Israel Journal of Ecology & Evolution, 2019
http://dx.doi.org/10.1163/22244662-20191064

BRILL

Measuring natural selection on multivariate phenotypic traits: a protocol for verifiable and reproducible analyses of natural selection

Facundo Xavier Palacio[a,*], Mariano Ordano[b] and Santiago Benitez-Vieyra[c]

[a]*Sección Ornitología, División Zoología Vertebrados, Facultad de Ciencias Naturales y Museo, and Consejo Nacional de Investigaciones Científicas y Técnicas, Paseo del Bosque s/n (B1900FWA), La Plata, Buenos Aires, Argentina*
[b]*Unidad Ejecutora Lillo, Fundación Miguel Lillo, and Consejo Nacional de Investigaciones Científicas y Técnicas, Miguel Lillo 251 (T4000JFE), San Miguel de Tucumán, Tucumán, Argentina*
[c]*Laboratorio de Ecología Evolutiva y Biología Floral, Instituto Multidisciplinario de Biología Vegetal, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Córdoba, Av. Vélez Sársfield 1611 (X5016GCA), Córdoba, Argentina*

Abstract The use of multiple regression analysis to quantify the regime and strength of natural selection in nature has been an influential approach in evolutionary biology over the last 36 years. However, many studies fail to report the protocol of estimation of selection coefficients (selection gradients) and the specific model assumptions, thus failing to verify and reproduce the estimation of selection coefficients. We present a brief overview of the Lande and Arnold's approach and a step-by-step R routine to aid researchers to perform a verifiable and reproducible regression analysis of natural selection. The steps involved in the analysis include: (1) assessing collinearity between phenotypic traits, (2) testing normality of model residuals, and (3) testing multivariate normality of phenotypic traits. We also performed a series of simulations to test the effect of non-symmetrical (skewed) phenotypic traits on the estimation of linear selection gradients. These showed that the bias in the linear gradient increased with increased skewness in phenotypic traits for the quadratic model, whereas the linear gradient of a model with only linear terms was nearly independent of trait skewness. If none of the above assumptions are met, selection gradients need to be estimated from two separate equations, whereas standard errors must be computed using other methods (e.g. bootstrapping). We expect that the procedure outlined here and the availability of analytical codes motivate the verifiability and reproducibility of the Lande and Arnold's approach in the study of microevolution.

Keywords correlational selection; linear selection; non-linear selection; selection gradients

Introduction

The use of multiple regression analysis to quantify the mode and strength of natural selection in nature has been an influential approach in evolutionary biology over the last 36 years. This approach has allowed researchers to estimate natural selection on the mean (directional selection), variance (stabilizing/disruptive selection) and covariance (correlational selection or selection on trait combinations) on correlated phenotypic traits using a multiple linear regression model (Lande and Arnold, 1983). Despite its statistical formulation follows well established linear regression procedures, many studies fail to report the protocol of estimation of selection coefficients (Stinchcombe et al. 2008). Given the statistical assumptions of the model (i.e. uncorrelated phenotypic traits, normality of model residuals and multivariate normality of phenotypic traits; Lande and Arnold, 1983; Mitchell-Olds and Shaw, 1987), it is necessary to describe how to estimate and report these coefficients along with standard errors and *P*-values (Palacio, 2018). Overall, these assumptions indicate that fitting a Lande and Arnold's model is not always straightforward. For the same reason, Lande and Arnold's approach has spawned a plethora of works devoted to expand this framework (e.g. Arnold and Wade, 1984; Rausher, 1992; Schluter and Nychka, 1994; Janzen and Stern, 1998; Scheiner et al., 2000; Blows and Brooks, 2003; Shaw and Geyer, 2010; Calsbeek, 2012; Morrissey and Sakrejda, 2013; Morrissey, 2014). Researchers may feel overwhelmed by this myriad of methods and thus hesitate about how to measure and report the results of natural selection analyses. Thereby, we here present a brief overview of the Lande and Arnold's approach and a step-by-step R (R Core Team 2017) routine for a verifiable and reproducible regression analysis of natural selection.

The Lande and Arnold's approach and its assumptions

Lande and Arnold (1983) developed a multiple regression model relating individual fitness with phenotypic quantitative traits

$$w = \alpha + \sum_{i=1}^{n} \beta_i z_i + \frac{1}{2} \sum_{i=1}^{n} \gamma_{ii} z_i^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} z_i z_j + \varepsilon_i$$

or, in matrix notation

$$w = \alpha + \boldsymbol{\beta}^T \mathbf{z} + \frac{1}{2} \mathbf{z}^T \boldsymbol{\gamma} \mathbf{z} + \varepsilon$$

*Corresponding author. E-mail: facundo_palacio@fcnym.unlp.edu.ar

Considering a biological population, a set of phenotypic traits and a fitness estimator associated with each individual, w is the vector of fitness values divided by the mean population fitness (relative fitness), $\mathbf{z}$ is the vector of phenotypic values for n traits, α and ε are intercept and error terms, $\boldsymbol{\beta}$ is the vector of linear (directional) selection gradients β_i, and $\boldsymbol{\gamma}$ is the matrix of non-linear selection gradients, where γ_{ii} is the stabilizing/disruptive selection gradient for trait i, and γ_{ij} is the correlational selection gradient for traits i and j (Lande and Arnold, 1983; Lynch and Walsh, 1998). Gradients are estimated via ordinary least squares, and standardized selection gradients allow comparisons among populations and traits. The partial regression coefficients, β_i's and γ_{ii}'s, represent the change in relative fitness by changing z_i while holding remaining trait values constant (Lande and Arnold, 1983). From an evolutionary perspective, selection gradients estimate the direct effects of selection on phenotypic traits (i.e. independent from other measured traits), as it removes the effects of correlations among traits (Lande and Arnold, 1983).

In practice, however, several issues must be addressed to properly quantify the regime and strength of natural selection on phenotypic traits (Mitchell-Olds and Shaw, 1987). First, extreme correlations among phenotypic traits should be tested, as high collinearity leads to instable coefficient estimation, inflated standard errors and consequently inference on selection biased (Mitchell-Olds and Shaw, 1987; Graham, 2003; Dormann et al., 2013; but see Morrissey and Ruxton, 2018). Second, $\boldsymbol{\beta}$ not necessarily equals the vector of partial regression coefficients obtained by assuming a regression with only linear terms. This only holds when phenotypic traits meet multivariate normality (Lande and Arnold, 1983; Lynch and Walsh, 1998). A pragmatic solution (i.e. without testing multivariate normality) is thus to estimate linear selection from a model with only linear terms and to estimate nonlinear selection from a model with quadratic and cross-product terms (e.g. Benitez-Vieyra et al., 2006; Bentsen et al., 2006; Ordano et al., 2008; Sandring and Ågren, 2009; Palacio et al., 2014). Third, errors are assumed to follow a normal distribution. Although the normality assumption is robust to violations, some fitness proxies (e.g. survival) will be particularly affected by violations of this assumption (Janzen and Stern, 1998). Because the nature of the variables used as phenotypic traits is diverse (Kingsolver et al., 2001), all these assumptions indicate that fitting a Lande and Arnold's model is cumbersome, and needs specific procedures on the basis of the type of variables included.

Data simulation

The effect of extreme collinearity among predictors (traits) on model estimation and prediction via theoretical and simulation approaches has been extensively addressed in the literature (see Dormann et al., 2013 for a review). Here, we focus on the issue of skewed (non-normal) phenotypic trait distributions, which has received much less attention in natural selection studies. Theoretical demonstrations have shown that, if a phenotypic trait distribution is non-normal, the linear term in the quadratic model is a function of linear and quadratic selection, whereas the linear term in a linear model is only a function of linear selection (Lynch and Walsh, 1998). Therefore, it is expected that the higher the asymmetry (i.e. skewness), the higher the bias of the linear gradient in a quadratic model. By contrast, the bias of the linear gradient in a model with only linear terms is expected to be independent from trait skewness. This is what justifies using two separate models to estimate linear and quadratic gradients when multivariate normality of phenotypic traits is not met.

We simulated two uncorrelated ($r = 0.03$) phenotypic traits x_1, x_2 and 100 individuals per population following Poisson distributions with parameters λ_1 and λ_2 ($\lambda_1 = \lambda_2$), respectively. These parameters took 1000 evenly distributed values from 1 to 1000 (i.e. 1000 simulated populations). These values also represented decreased skewness, since skewness = $\lambda^{-1/2}$ and the Poisson distribution approximates to the normal when λ approaches infinity (and therefore, skewness approaches to zero). The relationship between standardized phenotypic traits and relative fitness was simulated using five (two linear, two quadratic and one correlational) random gradients drawn from an arbitrary uniform distribution with minimum = –1 and maximum = 1. Relative fitness was computed as a function of phenotypic traits plus arbitrary random normal noise (SD = 0.2). For each simulation, we fitted two separate Lande and Arnold's models (one including and one excluding quadratic terms), estimated gradients and differentials for each model and trait, and computed the difference between linear gradients and its corresponding linear differential. The selection differential S_i on a trait i quantifies the direct effects of selection on a trait i, as well as the indirect effects due to selection on correlated traits (Lande and Arnold, 1983):

$$S_i = \beta_i s_i^2 + \sum \beta_j s_{ij}$$

where S_i is the linear differential of trait i, s_i^2 is the variance of trait i, and s_{ij} is the covariance between trait i and j. Under no correlation between phenotypic traits ($s_{ij} = 0$), and, after trait standardization, the linear gradient equals the linear differential:

$$\frac{S_i}{s_i} = \frac{\beta_i s_i^2}{s_i} + \frac{\sum \beta_j s_{ij}}{s_i s_j}$$

$$\frac{S_i}{s_i} = \beta_i s_i + \sum \beta_j r_{ij}$$

$$S_i = \beta_i$$

where r_{ij} is the correlation between trait i and j. So, the linear differential is always unbiased and serves as a baseline to assess the bias of its respective linear gradient when including quadratic terms in the model. We therefore plotted skewness against the difference between linear gradients and differentials ($\beta_i - S_i$) for each model and trait. Under no correlation between traits, the difference between

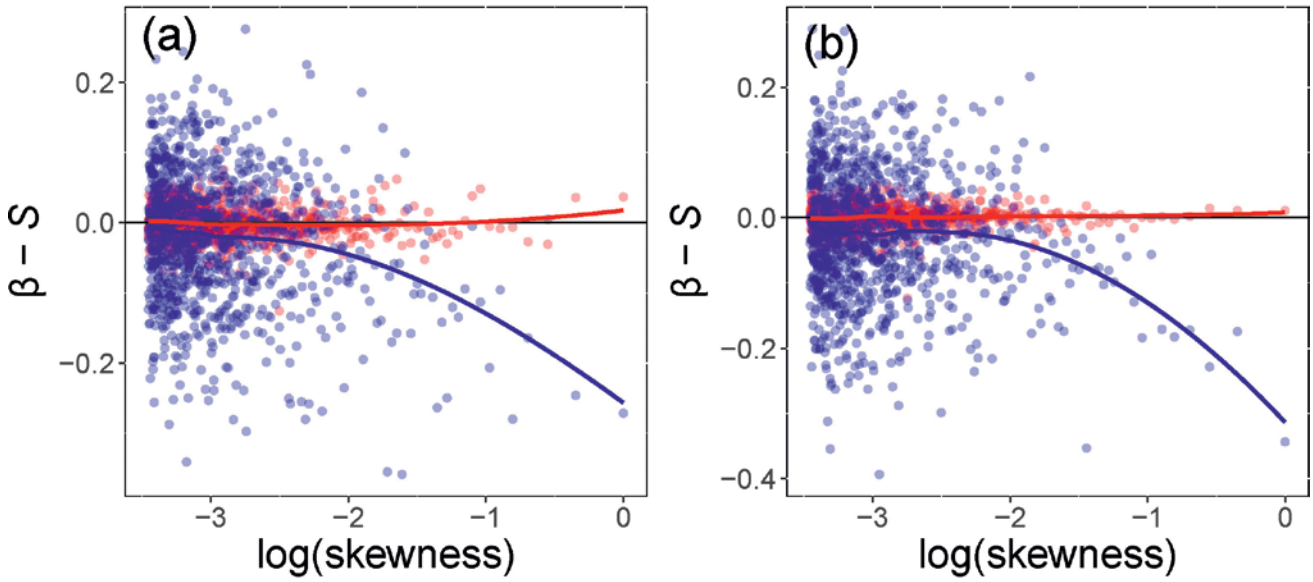


Figure 1. Effect of skewed phenotypic trait distributions z_1 (a) and z_2 (b) on estimation of linear selection gradients. Simulated phenotypic traits followed Poisson distributions with parameter λ, and skewness was computed as $\lambda^{-1/2}$ (n = 1000 simulations). Each point represents the difference between the linear gradient β and its corresponding linear differential S in a population of 100 individuals (red: model with only linear terms, blue: model with both linear and quadratic terms). Curves depict non-parametric LOESS fit.

a linear gradient and its corresponding linear differential should increase with increased skewness in the quadratic model, whereas this difference should remain constant in the linear model. The R code is available at Supplementary Material S1.

Simulations showed that, as expected, the bias in the linear gradient increased with increased skewness in phenotypic traits for the quadratic model (Fig. 1). For the maximum skewness value (skewness = 1), the linear gradient of x_1 in the quadratic model was 2.26 times higher than its linear differential, and, more importantly, of opposite sign (–0.15 vs 0.12). The linear gradient of x_2 in the quadratic model was 0.65 times lower than its linear differential (0.52 vs 0.18). On the other hand, the linear gradient of a model with only linear terms was nearly independent of trait skewness (Fig. 1).

Application

We applied the Lande and Arnold's approach to the dataset of Benitez-Vieyra et al. (2006), who studied the role of pollinators as natural selection agents on floral traits (flower number, nectary depth, and size of lower lip) in a population (n = 116 individuals) of the orchid *Cyclopogon elatus* (Orchidaceae) from central Argentina. Here we used two phenotypic traits (z_1 = standardized flower number and z_2 = standardized mean nectary depth per plant) and number of exported pollinaria as a (male) fitness measure to obtain selection gradients. The model is expressed as:

$$w = \alpha + \begin{pmatrix}\beta_1 & \beta_2\end{pmatrix}^T \begin{pmatrix}z_1 \\ z_2\end{pmatrix} + \frac{1}{2}\begin{pmatrix}z_1 & z_2\end{pmatrix}^T \begin{pmatrix}\gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22}\end{pmatrix}\begin{pmatrix}z_1 \\ z_2\end{pmatrix} + \varepsilon$$

$$w = \alpha + \beta_1 z_1 + \beta_2 z_2 + \frac{1}{2}\gamma_{11} z_1^2 + \frac{1}{2}\gamma_{22} z_2^2 + \gamma_{12} z_1 z_2 + \varepsilon$$

The first step is to relativize absolute fitness W to its population mean $w = \frac{W}{\bar{W}}$, and standardize phenotypic traits to mean zero and variance one. Second, it should be checked whether phenotypic traits come from a multivariate normal distribution. This may be a daunting task, as there are at least 50 statistical techniques for testing multivariate normality (Mecklin and Mundfrom, 2005). Mecklin and Mundfrom (2005) suggest using Henze-Zirkler's and Royston's tests due to their type I and type II error control and power, coupled with graphical methods. For this, we may plot histograms and density plots (Fig. 2a-b), and run multivariate normality tests along with Q-Q plots using the package MVN (Korkmaz et al., 2014). Both tests suggest little support for multivariate normality of phenotypic traits (HZ = 3.48, P < 0.0001; H = 34.79, P < 0.0001). Moreover, Q-Q plots show some deviations from the straight line (Fig. 2c-d), indicating departure from a multivariate normal distribution. Overall, rejecting multivariate normality suggests that two separate models (one linear and another quadratic) should be fitted, in order to obtain unbiased linear and nonlinear selection gradient estimates (Lande and Arnold, 1983).

To assess collinearity among phenotypic traits, we may compute both pairwise correlations and variance inflation factors (VIFs) on the linear model using the package car (Fox and Weisberg, 2011). Briefly stated, VIFs quantify the extent of correlation between one predictor and the rest of predictors in a model (O'brien 2007). A low Pearson correlation coefficient between both traits (r = 0.21) and VIF values lower than 10 (VIF = 1.04) suggest low collinearity effects (Dormann et al., 2013). Therefore, we are allowed to include both phenotypic traits into the Lande and Arnold's model.

We also need to assess whether model residuals are normally distributed. If residuals depart from a normal distribution, selection gradients will be still unbiased, but

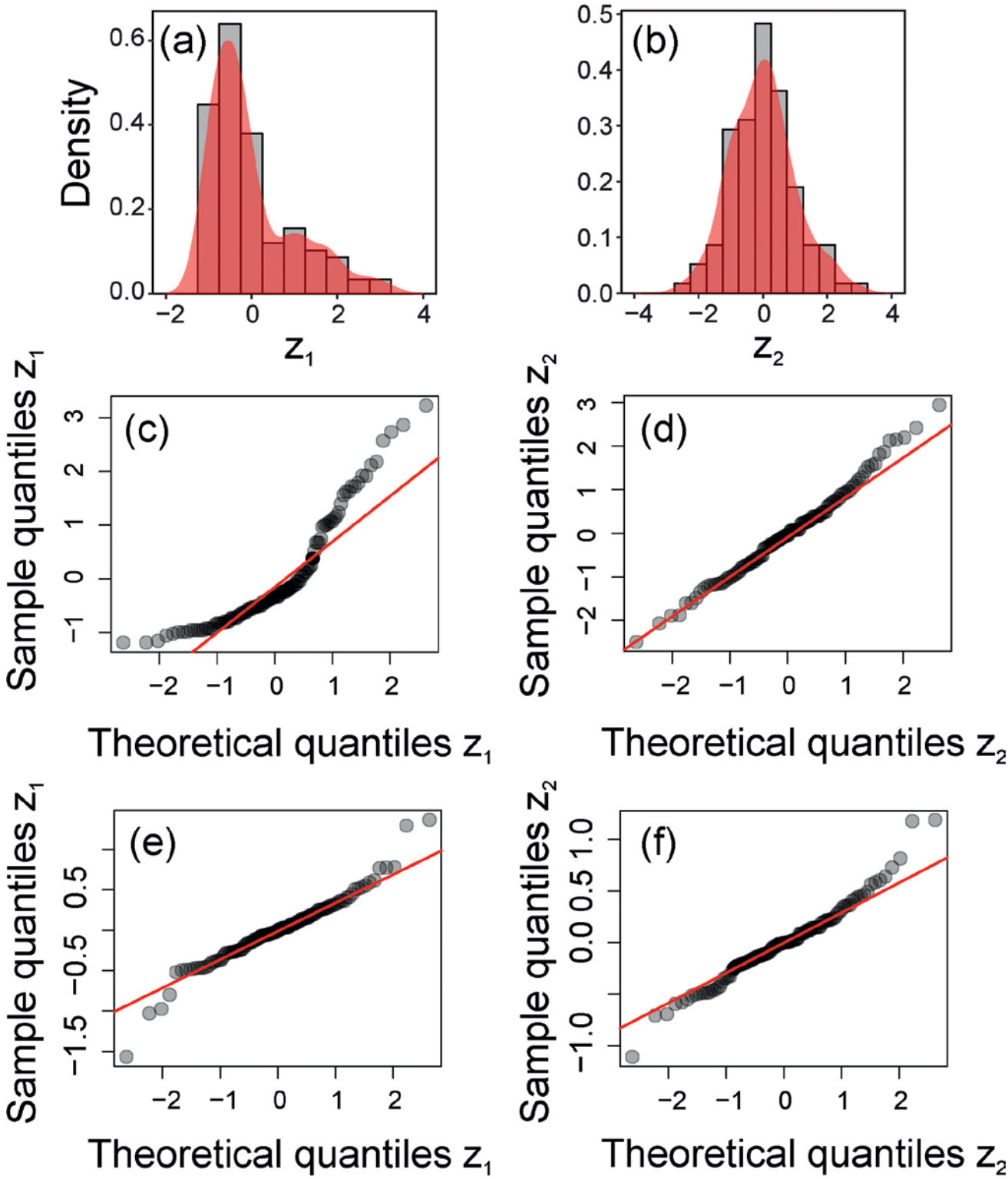


Figure 2. Data exploration and visual checking of Lande and Arnold's model assumptions. Histograms (a-b) and Q-Q plots (c-d) of flower number z_1 and mean nectary depth z_2 of a *Cyclopogon elatus* population (n = 116 plants) are shown. Q-Q plots to assess normality of residuals in the linear (e) and quadratic (f) model are also depicted.

standard errors and *P*-values will be not (Mitchell-Olds and Shaw, 1987). To check this, it is often recommended to rely on graphical methods, such as the Q-Q plot, rather than on normality tests (Zuur et al., 2010). Both the results of the Shapiro-Wilk test ($W_{\text{linear model}}$ = 0.96, P = 0.001; $W_{\text{quadratic model}}$ = 0.98, P = 0.03) and the Q-Q plots (Fig. 2e-f) show significant departure from normality, arguing for the use of a different method to estimate standard errors and *P*-values. Although transformations of the fitness measure are possible, they are not recommended, as they do not preserve the genetic interpretation of the analysis (Lande and Arnold, 1983; Mitchell-Olds and Shaw, 1987). This can be circumvented via resampling methods (Mitchell-Olds and Shaw, 1987; Morrissey and Sakrejda,

Table 1. Multivariate phenotypic selection mediated by pollinators on flower number (z_1) and mean nectary depth (z_2) in *Cyclopogon elatus* (n = 116 plants). Linear gradients (β_i) were estimated from a multiple regression with only linear terms, whereas nonlinear gradients (γ_{ii} and γ_{ij}) were estimated from a multiple regression with both linear and quadratic terms. Standard errors (SE) and 95% confidence intervals (CI) were computed using 999 bootstrap samples (see Methods). Significant gradients are shown in bold.

Phenotypic trait	β_i	SE	CI	γ_{ii} or γ_{ij}	SE	CI
z_1	**0.652**	0.039	0.505–0.777	**-0.200**	0.064	-0.354–-0.043
z_2	**0.198**	0.039	0.121 –0.291	-0.086	0.056	-0.239–0.049
$z_1 \times z_2$				**0.143**	0.038	0.024–0.267

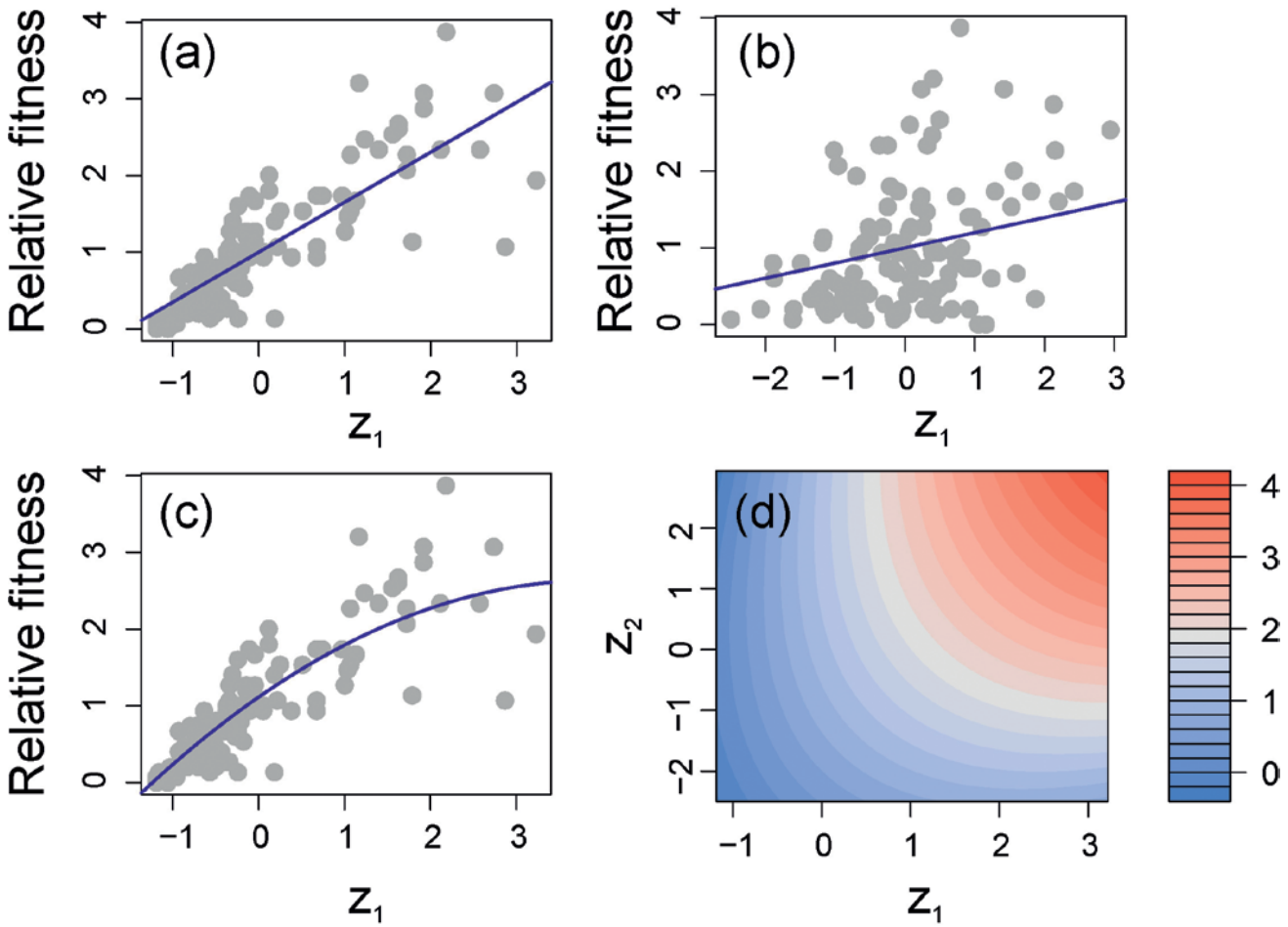


Figure 3. Insect-pollinated mediated selection on flower number (z_1) and nectary depth (z_2) in a *Cyclopogon elatus* population (n = 116 plants). (a) and (b) show regressions with linear terms only (linear gradients), whereas (c) and (d) show regressions with both linear and quadratic terms (non-linear gradients). (e) shows a fitness surface derived from the quadratic model, depicting correlational selection. The colored scale represents predicted relative fitness (number of exported pollinaria per plant).

2013). Here, we used ordinary nonparametric bootstrapping to estimate standard errors along with 95% bias-corrected bootstrap confidence intervals (Efron and Tibshirani, 1994) with the package boot (Canty and Ripley, 2017).

A significant positive linear gradient on both flower number and nectary depth was found (Table 1), which means positive directional selection on these traits (Fig. 3a-b). Besides, there was a significant negative quadratic gradient on flower number, as well as a positive correlational selection gradient on the combination of flower number and nectary depth (Table 1). However, plots depicting relationships between traits and fitness are still needed, because a significant quadratic gradient is a necessary, but not sufficient, condition to infer stabilizing or disruptive selection (Endler, 1986; Mitchell-Olds and Shaw, 1987). This is because quadratic gradients indicate curvature, rather than the presence of a local minimum or maximum (Schluter, 1988; Phillips and Arnold, 1989). Indeed, the negative quadratic gradient on flower number indicates a combination of linear and non-linear relationship between this trait and fitness, rather than pure stabilizing selection (Fig. 3c).

Similarly, we could barely interpret correlational gradients without visualizing the fitness surface, as it may take numerous shapes regardless of the sign of the gradient (Brodie et al., 1995). But still, visualizing the fitness surface is not trivial and may be misleading, an issue thoroughly addressed in the literature (Schluter, 1988; Phillips

and Arnold, 1989; Schluter and Nychka, 1994; Shaw and Geyer, 2010). This is beyond the scope of this paper, and we aim to provide general guidelines to perform a classical Lande and Arnold's approach instead. Thereby, here we only show how to visualize natural selection favoring specific trait combinations. Using the package visreg (Breheny and Burchett, 2017), 2D or 3D fitness surfaces depicting correlational selection can be plotted (Fig. 3e). This shows that insect-pollinated selection favors combinations of large flower numbers and deep nectaries. These surfaces can also be extended to include more than two phenotypic traits, in whose case one must keep one or more traits fixed to their means. The R code is available at Supplementary Material S2.

Discussion

Here, we show a verifiable and reproducible routine to perform a classical natural selection analysis in order to clarify some statistical issues. Further, we corroborate that ignoring multivariate normality of phenotypic traits may strongly bias selection gradients. Finally, we showed that under the worst scenario of statistical assumptions (no multivariate normality of the phenotype and non-normal residual distribution), the summary table describing the Lande and Arnold's model will be a mosaic of selection gradients, standard errors and confidence intervals estimated from different methods. These should be basic components of any selection study (Kingsolver et al., 2012). We expect that the general procedure outlined here and the availability of the code motivate a verifiable and reproducible use (Gandrud, 2013) of the Lande and Arnold's approach to the study of microevolution.

Other approaches are also useful for the estimation of selection gradients (e.g. structural equation modeling; Scheiner et al., 2000; generalized additive models; Morrissey and Sakrejda, 2013; optimal transport theory; Henshaw and Zemel, 2017). However, these approaches require large sample sizes. Although "more is better" in terms of sample size, it is common for natural selection studies to be carried out with sample sizes less than 100 individuals (Conner, 2001; Kingsolver et al., 2001). Sample size constraints lead to the use of the Lande and Arnold's approach, rather than to the use of modern analytical approaches (Morrissey and Sakrejda, 2013; Chevin et al., 2015; terHorst et al., 2015). A deeper work is necessary to fill the current requirements of verifiability and reproducibility of routines for this menu of methods for the study of natural selection. Hopefully, our proposal for a verifiable and reproducible implementation of the Lande and Arnold's approach will contribute to turn another page in the study of microevolution.

Acknowledgements

This study was partially funded by Fundación Miguel Lillo (Ministerio de Educación, Argentina) and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Ministerio de Ciencia y Tecnología, Argentina). The authors declare that they have no conflict of interest.

Supplementary material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.9601178

References

Arnold SJ, Wade MJ. (1984). On the measurement of natural and sexual selection: theory. *Evolution.* 38:709–719.

Benitez-Vieyra S, Medina AM, Glinos E, Cocucci AA. (2006). Pollinator-mediated selection on floral traits and size of floral display in *Cyclopogon elatus*, a sweat bee-pollinated orchid. *Funct Ecol.* 20:948–957.

Bentsen CL, Hunt J, Jennions MD, Brooks R. (2006). Complex multivariate sexual selection on male acoustic signaling in a wild population of *Teleogryllus commodus*. *Am Nat.* 167:E102–E116.

Blows MW, Brooks R. (2003). Measuring nonlinear selection. *Am Nat.* 162:815–820.

Breheny P, Burchett W. (2017). visreg: visualization of regression models. R package version 2.4-1. Available from https://CRAN.R-project.org/package=visreg (accessed June 2018)

Brodie ED, Moore AJ, Janzen FJ. (1995). Visualizing and quantifying natural selection. *Trends Ecol Evol.* 10:313–318.

Calsbeek B. (2012). Exploring variation in fitness surfaces over time or space. *Evolution.* 66:1126–1137.

Canty A, Ripley B. (2017). boot: Bootstrap R (S-Plus) functions. R package version 1.3-20. Available from https://cran.r-project.org/web/packages/boot/ (accessed June 2018)

Chevin LM, Visser ME, Tufto J. (2015). Estimating the variation, autocorrelation, and environmental sensitivity of phenotypic selection. *Evolution*. 69:2319–2332.

Conner JK. (2001). How strong is natural selection? *Trends Ecol Evol.* 16:215–217.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, García Marquéz JR, Gruber B., Lafourcade B., Leitão PJ, Münkemüller T, McClean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography.* 36:27–46.

Efron B, Tibshirani RJ. (1994). *An Introduction to the Bootstrap*. CRC press, Boca Raton.

Endler JA. (1986). *Natural Selection in the Wild*. Princeton University Press, Princeton.

Fox J, Weisberg S. (2011). *An R Companion to Applied Regression*. Thousand Oaks, Sage.

Gandrud C. (2013). *Reproducible Research with R and R Studio*. CRC Press, London.

Graham MH, (2003). Confronting multicollinearity in ecological multiple regression. *Ecology.* 84:2809–2815.

Henshaw JM, Zemel Y, (2017). A unified measure of linear and nonlinear selection on quantitative traits. *Methods Ecol Evol.* 8:604–614.

Janzen FJ, Stern HS, (1998). Logistic regression for empirical studies of multivariate selection. *Evolution.* 52:1564–1571.

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM. (2012). Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evol Ecol.* 26:1101–1118.

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P. (2001). The strength of phenotypic selection in natural populations. *Am Nat.* 157:245–261.

Korkmaz S, Goksuluk D, Zararsiz G. (2014). MVN: An R package for assessing multivariate normality. *R J.* 6:151–162.

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

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

Mecklin CJ, Mundfrom DJ. (2005). A Monte Carlo comparison of the Type I and Type II error rates of tests of multivariate normality. *J Stat Comput Simul.* 75:93–107.

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

Morrissey MB. (2014). In search of the best methods for multivariate selection analysis. *Methods Ecol Evol.* 5:1095–1109.

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

Morrissey MB, Ruxton GD. (2018). Multiple regression is not multiple regressions: the meaning of multiple regression and the non-problem of collinearity. *Philos Theory Pract Biol.* 10:3.

O'brien RM. (2007). A caution regarding rules of thumb for variance inflation factors. *Qual Quant.* 41:673–690.

Ordano M, Fornoni J, Boege K, Domínguez CA. (2008). The adaptive value of phenotypic floral integration. *New Phytol.* 179:1183–1192.

Palacio FX. (2018). *P*-values to determine whether natural selection operates in nature: seizing their potential. *Rev Mex Biodivers.* 89:327–330.

Palacio FX, Lacoretz M, Ordano M. (2014). Bird-mediated selection on fruit display traits in *Celtis ehrenbergiana* (Cannabaceae). *Evol Ecol Res.* 16:51–62.

Phillips PC, Arnold SJ. (1989). Visualizing multivariate selection. *Evolution.* 43:1209–1222.

R Core Team. (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from http:// www.R-project.org (accessed June 2018).

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

Sandring S, Ågren J. (2009). Pollinator-mediated selection on floral display and flowering time in the perennial herb *Arabidopsis lyrata*. *Evolution.* 63:1292–1300.

Scheiner SM, Mitchell RJ, Callahan, HS. (2000). Using path analysis to measure natural selection. *J Evol Biol.* 13:423–433.

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

Schluter D, Nychka D. (1994). Exploring fitness surfaces. *Am Nat.* 143:597–616.

Shaw RG, Geyer CJ (2010) Inferring fitness landscapes. *Evolution.* 64:2510–2520.

Stinchcombe JR, Agrawal AF, Hohenlohe PA, Arnold SJ, Blows MW. (2008). Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution.* 62:2435–2440.

terHorst CP, Lau JA, Cooper IA, Keller KR, Rosa RJL, Royer AM, Schultheis EH, Suwa T, Conner JK. (2015). Quantifying nonadditive selection caused by indirect ecological effects. *Ecology.* 96:2360–2369.

Zuur AF, Ieno EN, Elphick CS. (2010). A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol.* 1:3–14.

Israel Journal of Ecology & Evolution

Measuring natural selection on multivariate phenotypic traits: a protocol for verifiable and reproducible analyses of natural selection

Facundo Xavier Palacio[a,*], Mariano Ordano[b] and Santiago Benitez-Vieyra[c]

[a]Sección Ornitología, División Zoología Vertebrados, Facultad de Ciencias Naturales y Museo, and Consejo Nacional de Investigaciones Científicas y Técnicas, Paseo del Bosque s/n (B1900FWA), La Plata, Buenos Aires, Argentina

[b]Unidad Ejecutora Lillo, Fundación Miguel Lillo, and Consejo Nacional de Investigaciones Científicas y Técnicas, Miguel Lillo 251 (T4000JFE), San Miguel de Tucumán, Tucumán, Argentina

[c]Laboratorio de Ecología Evolutiva y Biología Floral, Instituto Multidisciplinario de Biología Vegetal, Consejo Nacional de Investigaciones Científicas y Técnicas, and Universidad Nacional de Córdoba, Av. Vélez Sársfield 1611 (X5016GCA), Córdoba, Argentina

*Corresponding author. E-mail: facundo_palacio@fcnym.unlp.edu.ar

Supplementary material

Effect of phenotypic trait skewness on selection gradients

Supplementary Material. Measuring natural selection on multivariate phenotypic traits. A protocol for verifiable and reproducible analyses of natural selection. Facundo Palacio, Mariano Ordano, and Santiago Benitez-Vieyra.

Load packages

```
library(ggplot2)
library(gridExtra)
```

Step 1. Simulation design

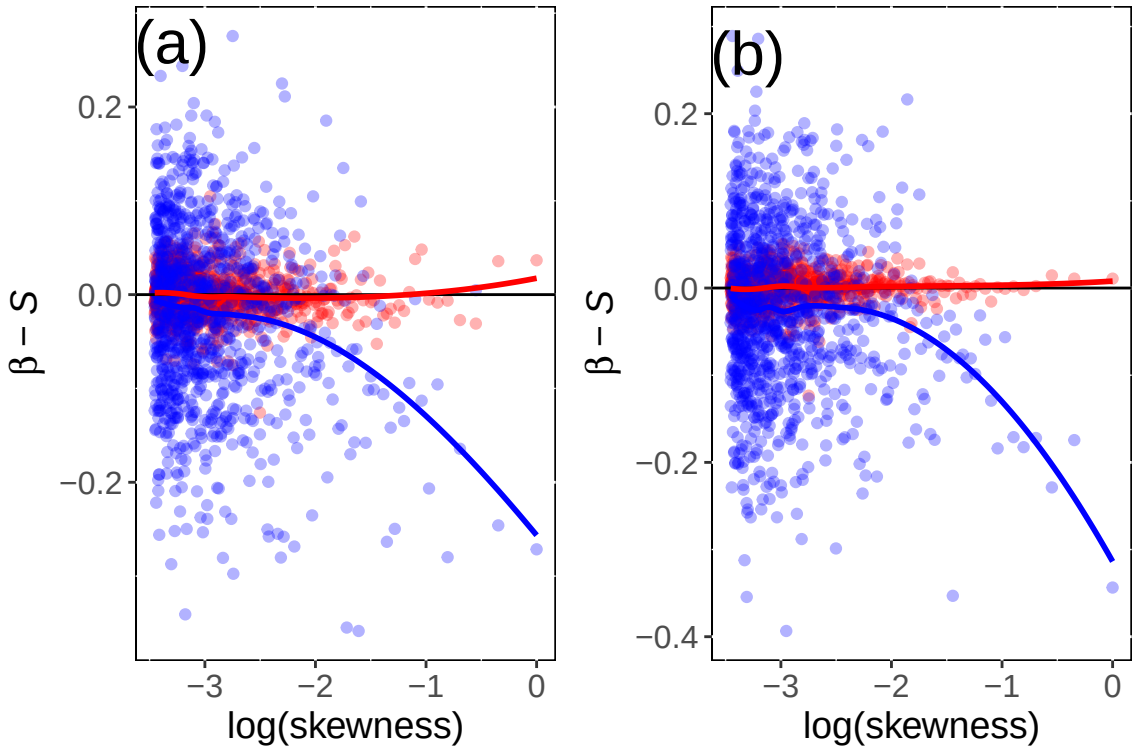
```
set.seed(77)
n.simulations <- 1000
beta <- runif(5, min = -1, max  = 1)
grad.LA.lin <- matrix(NA, nrow = n.simulations, ncol = 2)
grad.LA.quad <- matrix(NA, nrow = n.simulations, ncol = 2)
z1.skew <- c()
z2.skew <- c()
S1 <- c()
S2 <- c()
C1 <- c()
C2 <- c()

for(i in 1:n.simulations){
N <- 100

lambda <- 1:n.simulations
x1 <- rpois(N, lambda = lambda[i])
x2 <- rpois(N, lambda = lambda[i])
z1 <- (x1 - mean(x1))/sd(x1)
z2 <- (x2 - mean(x2))/sd(x2)
z1.skew[i] <- 1/sqrt(lambda[i])
z2.skew[i] <- 1/sqrt(lambda[i])

W <- 1 + beta[1]*z1 + beta[2]*z2 + beta[3]*z1^2 + beta[4]*z2^2 + beta[5]*z1*z2 +
     rnorm(N, mean = 0, sd = 0.2)
wrel <- W/mean(W)
S1[i] <- cov(wrel, z1)
S2[i] <- cov(wrel, z2)
C1[i] <- cov(wrel, z1^2)
C2[i] <- cov(wrel, z2^2)
# Models to estimate linear gradients
LA.lin <- lm(wrel ~ z1 + z2)
LA.quad <- lm(wrel ~ z1 + z2 + I((1/2)*z1^2) + I((1/2)*z2^2) + z1:z2)
grad.LA.lin[i,] <- LA.lin$coeff[2:3]
grad.LA.quad[i,] <- LA.quad$coeff[2:3]
}
```

Step 2. Gradient estimations

```
grad.LA.lin <- as.data.frame(grad.LA.lin)
names(grad.LA.lin) <- c("z1" , "z2")
grad.LA.quad <- as.data.frame(grad.LA.quad)
names(grad.LA.quad) <- c("z1" , "z2")
grad.LA.lin$z1.skew <- z1.skew
grad.LA.quad$z2.skew <- z2.skew
grad.LA.lin$diff.z1.S1 <- grad.LA.lin$z1 - S1
grad.LA.lin$diff.z2.S2 <- grad.LA.lin$z2 - S2
grad.LA.quad$diff.z1.S1 <- grad.LA.quad$z1 - S1
grad.LA.quad$diff.z2.S2 <- grad.LA.quad$z2 - S2
```

Step 3. Plot

```
p1 <- ggplot() + geom_point(data = grad.LA.lin, aes(log(z1.skew), diff.z1.S1),
                            col = "red", alpha = 0.3) +
  geom_point(data = grad.LA.quad, aes(log(z1.skew), diff.z1.S1),
                            col = "blue", alpha = 0.3) +
  geom_hline(yintercept = 0) +
  xlab("log(skewness)") +
  ylab(bquote(italic(beta) ~ "-" ~ italic(S))) +
  geom_smooth(data = grad.LA.lin, aes(log(z1.skew), diff.z1.S1),
              method = "loess", se = FALSE, col = "red") +
  geom_smooth(data = grad.LA.quad, aes(log(z1.skew), diff.z1.S1),
              method = "loess", se = FALSE, col = "blue") +
  theme(panel.background = element_rect(fill = "white",
                                        colour = "black"),
        axis.title = element_text(size = 14), axis.text = element_text(size = 12)) +
  annotate("text", label = "(a)", x = -3.3, y = 0.27, size = 8)

p2 <- ggplot() + geom_point(data = grad.LA.lin, aes(log(z2.skew), diff.z2.S2),
                            col = "red", alpha = 0.3) +
  geom_point(data = grad.LA.quad, aes(log(z2.skew), diff.z2.S2),
                            col = "blue", alpha = 0.3) +
  geom_hline(yintercept = 0) +
  xlab("log(skewness)") +
  ylab(bquote(italic(beta) ~ "-" ~ italic(S))) +
  geom_smooth(data = grad.LA.lin, aes(log(z2.skew), diff.z2.S2),
              method = "loess", se = FALSE, col = "red") +
  geom_smooth(data = grad.LA.quad, aes(log(z2.skew), diff.z2.S2),
              method = "loess", se = FALSE, col = "blue") +
  theme(panel.background = element_rect(fill = "white", colour = "black"),
        axis.title = element_text(size = 14), axis.text = element_text(size = 12)) +
  annotate("text", label = "(b)", x = -3.3, y = 0.27, size = 8)

grid.arrange(p1, p2, nrow = 1)
```


(a)
(b)
β − s
log(skewness)
0.2
0.0
−0.2
−0.4
−3
−2
−1
0

Protocol for natural selection analysis

Supplementary Material. Measuring natural selection on multivariate phenotypic traits. A protocol for verifiable and reproducible analyses of natural selection. Facundo Palacio, Mariano Ordano, and Santiago Benitez-Vieyra.

Load packages and data

```
library(MVN)
library(car)
library(boot)
library(visreg)
library(ggplot2)
library(knitr)

data <- read.table("cyclop.txt", header = TRUE)
data <- na.omit(data)
attach(data)
```

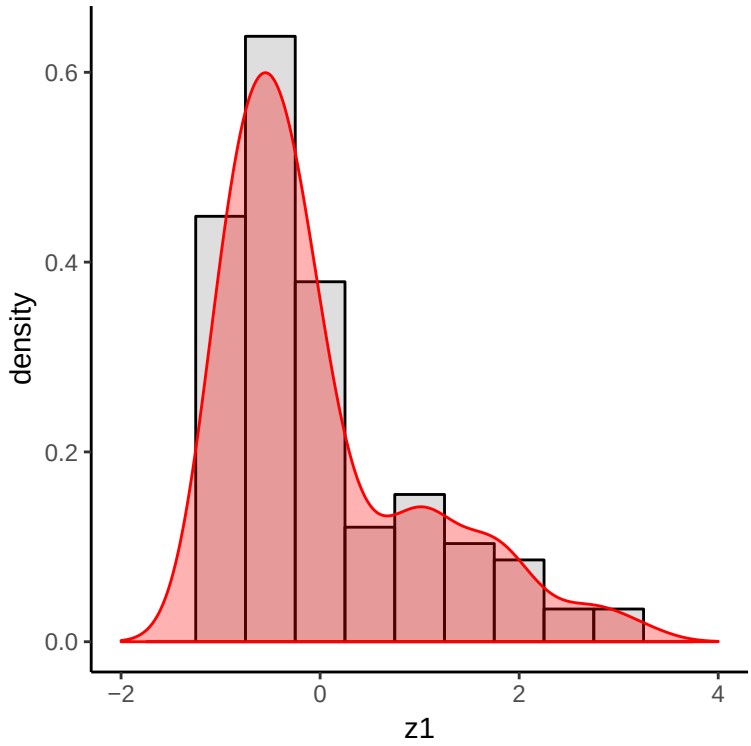
Step 1. Relative fitness and standardization of phenotypic traits (mean = 0, variance = 1)

```
W <- exported.pollinaria
wrel <- W/mean(W)
x1 <- flower.number
x2 <- nectary.depth
z1 <- (x1 - mean(x1))/sd(x1)
z2 <- (x2 - mean(x2))/sd(x2)
```

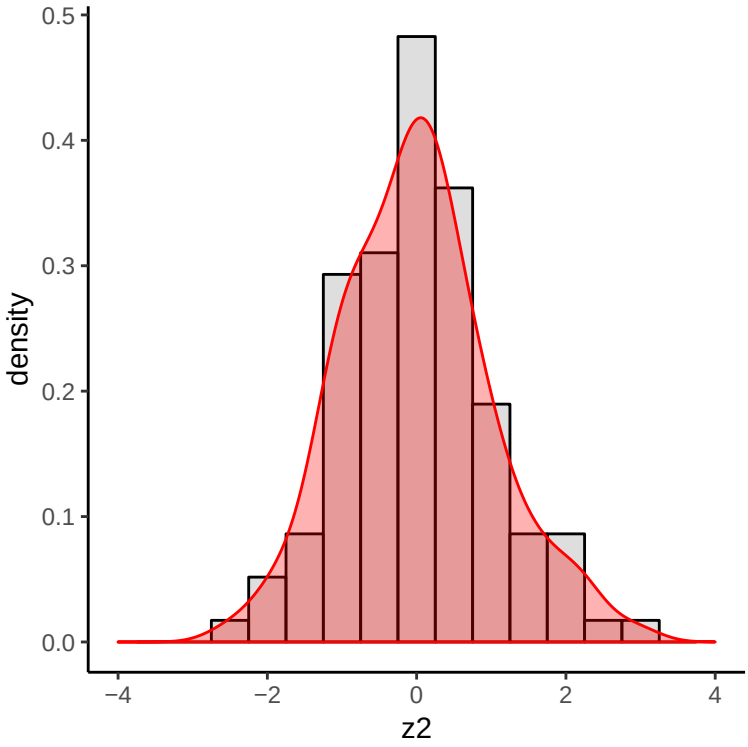
Step 2. Testing multivariate normality of phenotypic traits

2.1. Histograms

```
ggplot(data = data, aes(x = z1)) + xlim(-2, 4) +
  geom_histogram(aes(y = ..density..), col = "black",
                 fill = "gray", binwidth = 0.5, alpha = 0.5) +
  geom_density(col = "red", fill = "red", alpha = 0.3) + theme_classic()
```




```
ggplot(data = data, aes(x = z2)) + xlim(-4, 4) +
  geom_histogram(aes(y = ..density..), col = "black",
                 fill = "gray", binwidth = 0.5, alpha = 0.5) +
  geom_density(col = "red", fill = "red", alpha = 0.3) + theme_classic()
```



2.2. Multivariate normality tests

```
kable(mvn(data.frame(z1, z2), mvnTest = "hz")$multivariateNormality, digits = 3)
```

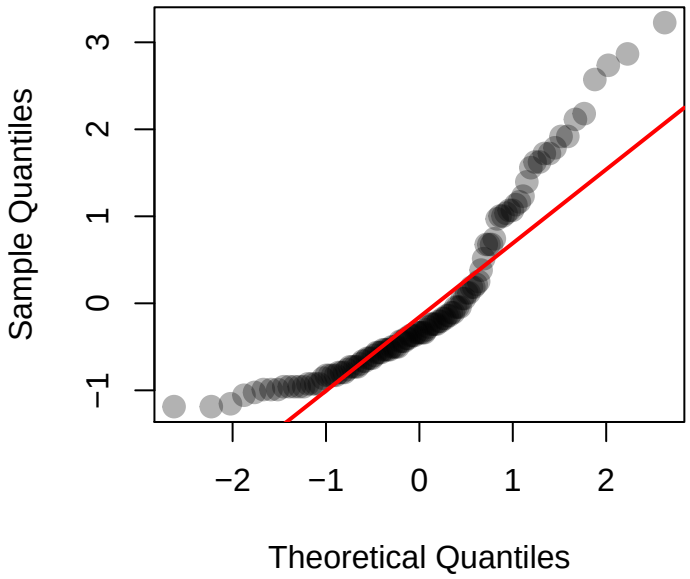
Test	HZ	p value	MVN
Henze-Zirkler	3.482	0	NO

```
kable(mvn(data.frame(z1, z2), mvnTest = "royston")$multivariateNormality, digits = 3)
```

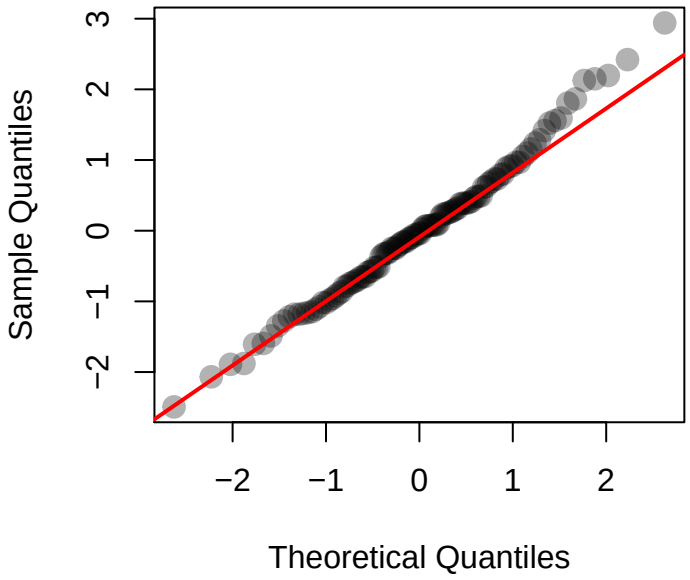
Test	H	p value	MVN
Royston	34.789	0	NO

2.3. Q-Q plots

```
qqnorm(z1, cex = 1.5, pch = 19, col = rgb(red = 0, green = 0, blue = 0, alpha = 0.3), main = "")
qqline(z1, col = "red", lwd = 2)
```




```
qqnorm(z2, cex = 1.5, pch = 19, col = rgb(red = 0, green = 0, blue = 0, alpha = 0.3), main = "")
qqline(z2, col = "red", lwd = 2)
```



Step 3. Assessing collinearity of phenotypic traits

3.1. Pearson correlation between traits

```
cor(z1, z2)
```

```
## [1] 0.2101184
```

3.2. Variance inflation factor on (linear) Lande and Arnold's model

```
lin.grad <- lm(wrel ~ z1 + z2)
nonlin.grad <- lm(wrel ~ z1 + z2 + I((1/2)*z1^2) + I((1/2)*z2^2) + z1:z2)

kable(vif(lin.grad), digits = 3)
```

	x
z1	1.046
z2	1.046

```
kable(summary(lin.grad)$coeff, digits = 3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.000	0.038	26.409	0
z1	0.652	0.039	16.752	0
z2	0.198	0.039	5.082	0

```
kable(summary(nonlin.grad)$coeff, digits = 3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.112	0.051	21.892	0.000
z1	0.780	0.055	14.289	0.000
z2	0.173	0.037	4.717	0.000
I((1/2) * z1^2)	-0.200	0.064	-3.109	0.002
I((1/2) * z2^2)	-0.086	0.056	-1.524	0.130
z1:z2	0.143	0.038	3.822	0.000

Step 4. Checking model residuals

4.1. Shapiro test

```
shapiro.test(resid(lin.grad))
```

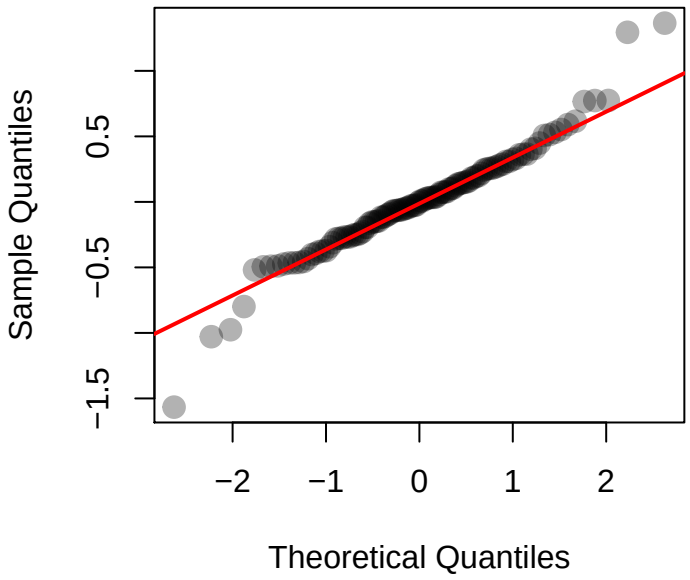
```
##
##  Shapiro-Wilk normality test
##
## data:  resid(lin.grad)
## W = 0.95743, p-value = 0.001003
```

```
shapiro.test(resid(nonlin.grad))
```

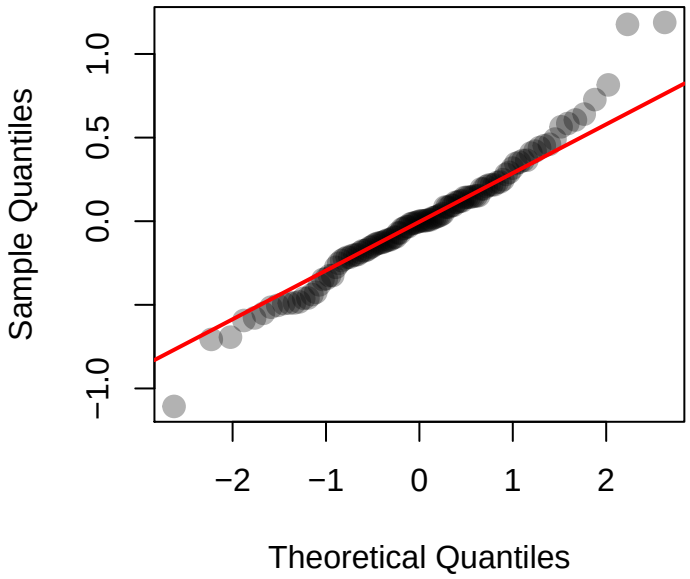
```
##
##  Shapiro-Wilk normality test
##
## data:  resid(nonlin.grad)
## W = 0.97573, p-value = 0.03342
```

4.2. Q-Q plots

```
qqnorm(resid(lin.grad), cex = 1.5, pch = 19,
       col = rgb(red = 0, green = 0, blue = 0, alpha = 0.3), main = "")
qqline(resid(lin.grad), col = "red", lwd = 2)
```




```
qqnorm(resid(nonlin.grad), cex = 1.5, pch = 19,
       col = rgb(red = 0, green = 0, blue = 0, alpha = 0.3), main = "")
qqline(resid(nonlin.grad), col = "red", lwd = 2)
```



Step 5. Standard error and confidence interval estimation

5.1. For the grad function, the data must have the relative fitness in the first column and standardized variables in the remaining columns. The function returns a vector with linear, quadratic and correlational gradients, and represents the input for the function boot.

```
grad <- function(data, original = c(1:nrow(data))){
  data <- data[original, ]
  vars  <- colnames(data)[-1]
  colnames(data)[1] <- "Wrel"
  model.lin <- as.formula(paste("Wrel", paste(vars, collapse=" + "), sep=" ~ "))
  m1 <- lm(formula = model.lin, data = data)
  part1 <- paste("(", paste(vars, collapse=" + "), ")^2", sep = "")
```

```
  part2 <- paste("I(0.5*(", vars, "^2))", sep = "", collapse = " + ")
  model.qua <- as.formula <- paste("Wrel", paste(part1, part2, sep = " + "), sep = " ~ ")
  m2 <- lm(formula = model.qua, data = data)
  sel.grad<-c(m1$coefficients[-1], m2$coefficients[-c(1:ncol(data))])
  return(sel.grad)
}

newdata <- data.frame(wrel, z1, z2)
selection.gradients <- grad(data = newdata)
boot.grad <- boot(data = newdata, statistic = grad, R = 999)
```

5.2. Create a list with 95% bias-corrected bootstrap confidence intervals for each gradient.

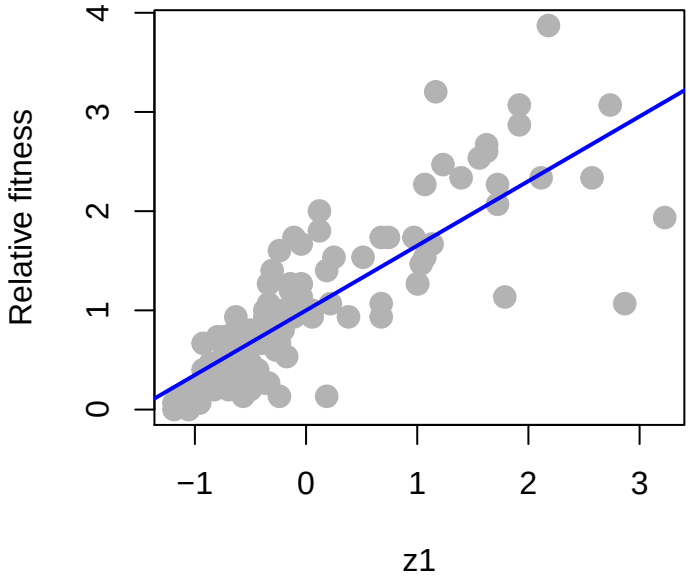
```
CI <- list()
for(i in 1:length(boot.grad$t0)){
  CI[[i]] <- boot.ci(boot.grad, conf = 0.95, type = "bca", index = i)$bca[4:5]
}

names(CI) <- names(boot.grad$t0)
CI <- do.call(rbind, CI)
colnames(CI) <-c("lower.ci", "upper.ci")
kable(CI, digits = 3)
```

	lower.ci	upper.ci
z1	0.505	0.777
z2	0.121	0.291
I(0.5 * (z1^2))	-0.354	-0.043
I(0.5 * (z2^2))	-0.239	0.049
z1:z2	0.024	0.267

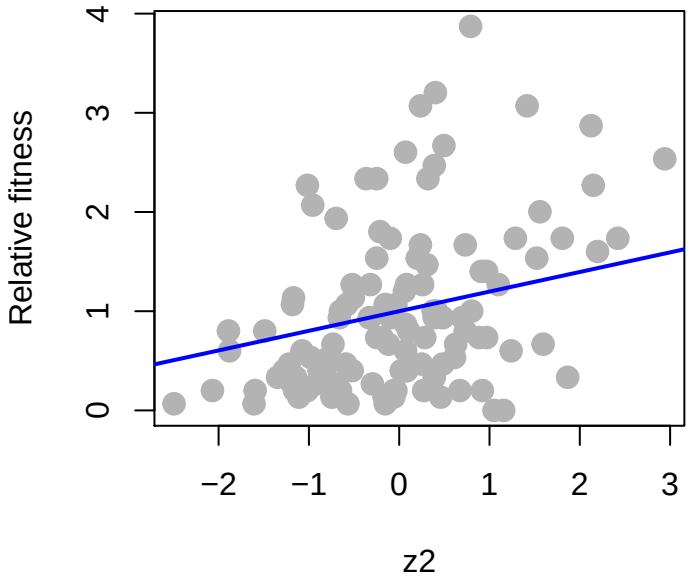
Step 6. Plotting Lande & Arnold's model results

6.1. Linear selection

```
new.z1 <- seq(-4, 4, length = 500)
plot(z1, wrel, pch = 19, cex = 1.5, col = "gray70", ylab = "Relative fitness")
pred.z1 <- predict(lin.grad, newdata = data.frame(z1 = new.z1, z2 = mean(z2)))
lines(new.z1, pred.z1, lwd = 2, col = "blue")
```

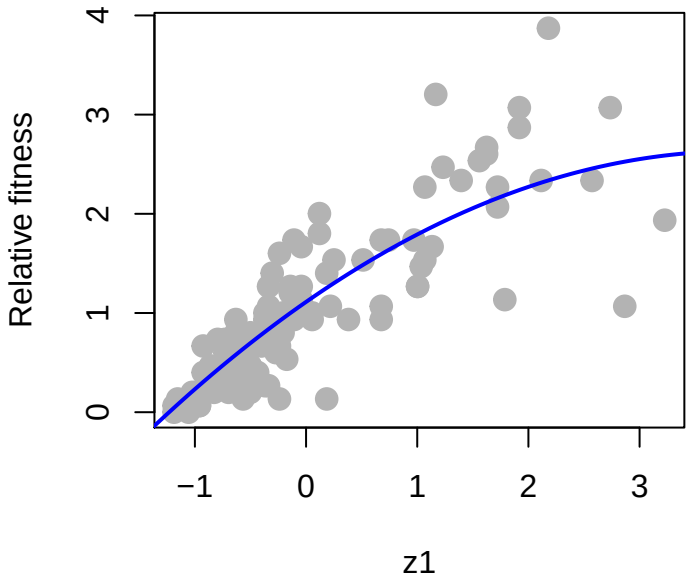



```
new.z2 <- seq(-4, 4, length = 500)
plot(z2, wrel, pch = 19, cex = 1.5, col = "gray70", ylab = "Relative fitness")
pred.z2 <- predict(lin.grad, newdata = data.frame(z1 = mean(z1), z2 = new.z2))
lines(new.z2, pred.z2, lwd = 2, col = "blue")
```

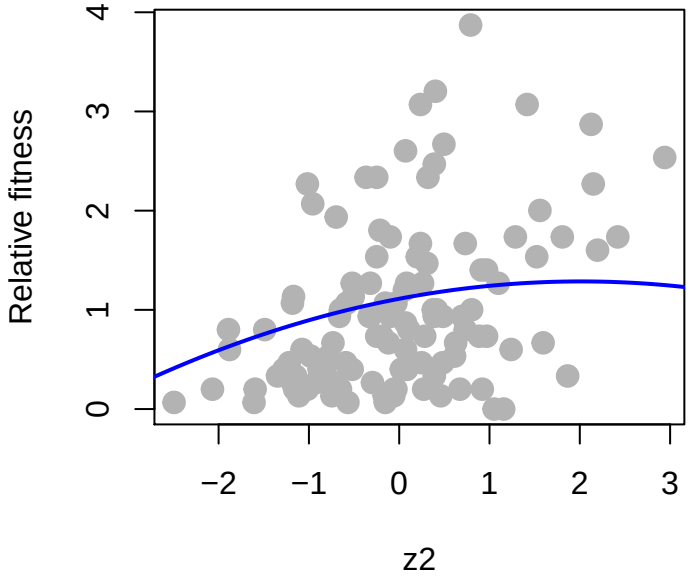


6.3. Quadratic selection

```
new.z1 <- seq(-4, 4, length = 500)
plot(z1, wrel, pch = 19, cex = 1.5, col = "gray70", ylab = "Relative fitness")
pred.z1 <- predict(nonlin.grad, newdata = data.frame(z1 = new.z1, z2 = mean(z2)))
lines(new.z1, pred.z1, lwd = 2, col = "blue")
```




```
new.z2 <- seq(-4, 4, length = 500)
plot(z2, wrel, pch = 19, cex = 1.5, col = "gray70", ylab = "Relative fitness")
pred.z2 <- predict(nonlin.grad, newdata = data.frame(z1 = mean(z1), z2 = new.z2))
lines(new.z2, pred.z2, lwd = 2, col = "blue")
```



6.4. Correlational selection

```
visreg2d(fit = nonlin.grad, xvar = "z1", yvar = "z2", scale = "response", plot.type = "image")
```