits on different natural scales is how changes in different traits are to be compared. Clearly, all traits must be put on a comparable scale, such as a mean- or variance-standardized scale. Although we have developed the theory for both of these cases, we believe that mean standardization is usually preferable for data that are on a ratio or log-interval scale. Mean standardization preserves the full set of comparisons that are valid for ratio- or log-interval-scale data, whereas variance standardization does not. More specifically, evolvability depends on variation, and if we use a variance scale, we are including some of what we want to measure in our measuring stick. The result is measures of evolvability and response to selection that are not interpretable without reference to the variance used for the standardization. One cannot fully interpret a response of a 10th of a standard deviation without knowing whether the standard deviation itself is large or small. As discussed elsewhere, this leads to a variety of interpretational problems (Houle, 1992; Hansen *et al.*, 2003b; Hansen & Houle, 2004; Hereford *et al.*, 2004). On a log-interval scale, variance scaling is nonsensical.

These problems are particularly acute when one tries to interpret the elements of the breeder's equation as measures of evolvability and selection. In the univariate Lande equation, $\Delta\bar{z} = G\beta = G(\mathrm{cov}[w,z]/P)$, G measures the amount of evolutionarily relevant variation, whereas β measures how mean fitness changes with trait mean. The selection gradient is the ratio of the covariance between the trait z and relative fitness to the total phenotypic variance, P. The univariate breeder's equation $\Delta\bar{z} = h^2S = (G/P)\mathrm{cov}[w,z]$ is a rearrangement of the Lande equation into a dimensionless quantity, h^2, that expresses the proportion of the total variation that is evolutionarily relevant and the selection differential S, which is the covariance between the selected trait and relative fitness. S contains both information about the form of directional selection and information about the amount of variation, thereby confounding variation and selection. Both a weak selection gradient in a population with abundant variation and a strong selection gradient in a population with little variation would lead to the same S. As expected from these elementary facts about the breeder's equation and the fact that there are very strong correlations between additive variances and phenotypic variances, heritabilities and selection differentials are also strongly correlated. Heritabilities are therefore poor predictors of evolvability (Houle, 1992; Hansen *et al.*, 2003b).

All of these problems carry over to the multivariate breeder's equation $\boldsymbol{\Delta\bar{z}} = (\mathbf{GP}^{-1})\mathbf{S}$, where the elements of $\mathbf{GP}^{-1}$are dimensionless and all scale information is in **S**. In addition, as noted by Lande & Arnold (1983), the elements of **S** reflect both direct and indirect effects of selection, whereas **β** corrects for indirect effects of known characters. The selection gradient is thus better suited to the representation of selection as a cause of change; **S** gives the effects of selection. For example, when selection acts only on a single trait, **β** will have a single nonzero element, whereas, in general, every element of **S** will be nonzero.

The multivariate breeder's equation is nevertheless favoured by some. For example, Klingenberg & Leamy (2001) suggested its use because **S** can be precisely represented as deformations of a mean shape in the context of geometric morphometric data. We do not find this a compelling advantage, as any vector, including **β**, can also be represented relative to a mean shape. Although we do not favour the use of the breeder's equation, a standardized version can readily be obtained with the phenotypic standard deviation vector **σ** as

$$\begin{aligned}\boldsymbol{\Delta\bar{z}} \oslash \boldsymbol{\sigma} &= ((\mathbf{G} \oslash (\boldsymbol{\sigma}\boldsymbol{\sigma}'))(\mathbf{P} \oslash (\boldsymbol{\sigma}\boldsymbol{\sigma}'))^{-1})(\mathbf{S} \oslash \boldsymbol{\sigma}) \\ &= (\mathbf{G}_{\boldsymbol{\sigma}}\boldsymbol{\rho}^{-1})(\mathbf{S} \oslash \boldsymbol{\sigma}),\end{aligned}$$

where **ρ** is the phenotypic correlation matrix. In the univariate case, $\rho = 1$, and $\beta\sigma = S/\sigma = i$, where i is 'intensity of selection', but in the multivariate case, $\boldsymbol{\beta} \odot \boldsymbol{\sigma} \neq \mathbf{S} \oslash \boldsymbol{\sigma}$ and $\boldsymbol{\rho} \neq 1$.

When the data obtained are naturally on an interval scale, such that the origin is arbitrary, and ratios of trait values are meaningless, mean scaling (although possible)

does not lead to more interpretable statistics. In these cases, variance standardization may be more informative. Still, having an independent yardstick by which to judge magnitudes would be desirable. One natural possibility is to use the mean difference between populations or species of interest, because mean differences are meaningful.

In deriving these measures of evolvability and constraint, we have focused exclusively on their meaning, neglecting statistical considerations entirely. Measures of uncertainty in our evolvability measures need to be developed. The sampling properties of statistics based on conditional evolvability are likely to be problematic, as they are heavily influenced by the smallest eigenvalues of a **G** matrix, which will have larger relative errors. Furthermore, the means or variances used to standardize data are themselves statistics with their own sampling errors. Additional work is needed to elucidate the sampling properties of the measures we outline.

We have sought to make two major points in this paper. First, we derived measures of evolvability that capture evolutionarily meaningful attributes of the **G** matrix. Our conditional and unconditional evolvabilities measure the potential of a population to evolve in a selected direction under different selective regimes. Our second and broader point is that measurements of quantities that have a clear, quantitative relationship to theory are far more useful than those that do not. Requiring this kind of meaning in our measures raises theoretical and statistical challenges, but these seem a small price to pay for meaning.

Acknowledgments

We thank Bradley Anholt, Mark Blows, Mark Kirkpatrick, Mihaela Pavlicev, Christophe Pélabon and two anonymous reviewers for comments on the manuscript, and Günter Wagner for inspiration and encouragement. This research was supported by grants by NSF grants DEB-0344417 and DEB-0129219.

References

Airoldi, J.P. & Flury, B.D. 1988. An application of common principal component analysis to cranial morphometry of *Microtus californicus* and *M. ochrogaster* (Mammalia, Rodentia). *J. Zool.* **216**: 21–36.

Anderson, T.W. 1984. *An Introduction to Multivariate Statistical Analysis*, 2nd edn. Wiley, New York.

Blows, M.W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. *J. Evol. Biol.* **20**: 1–8.

Blows, M.W. & Higgie, M. 2003. Genetic constraints on the evolution of mate recognition under natural selection. *Am. Nat.* **161**: 240–253.

Blows, M.W., Chenoweth, S.F. & Hine, E. 2004. Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. *Am. Nat.* **163**: 329–340.

Cheetham, A.H., Jackson, J.B.C. & Hayek, L.C. 1993. Quantitative genetics of bryozoan phenotypic evolution. I. Rate tests for random change versus selection in differentiation of living species. *Evolution* **47**: 1526–1538.

Chernoff, B. & Magwene, P.M. 1999. Morphological integration: forty years later. Afterword in: *Morphological Integration* (E.C. Olson & R.L. Miller, eds), pp. 316–360. University of Chicago Press, Chicago, IL.

Cheverud, J.M. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddleback (*S. fuscicollis*) tamarins. *J. Evol. Biol.* **9**: 5–42.

Cheverud, J.M. & Marroig, G. 2007. Comparing covariance matrices: random skewers method compared to the common principal components model. *Genet. Mol. Biol.* **30**: 461–469.

Cheverud, J.M., Rutledge, J.J. & Atchley, W.R. 1983. Quantitative genetics of development: genetic correlations among age-specific trait values and the evolution of ontogeny. *Evolution* **37**: 895–905.

Cheverud, J.M., Wagner, G.P. & Dow, M.M. 1989. Methods for the comparative analysis of variation patterns. *Syst. Zool.* **38**: 201–213.

Coltman, D.W., O'Donoghue, P., Hogg, J.T. & Festa-Bianchet, M. 2005. Selection and genetic (CO)variance in bighorn sheep. *Evolution* **59**: 1372–1382.

Eaton, M.L. 1983. *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.

Estes, S. & Arnold, S.J. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* **169**: 227–244.

Flury, B.D. 1984. Common principal components in *k* groups. *J. Am. Stat. Assoc.* **79**: 892–898.

Flury, B. 1988. *Common Principal Components and Related Multivariate Models*. Wiley, New York.

Foerster, K., Coulson, T., Sheldon, B.C., Pemberton, J.M., Clutton-Brock, T.H. & Kruuk, L.E.B. 2007. Sexually antagonistic genetic variation for fitness in red deer. *Nature* **447**: 1107–1109.

Galis, F., Van Dooren, T.J.M., Feuth, J.D., Metz, J.A.J., Witkam, A., Ruinard, S., Steigenga, M.J. & Wijnaendts, L.C.D. 2006. Extreme selection in humans against homeotic transformations of cervical vertebrae. *Evolution* **60**: 2643–2654.

Hallgrímsson, B. & Hall, B.K. 2005. *Variation*. Elsevier Academic, Amsterdam.

Hand, D.J. 2004. *Measurement Theory and Practice: The World Through Quantification*. Arnold, London.

Hansen, T.F. 2003. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems* **69**: 83–94.

Hansen, T.F. 2006. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* **37**: 123–157.

Hansen, T.F. & Houle, D. 2004. Evolvability, stabilizing selection, and the problem of stasis. In: *The Evolutionary Biology of Complex Phenotypes* (M. Pigliucci & K. Preston, eds), pp. 130–150. Oxford University Press, Oxford.

Hansen, T.F., Armbruster, W.S., Carlson, M.L. & Pélabon, C. 2003a. Evolvability and genetic constraint in *Dalechampia blossoms*: genetic correlations and conditional evolvability. *J. Exp. Zool. B Mol. Dev. Evol.* **296B**: 23–39.

Hansen, T.F., Pélabon, C., Armbruster, W.S. & Carlson, M.L. 2003b. Evolvability and genetic constraint in *Dalechampia blossoms*: components of variance and measures of evolvability. *J. Evol. Biol.* **16**: 754–766.

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

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204.

Houle, D., Mezey, J. & Galpern, P. 2002. Interpreting the results of common principal component analyses. *Evolution* **56**: 433–440.

Houle, D., Mezey, J., Galpern, P. & Carter, A. 2003. Automated measurement of *Drosophila* wings. *BMC Evol. Biol.* **3**: 25.

Hunt, G. 2007. Evolutionary divergence in directions of high phenotypic variance in the ostracode genus *Poseidonamicus*. *Evolution* **61**: 1560–1576.

Jensen, H., Sæther, B.-E., Ringsby, T.H., Tufto, J., Griffith, S.G. & Ellegren, H. 2003. Sexual variation in heritability and genetic correlations of morphological traits in house sparrow (*Passer domesticus*). *J. Evol. Biol.* **16**: 1296–1307.

Jones, A.G., Arnold, S.J. & Bürger, R. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**: 1747–1760.

Jones, A.G., Arnold, S.J. & Bürger, R. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. *Evolution* **58**: 1639–1654.

Jones, A.G., Arnold, S.J. & Bürger, R. 2007. The mutation matrix and the evolution of evolvability. *Evolution* **61**: 727–745.

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

Kirkpatrick, M. 2008. Patterns of quantitative genetic variation in multiple dimensions. *Genetica*.

Klingenberg, C.P. & Leamy, L.J. 2001. Quantitative genetics of geometric shape in the mouse mandible. *Evolution* **55**: 2342–2352.

Lande, R. 1976. The maintenance of genetic variability by mutation in a polygenic character with linked loci. *Genet. Res.* **26**: 221–235.

Lande, R. 1977. On comparing coefficients of variation. *Syst. Zool.* **26**: 214–217.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain:body size allometry. *Evolution* **33**: 402–416.

Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**: 203–215.

Lande, R. 1988. Quantitative genetics and evolutionary theory. In: *Proceedings of the Second International Conference on Quantitative Genetics* (B.S. Weir, E.J. Eisen, M.M. Goodman & G. Namkoong, eds), pp. 71–84. Sinauer, Sunderland, MA.

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

van der Linde, K. & Houle, D. 2008. A phylogeny of the genus *Drosophila* and some closely related genera. *J. Insect Syst.*, accepted.

Lynch, M. & Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.

Magwene, P.M. 2001. New tools for studying integration and modularity. *Evolution* **55**: 1717–1733.

Marroig, G. & Cheverud, J.M. 2005. Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in new world monkeys. *Evolution* **59**: 1128–1142.

McGuigan, K., Chenoweth, S.F. & Blows, M.W. 2005. Phenotypic divergence along lines of genetic variance. *Am. Nat.* **165**: 32–43.

Mezey, J.G. & Houle, D. 2003. Comparing G matrices: are common principal components informative? *Genetics* **165**: 411–425.

Mezey, J.G. & Houle, D. 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* **59**: 1027–1038.

Mitteroecker, P. & Bookstein, F. 2007. The conceptual and statistical relationship between modularity and morphological integration. *Syst. Biol.* **56**: 818–836.

Olson, E.D. & Miller, R.L. 1958. *Morphological Integration*. University of Chicago Press, Chicago, IL.

Phillips, P.C. & Arnold, S.J. 1999. Hierarchical comparison of genetic variance–covariance matrices. I. Using the Flury hierarchy. *Evolution* **53**: 1506–1515.

Rønning, B., Jensen, H., Moe, B. & Bech, C. 2007. Basal metabolic rate: heritability and genetic correlations with morphological traits in the zebra finch. *J. Evol. Biol.* **20**: 1815–1822.

Rolff, J., Armitage, S.A.O. & Coltman, D.W. 2005. Genetic constraints and sexual dimorphism in immune defence. *Evol. Dev.* **59**: 1844–1850.

Sarle, W.S. 1997. Measurement theory: frequently asked questions. Available at: ftp://ftp.sas.com/pub/neural/measurement.html.

Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* **50**: 1766–1774.

Schluter, D. 2000. *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Steppan, S.J., Phillips, P.C. & Houle, D. 2002. Comparative quantitative genetics: evolution of the G matrix. *Trends Ecol. Evol.* **17**: 320–327.

Stevens, S.S. 1946. On the theory of scales of measurement. *Science* **103**: 677–680.

Stevens, S.S. 1959. Measurement, psychophysics and utility. In: *Measurement: Definitions and Theories* (C.W. Churchman & P. Ratoosh, eds), pp. 18–63. Wiley, New York.

Stevens, S.S. 1968. Measurement, statistics, and the schemiparic view. *Science* **161**: 849–856.

Turelli, M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. *Theor. Popul. Biol.* **25**: 138–193.

Turelli, M. 1985. Effects of pleiotropy on predictions concerning mutation–selection balance for polygenic traits. *Genetics* **111**: 165–195.

Turelli, M. 1988a. Phenotypic evolution, constant covariances and the maintenance of additive variance. *Evolution* **42**: 1342–1347.

Turelli, M. 1988b. Population genetic models for polygenic variation and evolution. In: *Proceedings of the Second International Conference on Quantitative Genetics* (B.S. Weir, E.J. Eisen, M.J. Goodman & G. Namkoong, eds), pp. 601–618. Sinauer, Sunderland, MA.

Van Valen, L. 1978. The statistics of variation. *Evol. Theory* **4**: 33–43.

Van Valen, L. 2005. The statistics of variation. In: *Variation* (B. Hallgrímsson & B.K. Hall, eds), pp. 29–47. Elsevier Academic, Amsterdam.

Wagner, G.P. 1984. On the eigenvalue distribution of genetic and phenotypic dispersion matrices: evidence for a nonrandom organization of quantitative character variation. *J. Math. Biol.* **21**: 77–95.

Wagner, G.P. & Altenberg, L. 1996. Perspective: complex adaptations and the evolution of evolvability. *Evolution* **50**: 967–976.

Zelditch, M.L. 1987. Evaluating models of developmental integration in the laboratory rat using confirmatory factor analysis. *Syst. Zool.* **36**: 368–380.

Zhivotovsky, L.A. 1988. Some methods of analysis of correlated characters. In: *Proceedings of the Second International Conference of Quantitative Genetics* (S.J. Arnold, B.S. Weir, E.J. Eisen, M.M. Goodman & G. Namkoong, eds), pp. 423–432. Sinauer, Sunderland, MA.

Received 10 January 2008; revised 23 May 2008; accepted 27 May 2008

Appendix 1: derivation of technical results

Notation. For simplicity, we will use the notation $E[\lambda]$ and $\mathrm{var}[\lambda]$ to denote the mean and variance of eigenvalues, λ. Furthermore, $H[\lambda] = 1/E[1/\lambda]$ is the harmonic mean of λ, $I[\lambda] = \mathrm{var}[\lambda]/E[\lambda]^2$ is the mean-standardized variance of λ and $\mathrm{cov}[\lambda,\delta]$ is the covariance between two sets of eigenvalues λ and δ.

Result 1. Let $\mathbf{z}$ be a k-dimensional trait vector, and let $\mathbf{G}$ be a positive definite additive genetic variance matrix for $\mathbf{z}$. Let $y = \boldsymbol{\beta}'\mathbf{z}$ and $\mathbf{x} = \mathbf{A}'\mathbf{z}$ be an orthogonal transformation of the coordinate system such that y is an index trait in the direction of the selection gradient, $\boldsymbol{\beta}$. When $|\boldsymbol{\beta}| = 1$, the response to selection in y when $\mathbf{x}$ is held constant is $\Delta\bar{y} = \boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}(1 - R^2) = (\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}$, where R is the multiple correlation coefficient between y and $\mathbf{x}$.

Proof. We transform the trait vector, $\mathbf{z}$, to a new coordinate system where one axis is the gradient, which is now a normed coordinate vector, $\mathbf{b}$, and the other axes are orthogonal to $\mathbf{b}$. Each of the $k - 1$ orthonormal axes can be defined by a normed coordinate vector $\mathbf{a}_i$ such that $\mathbf{a}_i'\mathbf{b} = 0$ and $\mathbf{a}_i'\mathbf{a}_j = 0$ for $i \neq j$. We can collect the $\mathbf{a}_i$ as columns in an $k \times (k - 1)$ matrix $\mathbf{A}$, such that the matrix $[\mathbf{b},\mathbf{A}]$ is an orthogonal transformation of the coordinate system for $\mathbf{z}$. The $\mathbf{A}$ matrix is not unique, but the results we are to derive are invariant with respect to $\mathbf{A}$.

Following Hansen (2003), the change in y under unit directional selection when $\mathbf{x}$ is kept fixed is given as $\mathbf{G}_{y|x} = \mathbf{G}_y - \mathbf{G}_{yx}\mathbf{G}_x^{-1}\mathbf{G}_{xy}$. In terms of the variance matrix, $\mathbf{G}$, of the original traits we have

$$\begin{aligned}\mathbf{G}_y &= \mathbf{b}'\mathbf{G}\mathbf{b}\\ \mathbf{G}_{yx} &= \mathbf{b}'\mathbf{G}\mathbf{A}\\ \mathbf{G}_{xy} &= \mathbf{A}'\mathbf{G}\mathbf{b}\\ \mathbf{G}_x &= \mathbf{A}'\mathbf{G}\mathbf{A}\end{aligned}$$

This gives

$$\begin{aligned}\Delta\bar{y} &= \mathbf{b}'\mathbf{G}\mathbf{b}(1 - (\mathbf{b}'\mathbf{G}\mathbf{b})^{-1}(\mathbf{b}'\mathbf{G}\mathbf{A})(\mathbf{A}'\mathbf{G}\mathbf{A})^{-1}(\mathbf{A}'\mathbf{G}\mathbf{b}))\\ &= \mathbf{b}'\mathbf{G}\mathbf{b}(1 - R^2),\end{aligned}$$

where R is the multiple correlation coefficient of y with $\mathbf{x}$. The multiple correlation coefficient is invariant under any affine transformation of the coordinate system (Eaton, 1983, Proposition 10.1) and therefore to the choice of $\mathbf{A}$ (provided orthogonality with $\mathbf{b}$ is preserved).

The second part of the equality follows from the observation that

$$1 - R^2 = \det[[\mathbf{b},\mathbf{A}]'\mathbf{G}[\mathbf{b},\mathbf{A}]]/(\det[\mathbf{A}'\mathbf{G}\mathbf{A}](\mathbf{b}'\mathbf{G}\mathbf{b}))$$

(see Anderson, 1984, p. 40). The determinant $\det[[\mathbf{b},\mathbf{A}]'\mathbf{G}[\mathbf{b},\mathbf{A}]] = \det[[\mathbf{b},\mathbf{A}]]^2\ \det[\mathbf{G}] = \det[\mathbf{G}]$, and $\det[\mathbf{A}'\mathbf{G}\mathbf{A}]$ can be computed from the fact that it is the first principal cofactor of $[\mathbf{b},\mathbf{A}]'\mathbf{G}[\mathbf{b},\mathbf{A}]$. If M_{ij} is the ij cofactor of a matrix $\mathbf{M}$, then $M_{ij} = [\mathbf{M}^{-1}]_{ij}\det[\mathbf{M}]$. To use this, we note that $([\mathbf{b},\mathbf{A}]'\mathbf{G}[\mathbf{b},\mathbf{A}])^{-1} = [\mathbf{b},\mathbf{A}]'\mathbf{G}^{-1}[\mathbf{b},\mathbf{A}]$, because $[\mathbf{b},\mathbf{A}]$ is orthogonal, so that $[[\mathbf{b},\mathbf{A}]'\mathbf{G}^{-1}[\mathbf{b},\mathbf{A}]]_{11} = \mathbf{b}'\mathbf{G}^{-1}\mathbf{b}$. This gives $\det[\mathbf{A}'\mathbf{G}\mathbf{A}] = \det[\mathbf{G}](\mathbf{b}'\mathbf{G}^{-1}\mathbf{b})$. Therefore,

$$\begin{aligned}(1 - R^2) &= \det[\mathbf{G}]/(\det[\mathbf{G}](\mathbf{b}'\mathbf{G}^{-1}\mathbf{b})(\mathbf{b}'\mathbf{G}\mathbf{b}))\\ &= 1/((\mathbf{b}'\mathbf{G}^{-1}\mathbf{b})(\mathbf{b}'\mathbf{G}\mathbf{b})).\end{aligned}$$

Result 2. If the vector $\boldsymbol{\beta}$ is symmetrically distributed in k-dimensional space with $E[|\boldsymbol{\beta}|] = 1$, then $E[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] = \mathrm{Trace}[\mathbf{G}]/k = \sum_i\lambda_i/k$, where λ_i are eigenvalues of the positive definite matrix $\mathbf{G}$, and the summation is over all k eigenvalues.

Proof. This result follows from a slight modification of the proof of Lemma 1, if we note that only $E[|\boldsymbol{\beta}|] = 1$ and not $|\boldsymbol{\beta}| = 1$ is required for $E[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] = \sum_i\lambda_i/k$.

Result 3. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and $\mathbf{G}$ is of full rank, then

$$E[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}] \approx H[\lambda](1 + I[1/\lambda]/(k + 1)).$$

Proof. We begin with a standard approximation for the expectation of an inverse (Lynch & Walsh, 1998, appendix 1) to get

$$E[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}] \approx \frac{1}{E[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]}(1 + I[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]).$$

Fitting in the appropriate moments from Lemma 1 then gives the result.

Result 4. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and the $k \times k$ matrix $\mathbf{M}$ is of full rank, then

$$E\left[\sqrt{\boldsymbol{\beta}'\mathbf{M}\boldsymbol{\beta}}\right] \approx \sqrt{E[\lambda^2]}\left(1 - \frac{I[\lambda^2]}{8(k + 1)}\right),$$

where λs are the eigenvalues of $\mathbf{M}$.

Proof. A standard approximation for the expectation of the square root (Lynch & Walsh, 1998, appendix 1) gives

$$\mathrm{E}\left[\sqrt{\boldsymbol{\beta}'\mathbf{M}\boldsymbol{\beta}}\right] \approx \sqrt{\mathrm{E}[\boldsymbol{\beta}'\mathbf{M}\boldsymbol{\beta}]}\left(1-\frac{\mathrm{I}[\boldsymbol{\beta}'\mathbf{M}\boldsymbol{\beta}]}{8(k+1)}\right).$$

We can now fit in moments from Lemma 1 to get our result.

Result 5. If $\boldsymbol{\beta}$ is uniformly distributed on the surface of a unit hypersphere of dimension k, and $\mathbf{G}$ is of full rank, then

$$\mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}] \approx \frac{\mathrm{H}[\lambda]}{\mathrm{E}[\lambda]}\left(1+\frac{\mathrm{I}[\lambda]+\mathrm{I}[1/\lambda]+1-\mathrm{H}[\lambda]/\mathrm{E}[\lambda]}{k+1}\right).$$

Proof. A standard approximation for the expectation of a ratio (Lynch & Walsh, 1998, appendix 1) gives

$$\mathrm{E}\left[\frac{(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}}{\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}}\right] \approx \frac{\mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}]}{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]} \times\left(1+\frac{\mathrm{var}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]^2}-\frac{\mathrm{cov}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1},\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}{\mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}]\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}\right).$$

The covariance in this equation is

$$\begin{aligned}\mathrm{cov}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1},\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] &= \mathrm{E}\left[\frac{\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}}\right]-\mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}]\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]\\ &\approx \frac{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]}\left(1+\frac{\mathrm{var}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]}{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]^2}-\frac{\mathrm{cov}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta},\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}{\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta}]\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}]}\right)\\ &\quad-\mathrm{E}[(\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta})^{-1}]\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}],\end{aligned}$$

where we have used the same approximation for the expectation of a ratio. We can now fit in moments derived in Lemma 1 and Result 3 to obtain our result. In doing so, note that

$$\mathrm{cov}[\boldsymbol{\beta}'\mathbf{G}^{-1}\boldsymbol{\beta},\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] = (1-\mathrm{E}[\lambda]\mathrm{E}[1/\lambda])/(k+1).$$

Lemma 1. If the vector $\boldsymbol{\beta}$ is uniformly distributed on the unit k-sphere, $\mathbf{G}$ and $\mathbf{D}$ are positive definite matrices of rank k and $\mathbf{GD} = \mathbf{DG}$, then

$$\begin{aligned}\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] &= E[\lambda],\\ \mathrm{var}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] &= \frac{\mathrm{var}[\lambda]}{k+1},\\ \mathrm{cov}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta},\boldsymbol{\beta}'\mathbf{D}\boldsymbol{\beta}] &= \frac{\mathrm{cov}[\lambda,\delta]}{k+1},\end{aligned}$$

where λ are eigenvalues of $\mathbf{G}$, δ are eigenvalues of $\mathbf{D}$ and the summation is over all k eigenvalues.

Proof. We diagonalize $\mathbf{G}$ with the transformation $\mathbf{Ca} = \boldsymbol{\beta}$, where $\mathbf{C}$ is an orthogonal matrix, as

$$\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta} = \mathbf{a}'\mathbf{C}'\mathbf{G}\mathbf{C}\mathbf{a} = \mathbf{a}'\boldsymbol{\Lambda}\mathbf{a} = \sum_{\mathbf{i}}\mathbf{a_i^2}\lambda_{\mathbf{i}},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\mathbf{G}$. Then

$$\mathrm{E}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta}] = \sum_i \mathrm{E}[a_i^2]\lambda_i = \sum_i \lambda/k,$$

because $\mathrm{E}[|\boldsymbol{\beta}|] = 1$ implies $\sum_i \mathrm{E}[a_i^2] = 1$, and symmetry implies that $\mathrm{E}[a_i^2] = \mathrm{E}[a_j^2] = 1/k$ for all i and j. Note also that we only need $|\boldsymbol{\beta}| = 1$ in expectation for this result, and it therefore holds for any standardized symmetric distribution.

To compute the variance and covariance, we will need the moments $\mathrm{var}[a_i^2]$ and $\mathrm{cov}[a_i^2, a_j^2]$. We start with $\mathrm{var}[a_i^2] = \mathrm{E}[a_i^4] - \mathrm{E}[a_i^2]^2$. We can obtain the fourth moment by representing the uniform distribution on the surface of the sphere with a normalized spherical distribution (i.e. a standardized MVN). Let r_i be independent variables each with a standardized normal distribution. Then r_i^2 has a $\chi^2(1)$ distribution, and the variable $r_i^2/\sum_i r_i^2$, which has a Beta[1,$n-1$] distribution, will equal a_i^2. The fourth moment of a_i is thus the second moment of this beta distribution such that $\mathrm{E}[a_i^4] = 2/k(k+1)$, which gives

$$\mathrm{var}[a_i^2] = \frac{k-1}{k^2(k+1)}.$$

The covariance can be computed from the relation $a_i^2 = 1 - \sum_{j\neq i} a_j^2$ and from the fact that the covariances by symmetry must be the same for all i and j.

$$\begin{aligned}\mathrm{cov}[a_i^2, a_j^2] &= \sum_{j\neq i}\mathrm{cov}[a_j^2, a_j^2]/(k-1)\\ &= \mathrm{cov}\left[a_j^2, \sum_{j\neq i} a_j^2\right]/(k-1)\\ &= \mathrm{cov}[a_j^2, 1-a_i^2]/(k-1)\\ &= -\mathrm{var}[a_i^2]/(k-1) = \frac{-1}{k^2(k+1)}.\end{aligned}$$

If $\mathbf{G}$ and $\mathbf{D}$ commute, they can be simultaneously diagonalized with the same orthogonal matrix $\mathbf{C}$, and the covariance between the two quadratic forms can be written

$$\begin{aligned}\mathrm{cov}[\boldsymbol{\beta}'\mathbf{G}\boldsymbol{\beta},\boldsymbol{\beta}'\mathbf{D}\boldsymbol{\beta}] &= \sum_i\sum_j \lambda_i\delta_j\ \mathrm{cov}[a_i^2, a_j^2]\\ &= \sum_i \lambda_i\delta_i\ \mathrm{var}[a_i^2] + \sum_i\sum_{j\neq i}\lambda_i\delta_j\ \mathrm{cov}[a_i^2, a_j^2].\end{aligned}$$

Fitting in the above results for $\mathrm{var}[a_i^2]$ and $\mathrm{cov}[a_i^2, a_j^2]$, we get

$$\frac{\mathrm{cov}[\lambda,\delta]}{k+1},$$

and the variance follows by setting $\mathbf{G} = \mathbf{D}$.

Appendix 2: a worked example

To illustrate how our measures of evolvability are calculated, and how they can be interpreted, consider the two hypothetical three-trait **G** matrices in Table A1. We chose this simple example to represent some typical problems in inferring and comparing evolvability from **G** matrices. Traits 1 and 2 represent lengths of morphological features, and trait 3 represents the life-history trait fecundity. Comparison of the matrices themselves does not immediately suggest how each population will respond to selection. The diagonals suggest that traits 1 and 3 might respond more rapidly to selection in population 2 and that trait 2 would respond better in population 1. The degree of correlation among traits seems a bit higher in population 2 than population 1. It is clear, however, that more than a glance is necessary to ascertain which population would evolve more rapidly under particular circumstances. Furthermore, note that the units are not commensurate across all the traits; so, the raw numerical values cannot sensibly be compared when the directions of response are not the same. In addition, note that traits and populations vary in the relationships between trait means and variances; so, the appropriate standardization for each matrix is different.

Table A2 shows the example matrices standardized by trait means ($\mathbf{G}_\mu$), trait variances ($\mathbf{G}_\sigma$) and the square root of the **P** matrix ($\mathbf{G}_\mathbf{P}$). The diagonal of $\mathbf{G}_\mu$ are I_A values of each trait. The diagonals of $\mathbf{G}_\sigma$ are the heritabilities of each trait.

Table A3 gives some selection–response statistics when a single gradient is applied to both the example populations. The selection gradient giving change in relative fitness per unit change in trait is $\boldsymbol{\beta}' = [0.005/\text{mm}, -0.001/\text{mm}, 0.10/\text{egg}]$. These values were chosen to yield standardized $\boldsymbol{\beta}$ values that are in line with typical standardized strengths of selection (Hereford *et al.*, 2004). Table A4 shows the evolvability statistics developed in this paper, which are based on response to a $\boldsymbol{\beta}$ in the same direction, but standardized to length 1 on whichever scale the parameter estimates are on. Note that this means that the 'standard' strength of selection is different for each standardization.

The first section of each table gives the evolvability statistics for unstandardized data, where the units are a mixture of egg numbers and millimetres. Although the statistics are readily calculated, we see no useful interpretation of any of our statistics on this dog's breakfast scale. The dimensionless ratios of evolvabilities, *e*, and conditional evolvabilities, *c*, shown in the 'compare' column in Table A4, however, do have value in expressing the relative progress possible under selection. We do not show the ratio of respondabilities, *r*, as these are measured in different directions and are therefore not comparable. If the gradient $\boldsymbol{\beta}$ had included only one nonzero element, indicating that all the selection was on a single trait, the fact that the *e* and *c* each summarize only the response in the selection direction would give them the units of the single selected trait, and these values could be interpreted. The value of the response difference, *d*, is difficult to interpret, as it is a distance along a different direction in phenotype space from $\boldsymbol{\beta}$ and thus has different units.

The interpretability of these statistics increases on a mean-standardized scale. The individual elements of the response vector shown in Table A3 are in proportions of the mean of each trait. In this coordinate system, the selection gradient results in an average unconstrained change of 1.1% in population 1 and 1.8% in population 2 in the direction of $\boldsymbol{\beta}$. The respondability, *r*, in population 1 is thus 59% of that in population 2. Turning to evolvability, *e*, if the mean-standardized selection gradient had been of unit length in each population, corresponding to a strength of selection equal to that on fitness, and no stabilizing selection occurred, then the response would have been 1.2% in population 1 and 2.4% in population 2. The evolvability in population 1 would have been 50% of that in population 2. The ratio of the projections of $\Delta\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because populations 1 and 2 have different means; so, standardizing the selection gradients by their own means changes the relative size of the selection gradients applied. This suggests that it may sometimes be useful to standardize both gradients and **G** by common values, as outlined above for the expected response difference, $\bar{d}$.

On the mean scale, autonomies, $a(\boldsymbol{\beta})$, the ratio of conditional evolvability to evolvability, are 13% in population 1 but only 1.2% in population 2. The integration values, $i(\boldsymbol{\beta})$, are $1 - a(\boldsymbol{\beta})$; so, population 1 is 87% integrated and population 2 is 98.8% integrated in these directions. Despite the larger unconstrained evolvability of population 2 in direction $\boldsymbol{\beta}$, evolution would therefore be much more constrained by stabilizing selection on the remaining traits in population 2 than in 1. The conditional response is nearly 16 times as large in population 1 as in population 2. This is reflected by the difference in the angle of the unconstrained response relative to $\boldsymbol{\beta}$, 18° in population 1 and 36° in population 2. When the other traits are under stabilizing selection, this increased deflection will be counteracted and the constrained response reduced. The angle between the response vectors in the two populations is 18°. When we standardize by the average of the mean vectors response difference, $d(\boldsymbol{\beta})$, is 0.8%, which is a substantial proportion of the direct responses.

On the variance-standardized scale, element-wise standardization places the individual elements of the response vectors in standard deviation units. Respondability, *r*, is 3.1% of a standard deviation in this coordinate system in population 1 and 5.1% of a standard deviation in population 2 in direction $\boldsymbol{\beta}$. The ratio of the projections of $\Delta\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because of the different standardizations employed in the two populations. The autonomies, $a(\boldsymbol{\beta})$, on the standard deviation scale are

higher in both populations than those on the mean-standardized scale, particularly in population 2, with the result that conditional evolvabilities, $c(\boldsymbol{\beta})$, are very similar. The angles between the responses and $\boldsymbol{\beta}$ are higher in population 1 than on the mean-standardized scale, as is the angle between response vectors, θ_d. When we use element-wise variance standardization, the response difference, $d(\boldsymbol{\beta})$, is 5.1% of a standard deviation, about as large as the direct responses.

Standardization with the square root of the **P** matrix places the lengths of response vectors and evolvabilities in standard deviation units appropriate to their direction. For example, the evolvability, $e(\boldsymbol{\beta})$, in population 1 is 9.5% of the phenotypic standard deviation in direction $\boldsymbol{\beta}$. The oblique transformation of the coordinate system makes the elements of the response vectors difficult to interpret. In this case, **P** standardization results in similar vectors and scalar measures of evolvability to σ standardization.

The many differences between the statistics calculated on different scales make clear that the choice of scale can strongly influence the results. Each standardization gives a unique weighting of the traits that stretches or compresses each of the bases of phenotype space to a different degree. In addition, the square-root-of-**P** transformation also performs an oblique rotation of the bases.

Finally, we can compare the evolvability statistics over the entire phenotype space. Table A5 shows the mean evolvability, $\bar{e}$, conditional evolvability, $\bar{c}$, respondability, $\bar{r}$, and autonomy, $\bar{a}$, values for the two hypothetical populations. The average unconditional evolvability, $\bar{e}$, on a mean-standardized scale is 0.5% in population 1 and 1.4% in population 2. The average conditional evolvability, $\bar{c}$, is 0.18% of the mean in population 1 but just 0.03% in population 2. This difference is reflected in the lower autonomies, $\bar{a}$, in population 2. The raw and standard deviation scales show a similar pattern, in which population 2 is more unconditionally evolvable but also more constrained than population 1. The key cause of this result is that the eigenvalues of **G** matrix 2 are more uneven than those in **G** matrix 1.

Table A1 Example **G** and the residual matrix $\mathbf{E} = \mathbf{P}-\mathbf{G}$ matrices and trait means.

Population	Trait (units)	**G**	**E**	$\bar{\mathbf{z}}$
1	1 (mm) 2 (mm) 3 (eggs)	$\begin{bmatrix} 10 & 10 & 20 \\ & 30 & 20 \\ & & 80 \end{bmatrix}$	$\begin{bmatrix} 10 & 13 & 50 \\ & 30 & 40 \\ & & 890 \end{bmatrix}$	$\begin{bmatrix} 73 \\ 138 \\ 82 \end{bmatrix}$
2	1 (mm) 2 (mm) 3 (eggs)	$\begin{bmatrix} 20 & 16 & -10 \\ & 20 & 20 \\ & & 150 \end{bmatrix}$	$\begin{bmatrix} 20 & 20 & 20 \\ & 50 & 100 \\ & & 600 \end{bmatrix}$	$\begin{bmatrix} 80 \\ 152 \\ 64 \end{bmatrix}$

Table A2 Example **G** matrices from Table A1 standardized by trait means and variances.

Population	$\mathbf{G}_\mu \times 100$	$\mathbf{G}_\sigma$	$\mathbf{G}_\mathbf{P}$
1	$\begin{bmatrix} 0.19 & 0.10 & 0.33 \\ & 0.16 & 0.18 \\ & & 1.19 \end{bmatrix}$	$\begin{bmatrix} 0.50 & 0.29 & 0.14 \\ & 0.50 & 0.08 \\ & & 0.08 \end{bmatrix}$	$\begin{bmatrix} 0.60 & -0.09 & 0.08 \\ & 0.54 & 0.02 \\ & & 0.07 \end{bmatrix}$
2	$\begin{bmatrix} 0.31 & 0.13 & -0.20 \\ & 0.09 & 0.21 \\ & & 3.66 \end{bmatrix}$	$\begin{bmatrix} 0.50 & 0.30 & -0.06 \\ & 0.29 & 0.09 \\ & & 0.20 \end{bmatrix}$	$\begin{bmatrix} 0.42 & 0.15 & -0.11 \\ & 0.19 & 0.02 \\ & & 0.20 \end{bmatrix}$

Table A3 Standardized selection gradients and response vectors for example populations in Table A1 on the raw and three standardized scales.

Population		$\Delta\bar{\mathbf{z}}$	$\boldsymbol{\beta}_\mu$	$\Delta\bar{\mathbf{z}}_\mu$	$\boldsymbol{\beta}_\sigma$	$\Delta\bar{\mathbf{z}}_\sigma$	$\boldsymbol{\beta}_\mathbf{P}$	$\Delta\bar{\mathbf{z}}_\mathbf{P}$
1	Vector	$\begin{bmatrix} 0.24\text{ mm} \\ 0.22\text{ mm} \\ 0.88\text{ eggs} \end{bmatrix}$	$\begin{bmatrix} 0.37 \\ -0.14 \\ 0.82 \end{bmatrix}$	$\begin{bmatrix} 0.0033 \\ 0.0016 \\ 0.0107 \end{bmatrix}$	$\begin{bmatrix} 0.022 \\ -0.008 \\ 0.311 \end{bmatrix}$	$\begin{bmatrix} 0.054 \\ 0.028 \\ 0.028 \end{bmatrix}$	$\begin{bmatrix} 0.036 \\ 0.016 \\ 0.319 \end{bmatrix}$	$\begin{bmatrix} 0.047 \\ 0.013 \\ 0.025 \end{bmatrix}$
	Length	0.938	0.908	0.011	0.312	0.067	0.321	0.055
2	Vector	$\begin{bmatrix} -0.02\text{ mm} \\ 0.27\text{ mm} \\ 1.43\text{ eggs} \end{bmatrix}$	$\begin{bmatrix} 0.40 \\ -0.15 \\ 0.64 \end{bmatrix}$	$\begin{bmatrix} -0.0002 \\ 0.0018 \\ 0.0223 \end{bmatrix}$	$\begin{bmatrix} 0.032 \\ -0.008 \\ 0.274 \end{bmatrix}$	$\begin{bmatrix} -0.003 \\ 0.031 \\ 0.052 \end{bmatrix}$	$\begin{bmatrix} 0.025 \\ 0.042 \\ 0.268 \end{bmatrix}$	$\begin{bmatrix} -0.011 \\ 0.016 \\ 0.051 \end{bmatrix}$
	Length	1.454	0.770	0.022	0.276	0.061	0.273	0.054

The selection gradient is $\boldsymbol{\beta}' = [0.005/\text{mm}, -0.001/\text{mm}, 0.10/\text{egg}]$ and the length (norm) of this vector is 0.011 in a combination of egg and mm units.

Table A4 Evolvability statistics for the trait $\boldsymbol{\beta}' = [0.005/\text{mm}, -0.001/\text{mm}, 0.10\text{egg}]$.

	Standardization											
	None*			Mean			Standard deviation			Square root of **P**		
	Population			Population			Population			Population		
Statistic	1	2	Compare	1	2	Compare	1	2	Compare	1	2	Compare
$r(\boldsymbol{\beta})$	0.87	1.24	na†	0.0108	0.0181	0.59	0.031	0.051	0.62	0.030	0.051	0.60
$e(\boldsymbol{\beta})$	78	111	0.70	0.0119	0.0236	0.50	0.100	0.184	0.55	0.095	0.188	0.51
$c(\boldsymbol{\beta})$	29	3.8	7.67	0.0046	0.0003	15.91	0.043	0.038	1.14	0.056	0.158	0.36
$a(\boldsymbol{\beta})$	0.38	0.03		0.39	0.01		0.43	0.21		0.59	0.84	
θ‡	22	31	16	18	36	16	62	34	51	56	19	78
$d(\boldsymbol{\beta})$			0.61			0.008			0.051			0.069

The 'compare' column compares the responses in the two populations. For the respondabilities, $r(\boldsymbol{\beta})$, and evolvabilities, $e(\boldsymbol{\beta})$ and $c(\boldsymbol{\beta})$, the comparison is the ratio of the value in population 1 to that in population 2, when each population is standardized with its own vector or matrix. In other rows, both populations are standardized by the average of the standardization vectors or matrices in the two populations. The row labelled θ contains the angles between the response vectors in the two populations, θ_d.

*The units for the responses of each population are a mixture of mm and eggs, and therefore most of these statistics have no clear interpretation.

†The ratio of respondabilities is meaningless on the raw scale.

‡In the columns labelled 1 and 2, this is the angle between $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}\bar{\mathbf{z}}$. In the 'compare' column, it is the angle between the response vectors in the two populations, θ_d.

Table A5 Expectations of evolvability statistics over a uniform distribution of selection gradients in the entire phenotype space for the hypothetical populations.

	Standardization							
	None*		Mean		Variance		**P**	
Statistic	1	2	1	2	1	2	1	2
$\bar{e}$	40.00	63.33	0.0051	0.0135	0.361	0.329	0.402	0.269
$\bar{r}$	51.08	83.66	0.0070	0.0194	0.455	0.410	0.464	0.310
$\bar{c}$	12.40	3.87	0.0018	0.0003	0.123	0.049	0.151	0.178
$\bar{a}$	0.454	0.101	0.460	0.034	0.495	0.243	0.575	0.831

*The units for the responses of each population are a mixture of millimetres and eggs, and these statistics therefore have no clear interpretation.

Corrections to: Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. Journal of Evolutionary Biology 21:1201-1219.

Thomas F. Hansen[*,†] & David Houle[†]

[*]University of Oslo, Department of Biology, Center for Ecological and Evolutionary Synthesis, 0316 Oslo, Norway; e-mail: thomas.hansen@bio.uio.no
[†]Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA; e-mail: dhoule@bio.fsu.edu

This document contains corrected versions of all the material known to be in error in the original paper. Included here are new versions of Figures 2, 4 and 5, and Appendices 1 and 2. The preamble to Appendix 1 contains corrected version of four mean evolvability equations that appeared in the text of the paper.

Fig. 2 Approximation of mean conditional evolvability, $\bar{c}$. The plots show numerically computed $\bar{c}$ plotted against the analytical approximation in Result 3, Appendix 1, for 1000 random **G** matrices of various dimensionalities (k). In all cases, the matrices have random diagonal entries drawn from a uniform [0,1] distribution and zero off-diagonal elements. This is justified as the symmetry of the random selection gradients implies that the results are unaffected by diagonalization. The numerical mean is computed over 10,000 random unit selection gradients.

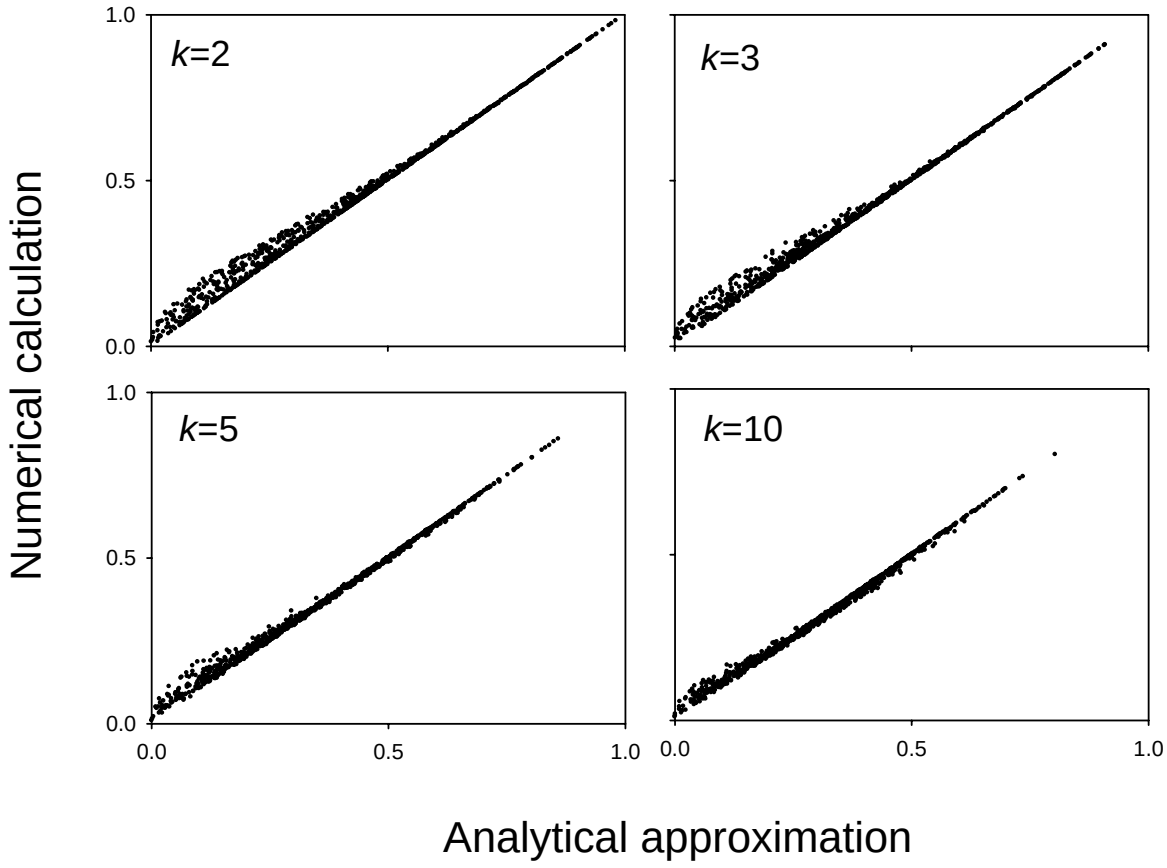


Fig. 4 Unconditional and conditional evolvabilities along the vector of differences in species means for wing shape between *Drosophila melanogaster* and other drosophilid species. The mean conditional and unconditional evolvabilities are shown as dashed horizontal lines. The evolvabilities are in units of centroid size.

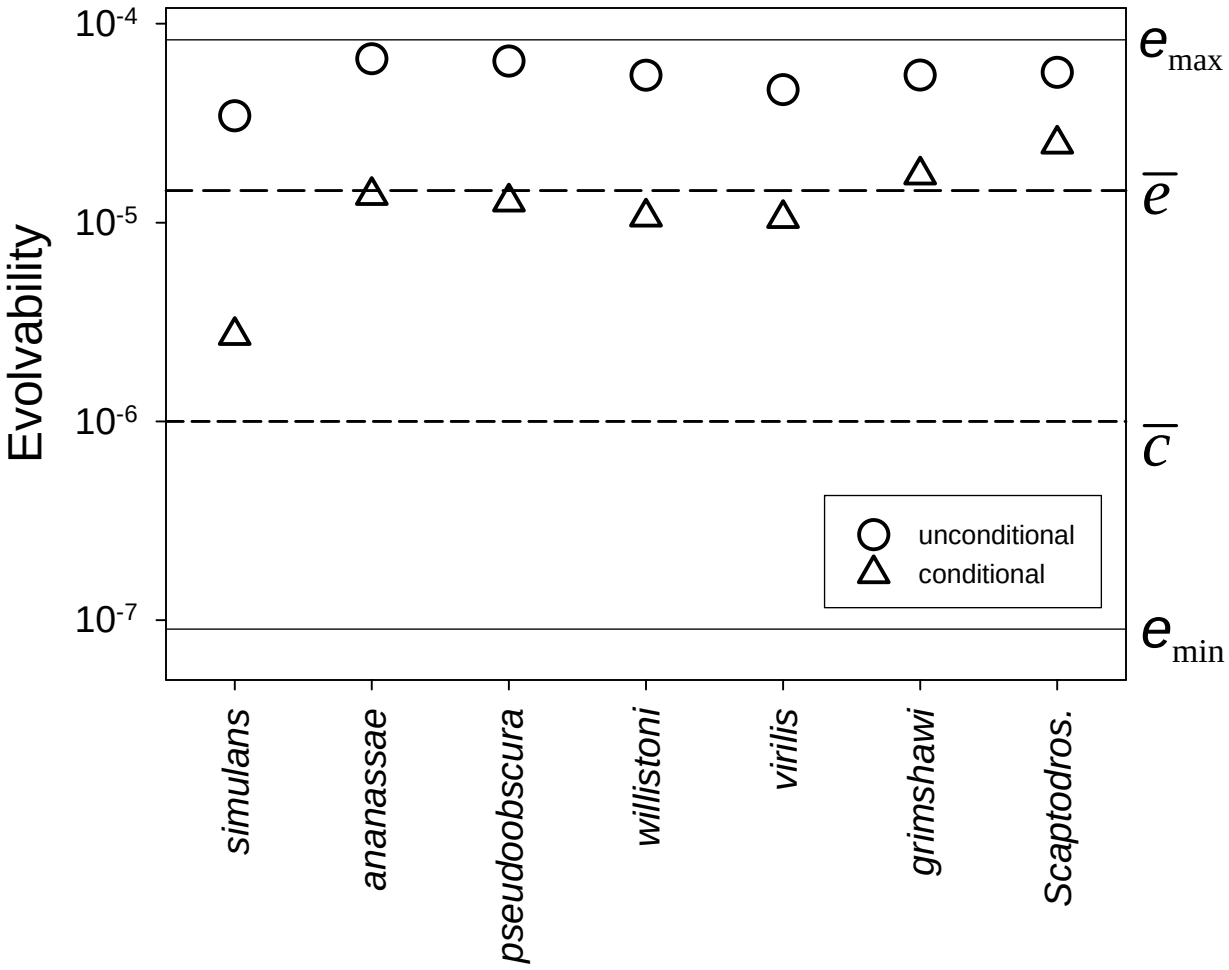
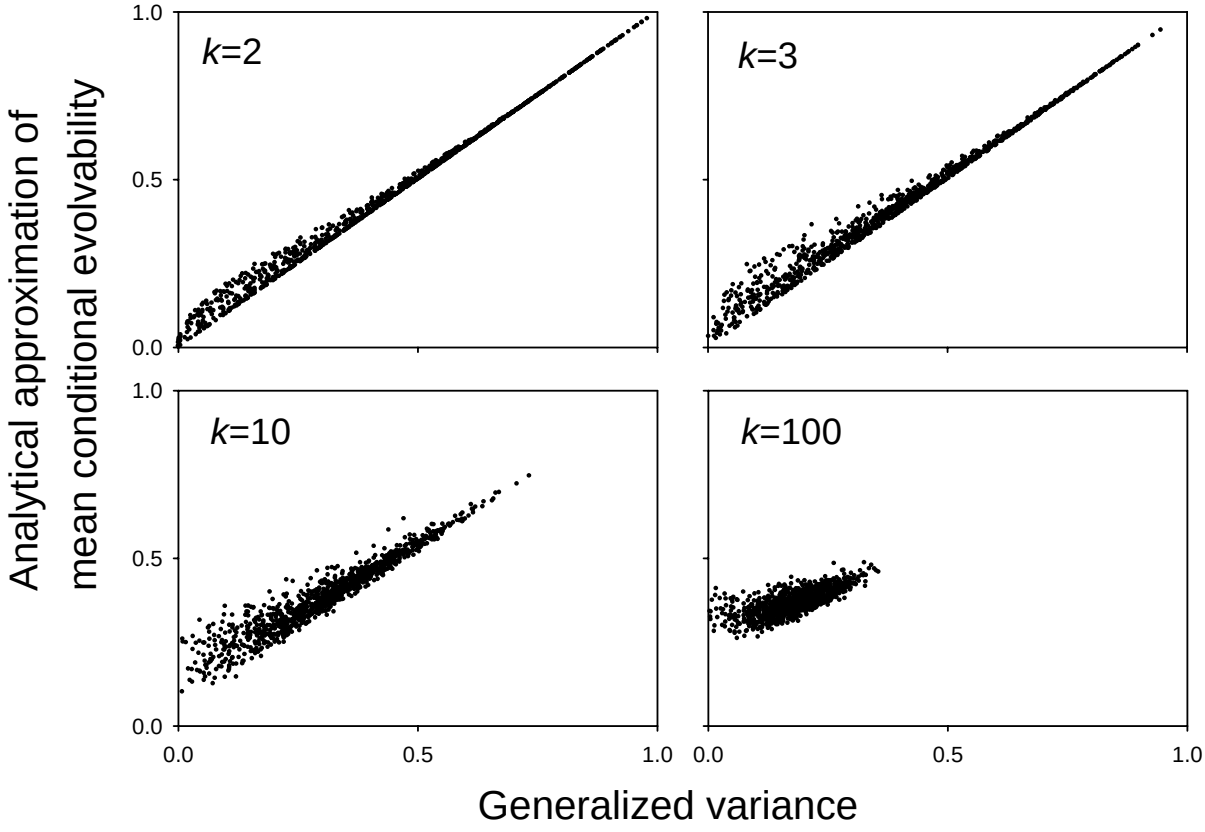

Evolvability
10-4
10-5
10-6
10-7
e_{max}
$\overline{e}$
$\overline{c}$
e_{min}
unconditional
conditional
simulans
ananassae
pseudoobscura
willistoni
virilis
grimshawi
Scaptodros.

Fig. 5 Mean conditional evolvability, $\bar{c}$, versus generalized variance for 1000 random **G** matrices of various dimensionalities (k). Calculation of $\bar{c}$ is based on the analytical approximation in Result 3 (Appendix 1). For the two-dimensional **G** matrices (k=2) the generalized variance is exactly equal to $\bar{c}$ and the error is due to our approximation. See the legend of Fig. 2 for an explanation of how the random matrices were generated.



Appendix 2 (Corrected Jan. 19, 2009)

This is a corrected version of Appendix 2 from Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J. Evol. Biol. 21:1201-1219. John Stinchcombe brought several errors in this Appendix to our attention. In reworking the results we discovered additional errors, including those in Appendix 1. This version corrects all these errors. Changes are highlighted in yellow.

First, the β_3 value used for all calculations was given incorrectly, and was 0.01, rather than 0.1.

Second, the evolvability statistics $\overline{r}$, $\overline{c}$, and $\overline{a}$ in Table A5 were incorrect. One cause of this is that the formulas given in the original paper for $\overline{r}$, $\overline{c}$, and $\overline{a}$ were incorrect, as noted in the correction to Appendix 1. The other cause is that we used a sample-size correction in calculating the variances of functions of eigenvalues. The correct formula to use in such calculations is $\mathrm{Var}\left(f\left(\lambda\right)\right) = \sum_i \left[f\left(\lambda_i\right)\right]^2 \Big/ k - \left[\sum_i f\left(\lambda_i\right) \Big/ k\right]^2$, where λ_i is the *i*th eigenvalue, *f(λ)* is the function of the λs (e.g. $1/\lambda$) whose variance is calculated, and *k* is the dimension of the **G** matrix.

Finally, other calculation errors are corrected in the respondabilities and response differences in Table A4. Several rounding errors have also been corrected, but not highlighted.

Appendix 2: a worked example

To illustrate how our measures of evolvability are calculated, and how they can be interpreted, consider the two hypothetical three-trait **G** matrices in Table A1. We chose this simple example to represent some typical problems in inferring and comparing evolvability from **G** matrices. Traits 1 and 2 represent lengths of morphological features, and trait 3 represents the life-history trait fecundity. Comparison of the matrices themselves does not immediately suggest how each population will respond to selection. The diagonals suggests that traits 1 and 3 might respond more rapidly to selection in population 2 and that trait 2 would respond better in population 1. The degree of correlation among traits seems a bit higher in population 2 than population 1. It is clear, however, that more than a glance is necessary to ascertain which population would evolve more rapidly under particular circumstances. Furthermore, note that the units are not commensurate across all the traits, so the raw numerical values cannot sensibly be compared when the directions of response are not the same. In addition, note that traits and populations vary in the relationships between trait means and variances, so the appropriate standardization for each matrix is different.

Table A2 shows the example matrices standardized by trait means ($\mathbf{G}_{\boldsymbol{\mu}}$), trait variances ($\mathbf{G}_{\boldsymbol{\sigma}}$), and the square root of the **P** matrix ($\mathbf{G}_{\mathbf{P}}$). The diagonal of $\mathbf{G}_{\boldsymbol{\mu}}$ are I_A values of each trait. The diagonals of $\mathbf{G}_{\boldsymbol{\sigma}}$ are the heritabilities of each trait.

Table A3 gives some selection-response statistics when a single gradient is applied to both of the example populations. The selection gradient giving change in relative fitness per unit change in trait is $\boldsymbol{\beta}'$ = [0.005/mm, –0.001/mm, 0.01/egg]. These values were chosen to yield standardized $\boldsymbol{\beta}$ values that are in line with typical standardized strengths of selection (Hereford *et al.,* 2004). Table A4 shows the evolvability statistics developed in this paper, which are based on response to a $\boldsymbol{\beta}$ in the same direction, but standardized to length 1 on whichever scale the parameter estimates are on. Note that this means that the 'standard' strength of selection is different for each standardization.

The first section of each table gives the evolvability statistics for unstandardized data, where the units are a mixture of egg numbers and millimeters. Although the statistics are readily calculated, we see no useful interpretation of any of our statistics on this dog's-breakfast scale. The dimensionless ratios of evolvabilities, e, and conditional evolvabilities, c, shown in the 'compare' column in A4, however, do have value in expressing the relative progress possible under selection. We do not show the ratio of respondabilities, r, as these are measured in different directions and are therefore not comparable. If the gradient $\boldsymbol{\beta}$ had included only one non-zero element, indicating that all the selection was on a single trait, the fact that the e and c each summarize only the response in the selection direction would give them the units of the single selected trait, and these values could be interpreted. The value of the response difference, d, is difficult to interpret, as it is a distance along a different direction in phenotype space from $\boldsymbol{\beta}$ and thus has different units.

The interpretability of these statistics increases on a mean-standardized scale. The individual elements of the response vector shown in Table A3 are in proportions of the mean of each trait. In this coordinate system, the selection gradient results in an average unconstrained change of 1.3% in population 1 and 2.9% in population 2 in the direction of $\boldsymbol{\beta}$. The

respondability, r, in population 1 is thus 43% of that in population 2. Turning to evolvability, e, if the mean-standardized selection gradient had been of unit length in each population, corresponding to a strength of selection equal to that on fitness, and no stabilizing selection occurred, then the response would have been 1.2% in population 1 and 2.4% in population 2. The evolvability in population 1 would have been 50% of that in population 2. The ratio of the projections of $\boldsymbol{\Delta}\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because populations 1 and 2 have different means, so standardizing the selection gradients by their own means changes the relative size of the selection gradients applied. This suggests that it may sometimes be useful to standardize both gradients and $\mathbf{G}$ by common values, as outlined in the text for the expected response difference, $\bar{d}$.

On the mean scale, autonomies, $a(\boldsymbol{\beta})$, the ratio of conditional evolvability to evolvability, are 39% in population 1 but only 1.2% in population 2. The integration values, $i(\boldsymbol{\beta})$, are 1- $a(\boldsymbol{\beta})$, so population 1 is 61% integrated and population 2 is 98.8% integrated in these directions. Despite the larger unconstrained evolvability of population 2 in direction $\boldsymbol{\beta}$, evolution would therefore be much more constrained by stabilizing selection on the remaining traits in population 2 than in 1. The conditional response is nearly 16 times as large in population 1 as in population 2. This is reflected by the difference in the angle of the unconstrained response relative to $\boldsymbol{\beta}$, 18° in population 1 and 36° in population 2. When the other traits are under stabilizing selection, this increased deflection will be counteracted and the constrained response reduced. The angle between the response vectors in the two populations is 18°. When we standardize by the average of the mean vectors response difference, $d(\boldsymbol{\beta})$, is 1.0%, which is a substantial proportion of the direct responses.

On the variance-standardized scale, elementwise standardization places the individual elements of the response vectors in standard-deviation units. Respondability, r, is 21% of a standard deviation in this coordinate system in population 1 and 22% of a standard deviation in population 2 in direction $\boldsymbol{\beta}$. The ratio of the projections of $\boldsymbol{\Delta}\bar{\mathbf{z}}$ on $\boldsymbol{\beta}$ does not equal the ratio of the evolvabilities because of the different standardizations employed in the two populations. The autonomies, $a(\boldsymbol{\beta})$, on the standard-deviation scale are higher in both populations than those on the mean-standardized scale, particularly in population 2, with the result that conditional evolvabilities, $c(\boldsymbol{\beta})$, are very similar. The angles between the responses and $\boldsymbol{\beta}$ are higher in population 1 than on the mean-standardized scale, as is the angle between response vectors, θ_d. When we use elementwise variance standardization, the response difference, $d(\boldsymbol{\beta})$, is 17% of a standard deviation, about as large as the direct responses.

Standardization with the square root of the $\mathbf{P}$ matrix places the lengths of response vectors and evolvabilities in standard deviation units appropriate to their direction. For example, the evolvability, $e(\boldsymbol{\beta})$, in population 1 is 9.5% of the phenotypic standard deviation in direction $\boldsymbol{\beta}$. The oblique transformation of the coordinate system makes the elements of the response vectors difficult to interpret. In this case, $\mathbf{P}$ standardization results in similar vectors and scalar measures of evolvability to $\boldsymbol{\sigma}$ standardization.

The many differences between the statistics calculated on different scales make clear that the choice of scale can strongly influence the results. Each standardization gives a unique

weighting of the traits that stretches or compresses each of the bases of phenotype space to a different degree. In addition, the square-root-of-**P** transformation also performs an oblique rotation of the bases.

Finally, we can compare the evolvability statistics over the entire phenotype space. Table A5 shows the mean evolvability, $\bar{e}$,conditional evolvability, $\bar{c}$, respondability, $\bar{r}$, and autonomy, $\bar{a}$, values for the two hypothetical populations. The average unconditional evolvability, $\bar{e}$, on a mean-standardized scale is 0.5% in population 1 and 1.4% in population 2. The average conditional evolvability, $\bar{c}$, is 0.16% of the mean in population 1 but just 0.03% in population 2. This difference is reflected in the lower autonomies, $\bar{a}$, in population 2. The raw and standard-deviation scales show a similar pattern, in which population 2 is more unconditionally evolvable but also more constrained than population 1. The key cause of this result is that the eigenvalues of **G** matrix 2 are more uneven than those in **G** matrix 1.

Table A1 Example **G** and the residual matrix **E = P – G** matrices and trait means.

Population	Trait (Units)	**G**	**E**	$\bar{z}$
1	1 (mm)	$\begin{bmatrix} 10 & 10 & 20 \\ & 30 & 20 \\ & & 80 \end{bmatrix}$	$\begin{bmatrix} 10 & 13 & 50 \\ & 30 & 40 \\ & & 890 \end{bmatrix}$	$\begin{bmatrix} 73 \\ 138 \\ 82 \end{bmatrix}$
	2 (mm)			
	3 (eggs)			
2	1 (mm)	$\begin{bmatrix} 20 & 16 & -10 \\ & 20 & 20 \\ & & 150 \end{bmatrix}$	$\begin{bmatrix} 20 & 20 & 20 \\ & 50 & 100 \\ & & 600 \end{bmatrix}$	$\begin{bmatrix} 80 \\ 152 \\ 64 \end{bmatrix}$
	2 (mm)			
	3 (eggs)			

Table A2 Example **G** matrices from Table A1 standardized by trait means and variances.

Population	$\mathbf{G}_{\mu} \times 100$	$\mathbf{G}_{\sigma}$	$\mathbf{G}_{P}$
1	$\begin{bmatrix} 0.188 & 0.099 & 0.334 \\ & 0.158 & 0.177 \\ & & 1.190 \end{bmatrix}$	$\begin{bmatrix} 0.500 & 0.289 & 0.144 \\ & 0.500 & 0.083 \\ & & 0.082 \end{bmatrix}$	$\begin{bmatrix} 0.602 & -0.086 & 0.084 \\ & 0.536 & 0.023 \\ & & 0.067 \end{bmatrix}$
2	$\begin{bmatrix} 0.313 & 0.132 & -0.195 \\ & 0.087 & 0.206 \\ & & 3.662 \end{bmatrix}$	$\begin{bmatrix} 0.500 & 0.302 & -0.058 \\ & 0.286 & 0.087 \\ & & 0.200 \end{bmatrix}$	$\begin{bmatrix} 0.424 & 0.154 & -0.105 \\ & 0.188 & 0.016 \\ & & 0.196 \end{bmatrix}$

Table A3 Standardized selection gradients and response vectors for example populations in Table A1 on the raw and three standardized scales.

Pop.		$\Delta\bar{\mathbf{z}}$	$\boldsymbol{\beta}_{\mu}$	$\Delta\bar{\mathbf{z}}_{\mu}$	$\boldsymbol{\beta}_{\sigma}$	$\Delta\bar{\mathbf{z}}_{\sigma}$	$\boldsymbol{\beta}_{P}$	$\Delta\bar{\mathbf{z}}_{P}$
1	vector	$\begin{bmatrix} 0.24\text{ mm} \\ 0.22\text{ mm} \\ 0.88\text{ eggs} \end{bmatrix}$	$\begin{bmatrix} 0.37 \\ -0.14 \\ 0.82 \end{bmatrix}$	$\begin{bmatrix} 0.0033 \\ 0.0016 \\ 0.0107 \end{bmatrix}$	$\begin{bmatrix} 0.022 \\ -0.008 \\ 0.311 \end{bmatrix}$	$\begin{bmatrix} 0.054 \\ 0.028 \\ 0.028 \end{bmatrix}$	$\begin{bmatrix} 0.036 \\ 0.016 \\ 0.319 \end{bmatrix}$	$\begin{bmatrix} 0.047 \\ 0.013 \\ 0.025 \end{bmatrix}$
	length	0.938	0.908	0.011	0.312	0.067	0.321	0.055
2	vector	$\begin{bmatrix} -0.02\text{ mm} \\ 0.26\text{ mm} \\ 1.43\text{ eggs} \end{bmatrix}$	$\begin{bmatrix} 0.40 \\ -0.15 \\ 0.64 \end{bmatrix}$	$\begin{bmatrix} -0.0002 \\ 0.0017 \\ 0.0223 \end{bmatrix}$	$\begin{bmatrix} 0.032 \\ -0.008 \\ 0.274 \end{bmatrix}$	$\begin{bmatrix} -0.003 \\ 0.031 \\ 0.052 \end{bmatrix}$	$\begin{bmatrix} 0.025 \\ 0.042 \\ 0.268 \end{bmatrix}$	$\begin{bmatrix} -0.011 \\ 0.016 \\ 0.051 \end{bmatrix}$
	length	1.454	0.770	0.022	0.276	0.061	0.273	0.054

The selection gradient is **β'** = [0.005/mm, –0.001/mm, 0.01/egg], and the length (norm) of this vector is 0.011 in a combination of egg and mm units.

Table A4 Evolvability statistics for the trait **β'** = [0.005/mm, –0.001/mm, 0.01/egg].

	Standardization											
	None*			Mean			Standard deviation			Square root of **P**		
	Population			Population			Population			Population		
Statistic	1	2	compare	1	2	compare	1	2	compare	1	2	compare
$r(\boldsymbol{\beta})$	84	129	na†	0.0125	0.0291	0.43	0.214	0.221	0.97	0.170	0.199	0.85
$e(\boldsymbol{\beta})$	78	111	0.70	0.0119	0.0236	0.50	0.100	0.184	0.55	0.095	0.188	0.50
$c(\boldsymbol{\beta})$	29	3.8	7.67	0.0046	0.0003	15.91	0.043	0.038	1.14	0.056	0.158	0.35
$a(\boldsymbol{\beta})$	0.38	0.03		0.39	0.01		0.43	0.21		0.59	0.84	
θ‡	22	31	16	18	36	15	62	34	51	56	19	78
$d(\boldsymbol{\beta})$§			54			0.010			0.174			0.234

The 'compare' column compares the responses in the two populations. For the respondabilities, $r(\boldsymbol{\beta})$, and evolvabilities, $e(\boldsymbol{\beta})$ and $c(\boldsymbol{\beta})$, the comparison is the ratio of the value in population 1 to that in population 2, when each population is standardized with its own vector or matrix. For θ and $d(\boldsymbol{\beta})$, both populations are standardized by the average of the standardization vectors or matrices in the two populations.

*The units for the responses of each population are a mixture of mm and eggs, and therefore most of these statistics have no clear interpretation.

†The ratio of respondabilities is meaningless on the raw scale.

‡In the columns labeled 1 and 2, this is the angle between $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}\bar{\mathbf{z}}$. In the 'compare' column it is the angle between the response vectors in the two populations, θ_d.

§Response differences were calculated from a standard length $\boldsymbol{\beta}$ under each standardization. In the original paper we calculated response differences using the unstandardized **β.**

Table A5 Expectations of evolvability statistics over a uniform distribution of selection gradients in the entire phenotype space for the hypothetical populations.

	Standardization							
	None[a]		Mean		Variance		**P**	
Statistic	1	2	1	2	1	2	1	2
$\overline{e}$	40.00	63.33	0.0051	0.0135	0.361	0.329	0.402	0.269
$\overline{r}$	50.79	83.17	0.0070	0.0193	0.453	0.408	0.463	0.309
$\overline{c}$	13.47	4.77	0.0016	0.0003	0.134	0.062	0.192	0.182
$\overline{a}$	0.389	0.091	0.388	0.033	0.422	0.207	0.470	0.708

[a]The units for the responses of each population are a mixture of millimeters and eggs, and these statistics therefore have no clear interpretation.