CrossMark
← click for updates

SOFTWARE TOOL ARTICLE

REVISED EvolQG - An R package for evolutionary quantitative genetics [version 3; referees: 1 approved, 2 approved with reservations]

Diogo Melo[1], Guilherme Garcia[1], Alex Hubbe[2], Ana Paula Assis[1], Gabriel Marroig[1]

[1]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil
[2]Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Salvador, Brazil

Abstract

We present an open source package for performing evolutionary quantitative genetics analyses in the R environment for statistical computing. Evolutionary theory shows that evolution depends critically on the available variation in a given population. When dealing with many quantitative traits this variation is expressed in the form of a covariance matrix, particularly the additive genetic covariance matrix or sometimes the phenotypic matrix, when the genetic matrix is unavailable and there is evidence the phenotypic matrix is sufficiently similar to the genetic matrix. Given this mathematical representation of available variation, the \textbf{EvolQG} package provides functions for calculation of relevant evolutionary statistics; estimation of sampling error; corrections for this error; matrix comparison via correlations, distances and matrix decomposition; analysis of modularity patterns; and functions for testing evolutionary hypotheses on taxa diversification.

R This article is included in the RPackage channel.

Open Peer Review

Referee Status: ✓ ? ?

	Invited Referees		
	1	**2**	**3**
REVISED **version 3** published 08 Nov 2016			
REVISED **version 2** published 27 Jun 2016		? report	? report
		↑	
version 1 published 30 Sep 2015	✓ report	? report	

1 **Mark Grabowsky**, University of Olso Norway

2 **David Houle**, Florida State University USA

3 **Charles Roseman**, University of Illinois at Urbana-Champaign USA

Discuss this article

Comments (0)

Corresponding author: Diogo Melo (diogro@usp.br)

Competing interests: No competing interests were disclosed.

Introduction

Quantitative genetics deals with the evolution and inheritance of continuous traits, like body size, bone lengths, gene expressions or any other inheritable characteristic that can be measured on a continuous scale, or which can be transformed to a continuous scale. This framework has been used in selective breeding and in describing the different sources of variation in natural populations, as well as understanding the interaction of evolutionary processes with this variation[32]. Quantitative genetics has been successful in describing short term evolution, and is also useful in understanding diversification at a macroevolutionary level. The core development of modern evolutionary quantitative genetics started with the generalization of the univariate breeders equation to the multivariate response to selection equation, derived by Lande and also referred to as the Lande equation[28,55]. The Lande equation relates the evolutionary change in trait means of a given population (Δz) to the interaction between the additive genetic variation (**G**-matrix) of this population and the directional selection (β) acting on this population. The additive genetic variation of a population is represented by a symmetric square matrix called the **G**-matrix, which contains the additive genetic variance of each trait on the diagonal and the additive genetic covariance between traits on the off-diagonal elements. From the Lande equation, $\Delta z = \mathbf{G}\beta$, we can see that different populations may present markedly different responses (Δz) to the same directional selection (β) simply because these populations have distinct **G**-matrices. Other evolutionary forces affecting populations are also influenced by available variation, e. g., based on the **G**-matrix it is possible to test if morphological differentiation of extant taxa is compatible with genetic drift or stabilizing selection (e.g., 2,35). Thus, describing and understanding changes in standing variation among populations[7,31,34] as well as understanding constraints imposed by populations standing variation (e.g., 19,33,51,58) are major elements in evolutionary quantitative genetics.

In this article we describe the **EvolQG** package, developed to deal with the evolutionary quantitative genetics questions addressed above in the R environment for statistical computing[53]. Our goal was to provide a suite of tools in a single consistent source, and to standardize and facilitate the adoption of these tools.

Measurement error estimation

Before estimating a **G**-matrix, it is important to evaluate the influence of measurement error in data collection, since measurement error can drastically bias further analyses[11]. Measurement error can be estimated by measuring each individual at least twice and estimating the amount of variation associated with each individual, which is the measurement error, in relation to total variation (i.e., the sum of within and between individuals variation) using an analysis of variance. The proportion of variance associated with among individual variation, and not within individual variation, is called the repeatability[30]. A repeatability of 1 means that no variation is associated with measurement error. The function **CalcRepeatability()** performs the calculation described in 30 for a set of multivariate traits measured at least twice for each individual.

Matrix estimation

In the rest of this article we assume that the covariance matrix of interest has already been estimated by some procedure. This matrix can be a simple covariance of all the observed traits, or an estimated parameter from a more complicated linear model. The simplest case of a linear model approach would be using a multivariate analysis of covariance (MANCOVA) to control for differences in trait means that are not of immediate interest in the analyses (e.g., sexual dimorphism, geographic variation, etc.). The residual pooled within-group covariance matrix can be used in subsequent analysis[34]. The **EvolQG** function **CalculateMatrix()** uses R's **lm()** model object to calculate variance-covariance matrices adjusting for the proper degrees of freedom in a simple fixed-effects MANCOVA. More complicated methods may be used to obtain **G**-matrices, such as an animal model or a mixed model[32,57], and these can be used for further analysis using **EvolQG**.

EvolQG also provides a simple Bayesian model for phenotypic matrix estimation using a conjugate inverse Wishart prior, implemented in the function **BayesianCalculateMatrix()**. The default behavior is to use a prior covariance matrix with the observed variances (from the data) on the diagonal and zeros in the off-diagonal elements. This method uses a type of regularization prior which provides a "shrinkage" posterior estimate which has some attractive properties when compared to the maximum likelihood estimate, such a lower mean squared error and a well behaved inverse[29,44,59]. The maximum *a posteriori* estimate is calculated analytically, and optionally the function also provides a sample from the posterior distribution and the median of these samples. These samples can be used in other Bayesian methods described in the next sections, or to calculate confidence intervals for any analysis using the covariance matrices.

Ideally all the analysis we describe should be performed on **G**-matrices, but accurate **G**-matrix estimation can be hard, requiring large sample sizes, many families and known genealogies[60]. One alternative we advocate for, which is usually more feasible, is to use the phenotypic covariance matrix (the **P**-matrix) as a proxy of the population's **G**-matrix[8,56]. Conditions on where this approximation is reasonable depend on the structure of developmental and environmental effects, and testing for similarity is an empirical question that should be undertaken before using the **P**-matrix as a proxy for the **G**-matrix, ideally by direct comparison (e.g., 13). As a general rule, high similarity between populations' **P**-matrices is a good indicator of high similarity between **P** and **G**, and of a stable shared **G**-matrix pattern, since the similarity between populations must come from either a common genetic structure, or the unlikely scenario of a different genetic structure buffered by an exactly compensating environmental structure in each population that leads to high similarity between phenotypic covariation.

Some of the methods described below are not applicable to covariance matrices, only to correlation matrices. Correlations are standardized measures of association that are bounded between [–1, 1], and, unlike covariances, can be directly compared for pairs of traits with different scales. In most instances, correlation matrices can be obtained directly from covariance matrices by using the R function **cov2cor()**.

Matrix error and repeatabilities

A **G**-matrix will always be estimated with error[20,37,41], and it is important to take this error into account in further analyses. In some circumstances we want to compare two or more **G**-matrices, calculating the matrices correlations (see section **Matrix Comparison**). However, due to error in estimating these matrices, their correlations will never be one, even if the actual population parameter values are identical[8]. Thus, matrix repeatabilities are used to correct matrix correlations by taking sampling error into account. The basic premise of all the methods is that taking repeated samples from the same population and comparing the resulting matrices would still give correlations that are lower than 1. We estimate the maximum correlation between matrices taken from the same population and correct the observed correlation by this maximum value. The corrected correlation between two observed matrices will be given by the original correlation divided by the geometric mean of their repeatabilities. If the repeatability of both matrices is one, the observed correlation does not change under the correction, and lower repeatabilities yield larger corrections. Estimating error for **G**- and **P**-matrices require different methods. The objective when estimating error is usually to create a sample of matrices that reflects our uncertainty regarding the estimated matrix. Because **G**-matrix are estimated using mixed models or animal models[32,57] this model structure must be taken into account, and sampling strategies that work in P-matrices will severely underestimate error in **G**-matrices. Given these restrictions, we can generate a distribution of **G**-matrices in several ways: (i) mixed model packages often provide the functionality of bootstraping over the exchangeable units in the design (i.e. permuting sires in a half-sib or full-sib design) and generate error estimates using these samples; (ii) Houle & Meyer[23] use an asymptotic approximation for the the parametric sampling distribution of the estimated **G**-matrix to generate samples and estimate error, a method called MVN-REML available in the program WOMBAT[40]; (iii) in a fully Bayesian context, samples from the posterior distribution of the **G**-matrix can be used. These samples can then be used to calculate repeatabilities or confidence intervals for derived quantities. Using any of these sampling methods the repeatability can be estimated as the average comparison value for all pairs of sampled matrices or as the average comparison of the estimated matrix and the sampled matrices.

When working with **P**-matrices things are simpler, and repeatabilities are straight forward to calculate using **EvolQG**. We provide a number of methods for estimating repeatability in **P**-matrices, and their results can be passed on to the functions that calculate matrix correlations (section **Matrix Comparison**):

AlphaRep(): Cheverud[8] describes an analytical expression for the repeatability of a correlation matrix. This expression is asymptotically convergent, so it should be used only when sample sizes are large, at least larger than the number of traits.

BootstrapRep(): We may estimate the repeatability of the covariance (or correlation) structure of a given data set using a bootstrap procedure, sampling individuals with replacement from the data set and calculating a covariance (or correlation) matrix from each sample. The mean value of the correlation between the random sample matrix and the original estimated matrix is an estimate of the repeatability. This method has the advantage of not assuming any distribution on the data, but does provide inflated estimates of repeatabilities for small data sets. Even so, upwardly biased matrix repeatabilities are not so problematic, because they lead to conservative corrections of matrix correlations. However, users should be aware of this bias and should not interpret a high repeatability obtained from a small data set as indicating that the parameter is well estimated.

MonteCarloRep(): We can use the observed covariance matrix as the Σ parameter in a multivariate normal distribution, and produce samples from this distribution, using a fixed sample size. The covariance (or correlation) matrix for each sample is compared to the observed matrix, and the mean of these comparisons is an estimate of the repeatability[34]. This method has the advantage of being easy to apply to matrices coming from linear models with many fixed effects, and not requiring the original data; but can also lead to inflated repeatabilities for small samples.

Both **MonteCarloRep()** and **BootstrapRep()** are based on generic functions (**MonteCarloStat()** and **BootstrapStat()**) which can be used for general sampling from covariance matrices using normal distributions or to generate bootstrap samples from a set of individuals, respectively. These functions can be used to generate confidence intervals for **P**-matrices for any of the evolutionary statistics described below, or any user defined function.

Sometimes the question we are trying to answer does not involve matrix comparisons, so other methods of assessing and correcting for error are needed.

Rarefaction(): Rarefaction consists of taking progressively smaller samples with replacement from the original data set, calculating some statistic on each data set and comparing this with the full data set. These comparisons gives a general idea of how the inferences would change if we had smaller sample sizes, and how robust our data set is with respect to sampling, given that an appropriately large initial sample is available. The default operation is to calculate the covariance or correlation matrices and compare them using any of the matrix comparison methods (see section **Matrix Comparison**). The generic **Rarefaction-Stat()** can also be used to generate rarefaction curves for any user defined function.

ExtendMatrix(): Marroig et al.[37] showed that sampling error on covariance matrix estimation can have a dramatic effect on the reconstruction of net selection gradients using the multivariate response to selection equation[28]. One way to improve estimates is the simple procedure of "extending" the eigenvalues of the covariance matrix, where all the eigenvalues lower than a certain threshold are substituted by the smallest eigenvalue above the threshold. This substitution causes minimal changes in the distribution of phenotypes, but improves dramatically the estimates of net selection gradients.

Alternatives to the extension method involve using regularized estimators for the covariance matrix, like Bayesian or shrinkage estimators[29,59], available from package **corpcor**. See 37 for a thorough examination of the performance and consequences of the extension and shrinkage methods on simulated and real data sets.

Evolutionary statistics

Hansen & Houle[19] provide a suite of statistics that have fairly good biological interpretations for a given **G**- or **P**-matrix. Marroig *et al.*[38] is a comprehensive example of how these statistics may be used for interpreting morphological data.

The function **MeanMatrixStatistics()** calculates most of these statistics and their distributions, as shown below. **MonteCarloStat()** and **BootstrapStat()** can be used to generate confidence intervals for these statistics when using **P**-matrices. Also, the previously available R package **evolvability**[5] implements some of these functions and provides confidence intervals.

In the following, $E[\cdot]_\beta$ represents the expected value over many random β vectors with unit norm, $< \cdot, \cdot >$ represents the dot product between two vectors, $cos(\cdot, \cdot)$ is the cosine between two vectors, **G** is an arbitrary covariance matrix, $\mathbf{G}^{-1}$ is the inverse **G**, $tr(\mathbf{G})$ is the trace of **G**, and $\| \cdot \|$ the Euclidean norm of a vector. **MeanMatrixStatistics()** calculates:

- Mean squared correlation (r^2): Given a correlation matrix, the elements below the diagonal are squared and averaged resulting in a measure of integration, that is, overall association between traits (also see the section **Modularity and Integration** and 49).

- Coefficient of variation of eigenvalues (ICV): A measure of integration that is suitable for covariance matrices, as it takes the amount of variation into account. Notice that at least for mammals, mean squared correlations and ICV generally have high correlation, but can lead to different conclusions if the traits included in the analysis have very different variances (due to scale, for example). If σ_λ is the standard deviation of the eigenvalues of a covariance matrix, and $\bar{\lambda}$ is the mean of the eigenvalues, the ICV is:

$$ICV = \frac{\sigma_\lambda}{\bar{\lambda}}$$

- Percent of variation in first principal component: If $\lambda_1^{\mathbf{G}}$ is the leading eigenvalue of **G**, we calculate this percentage as:

$$PC1\% = \frac{\lambda_1^{\mathbf{G}}}{tr(\mathbf{G})}$$

- Evolvability (Figure 1): The mean projection of the response to random selection gradients with unit norm onto the selection gradient. This projection is a measure of a population's available variation in the direction of a particular selection gradient, averaged across all directions[19].

$$\bar{e} = E[< \mathbf{G}\beta, \beta >]_\beta$$

- Flexibility (Figure 1): The mean cosine of the angle between random selection gradients and the
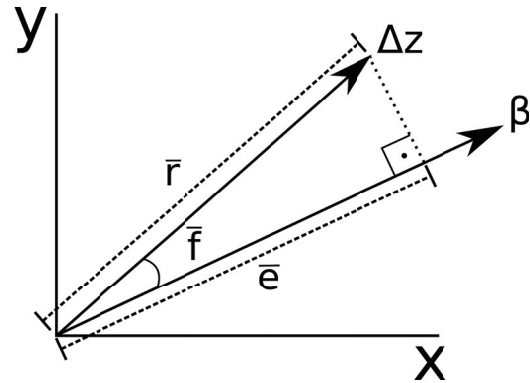

Figure 1. Graphical representation of evolvability ($\bar{e}$), respondability ($\bar{r}$) and flexibility ($\bar{f}$) for a single selection gradient (β) and the corresponding response (Δz) in the two dimensions defined by traits *x* and *y*.

corresponding responses. Flexibility measures on average how the response to selection aligns with the selection gradient[38].

$$\bar{f} = E[cos(\mathbf{G}\beta, \beta)]_\beta$$

- Respondability (Figure 1): Mean norm of the response to random selection gradients with unit norm. It also estimates how fast the population mean will change under directional selection[19,38].

$$\bar{r} = E[\|\mathbf{G}\beta\|]_\beta$$

- Conditional Evolvability: Measures the mean response to selection in the direction of a given β when other directions are under stabilizing selection[19].

$$\bar{c} = E\left[\left(< \mathbf{G}^{-1}\beta, \beta >\right)^{-1}\right]_\beta$$

- Autonomy: Measures the proportion of variance in the direction of a given β that is independent from variation in other directions. Therefore, mean Autonomy can also be calculated as the mean ratio between Conditional Evolvability ($\bar{c}$) and Evolvability ($\bar{e}$)[19].

$$\bar{a} = E\left[\left(< \mathbf{G}^{-1}\beta, \beta >\right)^{-1}\left(< \mathbf{G}\beta, \beta >\right)^{-1}\right]_\beta$$

- Constraints: The mean correlation between the response vector to random selection gradients and the matrix's first principal component[38]. If $\Lambda_1^{\mathbf{G}}$ is the first principal component of **G**, constraints are measured as:

$$E\left[cos\left(\mathbf{G}\beta, \Lambda_1^{\mathbf{G}}\right)\right]_\beta$$

Matrix comparison

A **G**-matrix describes how the variation in particular populations is structured, but frequently the relevant question is how similar or dissimilar two populations are with respect to this standing

variation. Because no two populations are identical, different patterns of variation are the norm. Depending on the evolutionary question at hand, different methods of comparing variation may be required. One possible application of matrix comparisons is when we wish to apply the Lande equation to micro and macro-evolution, because doing so requires some additional assumptions, such as a relative stability of the **G**-matrix over generations[24]. Comparing extant covariance matrices is a test of this required stability (e.g. [34]). For a thoughtful discussion on the biological relevance of statistical significance in matrix comparisons, see the discussion in [16].

Matrix correlations
One approach to estimate the similarity or dissimilarity between matrices is to calculate the correlation between these matrices. **EvolQG** provides several functions for pairwise matrix correlation.

RandomSkewers(): The Random Skewers (RS) method makes use of the Lande equation[28], $\Delta z = G\beta$, where Δz represents the vector of response to selection, **G** the **G**-matrix and β the directional selection vector, or selection gradient. In the RS method, the two matrices being compared are multiplied by a large number of normalized random selection vectors, and the resulting response vectors to the same selection vector are compared via a vector correlation (the cosine between the two vectors). The mean value of the correlation between the responses to the same selective pressure is used as a simple statistic of how often two populations respond similarly (in the same direction) to the same selective pressure:

$$RS(A,B) = E\left[cos\left(A\beta, B\beta\right)\right]_{\beta} \tag{1}$$

Where $E[\cdot]_{\beta}$ is the expected value over random selection vectors β. Significance in the random skewers comparison can be determined using a null expectation of correlation between random vectors. If the observed correlation between two matrices is above the 95% percentile of the distribution of correlations between random vectors, we consider the correlation significant and infer that there is evidence the two populations behave similarly under directional selection. Other implementations of the RS method sometimes resort to other forms of calculating significance, such as generating random matrices and creating a random distribution of correlations between matrices. Generating this distribution is difficult to do because generating random matrices with the properties of biological covariance structures is hard, see the **RandomMatrix()** function for a quick discussion on this problem. But perhaps more important than the significance test, the RS method measure the overall similarity of the responses predicted by the two matrices under directional selection, a property that is at the core of applications of quantitative genetics theory to evolutionary problems. In this sense, a significant similarity measure by Random Skewers might still be low for a given application, when small differences in the direction of response are important. The RS values range between -1 (the matrices have opposite structures) and 1 (the matrices share the same structure), and zero means the matrices have distinct structures.

MantelCor(): Correlation between matrices can be done using a simple Pearson correlation between the corresponding elements. Significance of this comparison must take the structure into account, so it is calculated by a permutation scheme, in which a null distribution is generated by permutation of rows and columns in one of the matrices and repeating the element-by-element correlation (i.e. Mantel test). The observed correlation is significant when it is larger than the 95% quantile of the permuted distribution. This method should only be used in correlation matrices, and cannot be used on covariance matrices because the variances might be very different, leading to large differences in the scale of the covariances. This scale difference can lead to a massive inflation in the correlation between matrices. The correlation between matrices range between -1 and 1, and higher correlations indicate matrices have more similar structures, null correlations indicate the matrices have distinct correlation structures. Correlations near zero can also occur if the elements of the matrices have nonlinear relations between them, as in all Pearson correlations. Negative correlations indicate the pattern of association between traits is reversed in the two matrices.

KzrCor(): The Krzanowski shared space, or Krzanowski correlation, measures the degree to which the first principal components (eigenvectors) span the same subspace[3,26], and is suitable for covariance or correlation matrices. If two $n \times n$ matrices are being compared, the first $k = \frac{n}{2} - 1$ principal components from one matrix are compared to the first k principal components of the other matrix using the square of the vector correlations, and the sum of the correlations is a measure of how congruent the spanned subspaces are. We can write the Krzanowski correlation in terms of the matrices' principal components (Λ_i^A being the i*th* principal component of matrix A):

$$KrzCor(A,B) = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k}cos^2(\Lambda_i^A, \Lambda_j^B) \tag{2}$$

The Krzanowski correlation values range between 0 (two subspaces are dissimilar) and 1 (two subspaces are identical).

PCAsimilarity(): The Krzanowski correlation compares only the subspace shared by roughly the first half of the principal components, but does not consider the amount of variation each population has in these directions of the morphological space[62]. In order to take the variation into account, we can add the eigenvalue associated with each principal component into the calculation, effectively weighting each correlation by the variance in the associated directions. If λ_i^A is the i*th* eigenvalue of matrix A, we have:

$$PCAsimilarity(A,B) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i^A\lambda_j^B cos^2(\Lambda_i^A, \Lambda_j^B)}{\sum_{i=1}^{n}\lambda_i^A\lambda_j^B} \tag{3}$$

Note the sum spans all the principal components, not just the first k as in the Krzanowski correlation method. This method gives correlations that are very similar to the RS method, but is much faster. The PCA similarity values range between 0 (the shared subspaces have no common variation) and 1 (the shared subspaces have identical variation).

Matrix decomposition
Some methods attempt to characterize how a set of matrices differ, going beyond simple correlations. **EvolQG** provides efficient implementations of some of these methods.

SRD(): The RS method can be extended to give information into which traits contribute to differences in terms of the pattern of correlated selection due to covariation between traits in two populations[36]. The Selection Response Decomposition does this by treating the terms of correlated response in the Lande equation as separate entities. Writing out the terms in the multivariate response to selection equation:

$$\mathbf{A}\beta = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix} = \qquad (4)$$

$$= \begin{pmatrix} A_{11}\beta_1 + A_{12}\beta_2 + \cdots + A_{1n}\beta_n \\ A_{21}\beta_1 + A_{22}\beta_2 + \cdots + A_{2n}\beta_n \\ \vdots \\ A_{n1}\beta_1 + A_{n2}\beta_2 + \cdots + A_{nn}\beta_n \end{pmatrix} = \Delta z$$

Separating the terms in the sums of the right hand side:

$$\begin{matrix} (A_{11}\beta_1 & A_{12}\beta_2 & \cdots & A_{1n}\beta_n) \\ (A_{21}\beta_1 & A_{22}\beta_2 & \cdots & A_{2n}\beta_n) \\ \vdots & \vdots & \ddots & \vdots \\ (A_{n1}\beta_1 & A_{n2}\beta_2 & \cdots & A_{nn}\beta_n) \end{matrix} \qquad (5)$$

Each of these row vectors $r_i^{A\beta} = (A_{ij}\beta_j)_{j=1\ldots n}$ are the components of the response to the selection gradient β on trait i. The term $A_{ii}\beta_i$ represents the response to direct selection on trait i, and the terms $(A_{ij}\beta_j)_{i\neq j}$ represent the response to indirect selection due to correlation with the other traits. Given two matrices, A and B, we can measure how similar they are in their pattern of correlated selection on each trait by calculating the correlation between the vectors r_i for each trait for random selection vectors of unit norm. The mean SRD score for trait i is then:

$$\mu SRD(A,B)_i = E\left[cor\left(r_i^{A\beta}, r_i^{B\beta}\right)\right]_\beta \qquad (6)$$

And the standard deviation of the correlations gives the variation in SRD scores:

$$\sigma SRD(A,B)_i = \sqrt{E\left[\left(cor\left(r_i^{A\beta}, r_i^{B\beta}\right) - \mu SRD_i\right)^2\right]_\beta} \qquad (7)$$

When the same trait in different matrices share a correlated response pattern, μSRD is high and σSRD is low; if the correlated response pattern is different, μSRD is low and σSRD is high. See 36 for details and examples.

RSprojection(): Aguirre et al.[3] used a modification of the Random Skewers method to examine differences in the magnitude of variation in different directions in a set of covariance matrices. This method is most useful when a posterior sample of covariance matrices is available, as this sample allows us to identify directions in the morphospace that show relevant differences in the amount of variation between any two matrices. The function **RSprojection()** uses a set of posterior samples to identify the directions that represent the most significant differences in variance, and to compare the amount of variation in each matrix for each of these directions. The posterior samples of covariance matrices can

be obtained using any Bayesian linear model package, such as MCMCglmm[17], or, for simple cases, **BayesianCalculateMatrix()**. This method identifies directions where there is a significant difference in the amount of variation, so it is most useful when the matrices being compared have similar scales, since matrices with identical structure might have different amounts of variation in all directions simply because the total amount of variation is larger in one of the matrices.

EigenTensorDecomposition(): Hine et al.[21] proposes using covariance tensors for characterizing covariance matrix variation, based on Basser & Pajevic[4]. Covariance tensors can be further decomposed into orthogonal eigentensors and eigenvalues; each covariance matrix in the sample can thus be represented as a combination of such eigentensors, in a manner analogous to Principal Component Analysis. Considering the non-Euclidean (Riemannian) nature of the space of symmetric positive-definite matrices (Figure 2), the implementation of this procedure in **EvolQG** relies on estimating a geometric mean matrix (which minimizes the sum of Riemannian distances among observations[43]; implemented as **MeanMatrix()**) and mapping such observations into an actual Euclidean space using the function

$$f(\mathbf{X}) = Log(\mathbf{M}^{-\frac{1}{2}}\mathbf{X}\mathbf{M}^{-\frac{1}{2}}) \qquad (8)$$

where **M** refers to the estimated geometric mean matrix and *Log* refers to the matrix logarithm operator. Only then a covariance tensor can be estimated, considering that the estimation of orthogonal eigentensors assumes that observations are contained within an Euclidean space, equipped with definitions for both angle and distance. The function **EigenTensorDecomposition()** implements these steps; furthermore, given an eigentensor decomposition estimated using this function, it is possible to project other covariance matrices of the same size as the original sample onto the obtained eigentensors (**ProjectMatrix()**), and also reconstruct covariance matrices based on their scores over eigentensors (**RevertMatrix()**), which may be useful for observing the direction associated with each eigentensor as actual covariance matrices. This function uses the inverse operation to equation 8, that is

$$f^{-1}(\mathbf{X}) = \mathbf{M}^{\frac{1}{2}}Exp(\mathbf{X})\mathbf{M}^{\frac{1}{2}}) \qquad (9)$$

where *Exp* refers to the matrix exponential operator. This operation maps the matrix **X** back onto the non-Euclidean space of symmetric positive-definite matrices. If a posterior sample of covariances matrices is available, this function can be used to implement the hypothesis-testing method described in Aguirre et al.[3] using covariance tensors, but this test also requires a null distribution of matrices that has to be tailored to the data at hand. Also, since the covariance tensor is calculated using the covariation between elements in the covariance matrix, these methods require two levels of replication in the data: individuals within populations for the calculation of the covariance matrices, and several populations for the calculation of the covariance tensor. This requirement should be kept in mind when interpreting the results, as small samples at either level can lead to poorly estimated parameters and misleading results.

Matrix distances
Another approach to estimate the similarity or dissimilarity between matrices is to calculate the distance between a pair of
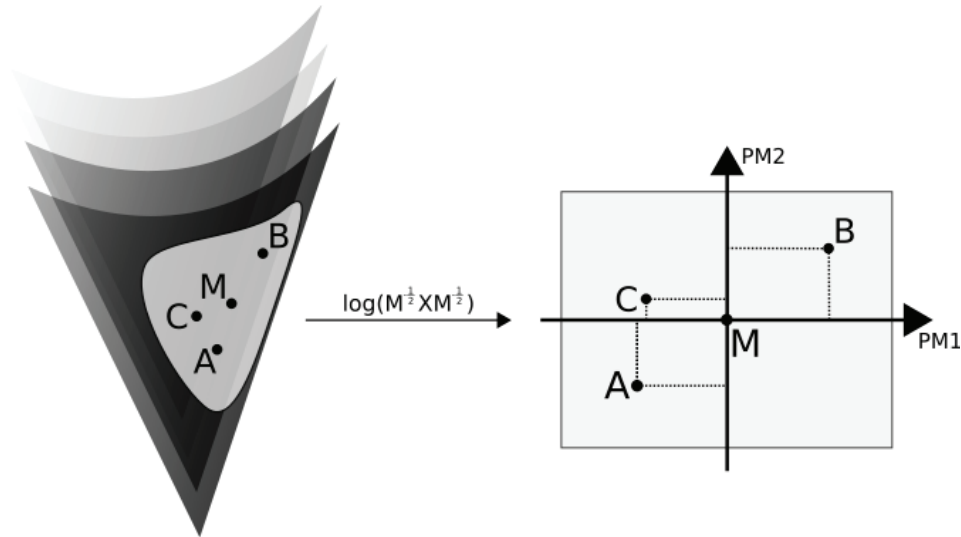
Figure 2. Graphical depiction of the eigentensor decomposition of covariance matrices A, B and C. The mean matrix M is estimated within the non-Euclidean space of symmetric positive-definite matrices; the transformation $f(X) = Log(M^{-\frac{1}{2}}XM^{-\frac{1}{2}})$ maps A, B and C into an Euclidean space centered on M. Only in this Euclidean space are the eigentensors (PM1 and PM2) estimated.

matrices. Matrices distances are different from correlations in that correlations are limited to $[-1, +1]$, while distances must only be positive. Also, smaller values of distances mean more similarity. Two distances are in use in the current evolutionary literature, and are implemented in the function **MatrixDistance()**.

- Overlap distance: Ovaskainen *et al.*[48] proposed a distance based on probability distributions, where two covariance matrices would have a distance proportional to how distinguishable they are. This distance is natural if we think of covariance matrices as describing the probability distribution of phenotypes or additive values in the population. The higher the probability of a random draw coming from the distribution defined by one of the matrices being misclassified as coming from the distribution defined by the other, the lower the distance. For two probability distributions f and g, the probability of misclassifying a draw from f as coming from g is:

$$q(f, g) = \int_{R^n} \frac{g(x)}{f(x) + g(x)} f(x) dx \qquad (10)$$

where n is the dimensionality of the space in which the distributions are defined. If the distributions are indistinguishable, $q(f, g) = 1/2$, if they are completely distinguishable $q(f, g) = 0$. We can then define the distance as:

$$d(f, g) = \sqrt{1 - 2q(f, g)} \qquad (11)$$

Since $q(f, g)$ is symmetrical, $d(f, g)$ is also symmetrical, and the square root guaranties that $d(f, g)$ satisfies the triangle inequality[48]. Calculation is straight forward and can be done with a simple sampling Monte Carlo scheme, see [48] for details.

- Riemann distance: Mitteroecker and Bookstein[42] use a Riemannian metric in the space of positive definite matrices (either covariance or correlation matrices), based on exponential mapping[43] to quantify transition in the ontogenetic trajectory of phenotypic covariance matrices. This metric is based on the eigenvalues of the product of one matrix to the inverse of the other. If λ_i are the eigenvalues of $A^{-1}B$ (or AB^{-1}), we have:

$$\|A, B\|_{cov} = \|B, A\|_{cov} = \sqrt{\sum_{i=1}^{p} \left[log(\lambda_i) \right]^2} \qquad (12)$$

This distance has the advantage of being invariable under changes in the base used to represent the matrices. See [42] for a discussion on the biological relevance of this distance.

Phylogenetic comparisons

PhyloW(): Given a set of covariance matrices for the terminal taxa in a phylogeny, we can estimate the covariance matrix for internal nodes by taking means over sister taxa, weighted by sample size. The mean matrix at the root node is the within-group covariance matrix in a MANCOVA with the terminal clades as the fixed effects. **PhyloW()** does this estimation by taking a tree and a set of measurements (covariance matrices) and returns means weighted by sample size for internal nodes. The implementation is generic, so this function can also be used to calculate weighted means for any numerical measurement with an addition operation implemented in R.

For G-matrices it might not be clear what the appropriate weight to use in each node, as the linear mixed models used in G-matrix estimation do not assign a clear degree of freedom for the covariance matrix. Number of families could plausibly be a good

weighing factor, but one should proceed with caution, or use a more direct approach of including the phylogeny in G-matrix estimation.

While using the within-group covariance matrix is a reasonable alternative as the estimator of an ancestral covariance matrix, it ignores branch lengths, and so should be used carefully when matrix differences are correlated to phylogenetic distance. An alternative when matrix evolution depends of branch lengths is to reconstruct every position of the covariance matrix independently via maximum likelihood, but this method can result in non positive-definite estimates.

PhyloCompare(): Sometimes it is not practical to pairwise compare every single population in a study, since for a large number of populations these results can be difficult to interpret. In these cases, comparing populations in a phylogeneticaly structured way can be helpful in detecting major transitions or differences between clades. **PhyloCompare()** takes estimates for all the nodes in a tree and compares sister groups by any comparison method, providing comparison values for every inner node.

Hypothesis testing
Modularity and integration
Modularity is a general concept in biology, and refers to a pattern of organization that is widespread in many biological systems. In modular systems, we find that some components of a given structure are more related or interact more between themselves than with other components. These highly related groups are termed modules. The nature of this interaction will depend on the components being considered, but may be of any kind, like physical contact between proteins, joint participation of enzymes in given biochemical pathways, or high correlation between quantitative traits in a population. This last kind of modularity is called variational modularity, and is characterized by high correlations between traits belonging to the same module and low correlation between traits in different modules[61]. In the context of morphological traits, variational modularity is associated with the concept of integration[47], that is, the tendency of morphological systems to exhibit correlations due to common developmental factors and functional demands[9,18].

Both modularity and integration may have important evolutionary consequences, since sets of integrated traits will tend to respond to directional selection in an orchestrated fashion due to genetic correlations between them; if these sets are organized in a modular fashion, they will also respond to selection independently of one another[38]. At the same time, selection can alter existing patterns of integration and modularity, leading to traits becoming more or less correlated[24,39]. The correlations between traits in a **G**-matrix then carries important information on the expected response to selection and on the history of evolutionary change of a given population.

TestModularity(): Variational modularity can be assessed by comparing a modularity hypothesis (derived from development and functional considerations) with the observed correlation matrix. If two traits are in the same variational module, we expect the correlation between them to be higher than between traits belonging to different modules. We test this partition by creating a modularity hypothesis matrix and comparing it via Pearson

correlation between the corresponding elements in the observed correlation matrix. The modularity hypothesis matrix consists of a binary matrix where each row and column corresponds to a trait. If the trait in row i is in the same module of the trait in column j, position (i, j) in the modularity hypothesis matrix is set to one, if these traits are not in the same module, position (i, j) is set to zero. Significant correlation between the hypothetical matrix representing a modularity hypothesis and the observed correlation matrix represents evidence of the existence of this variational module in the population. We also measure the ratio between correlations within a module (AVG+) and outside the module (AVG-). This ratio (AVG+/AVG-) is called the AVG Ratio, and measures the strength of the within-module association compared to the overall association for traits outside the module. The higher the AVG Ratio, the bigger the correlations within a module in relation to all other traits associations in the matrix (e.g., 50). **TestModularity()** also provides the Modularity Hypothesis Index, which is the difference between AVG+ and AVG- divided by the coefficient of variation of eigenvalues. Although the AVG Ratio is easier to interpret (how many times greater the within-module correlation is compared to the between-module correlation) than the Modularity Hypothesis Index, the AVG Ratio cannot be used when the observed correlation matrix presents correlations that differ in sign, and this is usually the case for residual matrices after size removal (for example with **RemoveSize()**, but see 25 for other alternatives). In these cases, the Modularity Hypothesis Index is useful and allows comparing results between raw and residual matrices[51].

LModularity(): If no empirical or theoretical information is available for creating a modularity hypothesis, such as functional or developmental data, we can try to infer the modular partition of a given population by looking only at the correlation matrix and searching for the trait partition that minimizes some indicator of modularity. Borrowing from network theory, we can treat a correlation matrix as a fully connected weighted graph, and define a Newman-like modularity index[45]. If A is a correlation matrix we define L modularity as:

$$L = \sum_{i \neq j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j) \tag{13}$$

The terms g_i and g_j represent the partition of traits, that is, in what modules the traits i and j belong to. The function $\delta(\cdot, \cdot)$ is the Kronecker delta, where:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \tag{14}$$

This means only traits in the same module contribute to the value of L. The term k_i represent the total amount of correlation attributed to trait i, or the sum of the correlation with trait i:

$$k_i = \sum_{j \neq i} A_{ij} \tag{15}$$

And m is the sum of all k ($m = \sum_i k_i$). The term $\frac{k_i k_j}{2m}$ plays the role of a null expectation for the correlation between the traits i and j. This choice for the null expectation is natural when we impose that it must depend on the values of k_i and k_j and must be symmetrical[45]. So, traits in the same module with correlations higher than the null

expectation will contribute to increase the value of L, while traits in the same module with correlation less than the null expectation will contribute to decrease L. With this definition of L, we use an optimization procedure to find the partition of traits (values of g_i) that maximizes L. This partition corresponds to the modularity hypothesis inferred from the correlation matrix, and the value of L is a measure of modularity comparable to the AVG Ratio. The igraph package[10] provides a number of community detection algorithms that can be used on correlation matrices using this function.

RemoveSize(): If the first principal component of a covariance or correlation matrix corresponds to a very large portion of its variation, and all (or most) of the entries of the first principal component are of the same sign (a *size* principal component, see 33), it is useful to look at the structure of modularity after removing this dominant integrating factor. This removal is done using the method described in 6. Porto *et al.*[51] show that modularity is frequently more easily detected in matrices where the first principal component variation was removed and provide biological interpretations for these results.

Drift

Selection is frequently invoked to explain morphological diversification, but the null hypothesis of drift being sufficient to explain current observed patterns must always be entertained. We can test the plausibility of drift for explaining multivariate diversification by using the regression method[1], or the correlation of principal component scores[1,35]. Since both these tests use drift as a null hypothesis, failure to reject the null hypothesis is not evidence that selection was not involved in the observed pattern of diversification, only that the observed pattern is compatible with drift. Also, these methods assume that the matrices involved share some degree of similarity, and should ideally be proportional to each other. We would be very weary of using these methods if the matrices are too dissimilar, or if the results change radically if different matrices are used as the ancestral matrix. Also, these tests rely on two levels of replication, taxa and traits. As a general guideline, at least 20 traits and at least 8 taxa should be sampled for using these methods with any confidence, and results should be analyzed in conjunction with other lines of evidence.

DriftTest(): Under drift and without gene flow, we expect that the current between group variance for many populations will be proportional to the ancestral population's covariance structure, which is approximated by the pooled within-group covariance matrix. This test assumes matrices remain proportional to the ancestral matrix, but this might not always be the case[14]. Conditions for the validity of these assumptions are reviewed in 52, and matrices for the extant groups should always be tested for similarity. Under these conditions, if B is the between group covariance matrix, and W is the within group covariance matrix, t is the time in number of generations and N_e is the effective population size, we have:

$$B \propto (t/N_e)W \qquad (16)$$

If we express all these matrices in terms of the eigenvectors of W, so that W is diagonal, we can write B as the variance of the scores of the means on these eigenvectors. The relationship between B and W can be expressed as a log regression, where B_i is the variance between groups in the projected means and λ_i^W are the eigenvalues of W:

$$log(B_i) = log(t/N_e) + \beta \, log(\lambda_i^W) \qquad (17)$$

where β is the regression coefficient. Under drift we expect β to be one. If β is significantly different from one, we have evidence that drift is not sufficient to explain currently observed diversification.

MultivDriftTest(): This drift test verifies the plausibility of drift in a multivariate context when only two populations are available, one ancestral (or reference) and one derived. Let z_0 represent a vector of means from m traits in an ancestral population. After t generations, the expected traits mean for n populations under drift would correspond to z_0 with variance given by $B = (t/N_e)W$, where B represents the expected between group covariance matrix, W is the genetic covariance matrix from the ancestral (or reference) population, and N_e is the effective population size[22,27,28]. So, given the ancestral population mean and G-matrix, we can use this model to estimate the B-matrix expected under drift. We can then use this B-matrix as the Σ parameter in a multivariate normal distribution and sample n populations from this distribution. Using this sample of random populations, we can assess the amount of divergence expected by drift, estimated as the norm of the difference vectors between ancestral (or reference) and simulated population means. Then, we can compare the observed amount of divergence between the ancestral and derived populations, calculated as the norm of the difference vector between them, taking into account the standard error of traits means. An observed divergence higher than the expectations under drift indicates that genetic drift is not sufficient to explain currently observed divergence, suggesting a selective scenario.

PCScoreCorrelation(): This test of drift relies on the correlation between principal component scores of different populations. Under drift alone, we expect the mean scores of different populations in the principal components of the within-group covariance matrix to be uncorrelated[35]. Significant correlations between the scores of the means on any two principal components is an indication of correlated directional selection[12].

Random matrices

RandomMatrix(): Generating realistic random covariance matrices for null hypothesis testing is a challenging task, because random matrices must adequately sample the space of biologically plausible evolutionary parameters, like integration and flexibility. Most common covariance and correlation matrix sampling schemes fail at this, producing matrices with unrealistically low levels of integration, unless the level of integration is supplied *a priori* (as in 15). The function **RandomMatrix()** implements the method described in 46, which provides correlation matrices with a reasonable range of evolutionary characteristics. However, the adequacy of the generated matrices in hypothesis testing has not been well established, and we recommend these random matrices be used only for informal tests requiring an arbitrary covariance or correlation matrix.

Summary

We have described a suite of functions dedicated to analyzing multivariate data sets within an evolutionary quantitative genetics framework. These functions focus on the central role that covariance and correlation matrices play in this framework; therefore, we provide functions that perform both descriptive statistics and hypothesis testing related to such matrices within an evolutionary context.

We have intentionally neglected to include techniques like phylogenetic regression or more extensive linear model functionality. Our reasons for this are twofold: the difficulty in transposing these methods efficiently to multiple traits, and the many different robust packages for performing some of these analyses, such as phytools, phylolm, pgls, nlme, MCMCglmm and others.

Some of the material implemented here is available in other sources or through custom implementations. We have attempted to create a single consistent source for these techniques. This is by no means an exhaustive effort, and we hope to expand it given demand from the community and further developments in the field. We hope to contribute to standardization and wide adoption of these tools, and, since we opted for an open source implementation under R, this also allows the involvement of the R community in using, debugging and complementing these tools, in an effort to contribute to an open scientific environment in which, for example, truly reproducible results are the norm rather than the exception.

Software availability

The most recent version of the **EvolQG** package can be installed from GitHub using the package **devtools**:

```
> library(devtools)
> install_github("lem-usp/evolqg")
```

A less up-to-date version is also available from CRAN:

```
> install.packages("evolqg")
```

1. Software available from: http://cran.r-project.org/web/packages/evolqg/

2. Latest source code: https://github.com/lem-usp/EvolQG

3. Archived source code as at time of publication: http://dx.doi.org/10.5281/zenodo.55121[63]

4. License: The MIT License (https://opensource.org/licenses/MIT)

References

1. Ackermann RR, Cheverud JM: **Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation.** *Am J Phys Anthropol.* 2002; **117**(3): 260–271.
   PubMed Abstract | **Publisher Full Text**

2. Ackermann RR, Cheverud JM: **Detecting genetic drift versus selection in human evolution.** *Proc Natl Acad Sci U S A.* 2004; **101**(52): 17946–17951.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

3. Aguirre JD, Hine E, McGuigan K, *et al.*: **Comparing G: multivariate analysis of genetic variation in multiple populations.** *Heredity (Edinb).* 2014; **112**(1): 21–29.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Basser PJ, Pajevic S: **Spectral decomposition of a 4th-order covariance tensor: Applications to diffusion tensor MRI.** *Signal Process.* 2007; **87**(2): 220–236.
   Publisher Full Text

5. Bolstad GH, Hansen TF, Pélabon C, *et al.*: **Genetic constraints predict evolutionary divergence in *dalechampia* blossoms.** *Philos Trans R Soc Lond B Biol Sci.* 2014; **369**(1649): 20130255.
   PubMed Abstract | **Publisher Full Text** | **Free Full Text**

6. Bookstein FL, Chernoff B, Elder RL, *et al.*: **Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes.** 1985.
   Reference Source

7. Cheverud JM: **Phenotypic, Genetic, and Environmental Morphological Integration in the Cranium.** *Evolution.* 1982; **36**(3): 499–516.
   Publisher Full Text

8. Cheverud JM: **A Comparison of Genetic and Phenotypic Correlations.** *Evolution.* 1988; **42**(5): 958–968.
   Publisher Full Text

9. Cheverud JM: **Developmental Integration and the Evolution of Pleiotropy.** *Integr Comp Biol.* 1996; **36**(1): 44–50.
   Publisher Full Text

10. Csardi G, Nepusz T: **The igraph software package for complex network research.** *Inter Journal.* Complex Systems:1695, 2006.
    Reference Source

11. Falconer DS, Mackay TF: **Introduction to Quantitative Genetics.** Benjamin Cummings, London, 4th edition, 1996.
    Reference Source

12. Felsenstein J: **Phylogenies And Quantitative Characters.** *Annu Rev Ecol Syst.* 1988; **19**(1): 445–471.
    Publisher Full Text

13. Garcia G, Hingst-Zaher E, Cerqueira R, *et al.*: **Quantitative Genetics and Modularity in Cranial and Mandibular Morphology of *Calomys expulsus*.**

Evol Biol. 2014; **41**(4): 619–636.
Publisher Full Text

14. Griswold CK, Logsdon B, Gomulkiewicz R: **Neutral evolution of multiple quantitative characters: a genealogical approach.** *Genetics.* 2007; **176**(1): 455–466.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

15. Haber A: **A Comparative Analysis of Integration Indices.** *Evol Biol.* 2011; **38**(4): 476–488.
Publisher Full Text

16. Haber A: **The Evolution of Morphological Integration in the Ruminant Skull.** *Evol Biol.* 2015; **42**(1): 99–114.
Publisher Full Text

17. Hadfield JD: **MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package.** *J Stat Softw.* 2010; **33**(2): 1–22.
Publisher Full Text

18. Hallgrímsson B, Jamniczky H, Young NM, *et al.*: **Deciphering the Palimpsest: Studying the Relationship Between Morphological Integration and Phenotypic Covariation.** *Evol Biol.* 2009; **36**(4): 355–376.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

19. Hansen TF, Houle D: **Measuring and comparing evolvability and constraint in multivariate characters.** *J Evol Biol.* 2008; **21**(5): 1201–1219.
PubMed Abstract | **Publisher Full Text**

20. Hill WG, Thompson R: **Probabilities of Non-Positive Definite between-Group or Genetic Covariance Matrices.** *Biometrics.* 1978; **34**(3): 429–439.
Publisher Full Text

21. Hine E, Chenoweth SF, Rundle HD, *et al.*: **Characterizing the evolution of genetic variance using genetic covariance tensors.** *Philos Trans R Soc Lond B Biol Sci.* 2009; **364**(1523): 1567–78.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Hohenlohe PA, Arnold SJ: **MIPoD: a hypothesis-testing framework for microevolutionary inference from patterns of divergence.** *Am Nat.* 2008; **171**(3): 366–385.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

23. Houle D, Meyer K: **Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood.** *J Evol Biol.* 2015; **28**(8): 1542–9.
PubMed Abstract | **Publisher Full Text**

24. Jones AG, Arnold SJ, Bürger R: **Evolution and stability of the G-matrix on a landscape with a moving optimum.** *Evolution.* 2004; **58**(8): 1639–1654.
PubMed Abstract | **Publisher Full Text**

25. Jungers WL, Falsetti AB, Wall CE: **Shape, relative size, and size-adjustments in morphometrics.** *Am J Phys Anthropol.* 1995; **38**(S21): 137–161.
Publisher Full Text

26. Krzanowski WJ: **Between-Groups Comparison of Principal Components.** *J Am Stat Assoc.* 1979; **74**(367): 703–707.
Publisher Full Text

27. Lande R: **Natural Selection and Random Genetic Drift in Phenotypic Evolution.** *Evolution.* 1976; **30**(2): 314–334.
Publisher Full Text

28. Lande R: **Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry.** *Evolution.* 1979; **33**(1): 402–416.
Publisher Full Text

29. Ledoit O, Wolf M: **A well-conditioned estimator for large-dimensional covariance matrices.** *J Multivar Anal.* 2004; **88**(2): 365–411.
Publisher Full Text

30. Lessells CM, Boag PT: **Unrepeatable repeatabilities: a common mistake.** *Auk.* 1987; **104**(1): 116–121.
Publisher Full Text

31. Lofsvold D: **Quantitative Genetics of Morphological Differentiation in Peromyscus. I. Tests of the Homogeneity of Genetic Covariance Structure Among Species and Subspecies.** *Evolution.* 1986; **40**(3): 559–573.
Publisher Full Text

32. Lynch M, Walsh B: **Genetics and analysis of quantitative traits.** Sinauer, Sunderland, MA., 1998.
Reference Source

33. Marroig G, Cheverud J: **Size as a line of least resistance II: Direct selection on size or correlated response due to constraints.** *Evolution.* 2010; **64**(5): 1470–1488.
PubMed Abstract | **Publisher Full Text**

34. Marroig G, Cheverud JM: **A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of new world monkeys.** *Evolution.* 2001; **55**(12): 2576–2600.
PubMed Abstract | **Publisher Full Text**

35. Marroig G, Cheverud JM: **Did natural selection or genetic drift produce the cranial diversification of neotropical monkeys?** *Am Nat.* 2004; **163**(3): 417–428.
PubMed Abstract | **Publisher Full Text**

36. Marroig G, Melo D, Porto A, *et al.*: **Selection Response Decomposition (SRD): A New Tool for Dissecting Differences and Similarities Between Matrices.** *Evol Biol.* 2011; **38**(2): 225–241.
Publisher Full Text

37. Marroig G, Melo DA, Garcia G: **Modularity, noise, and natural selection.** *Evolution.* 2012; **66**(5): 1506–1524.
PubMed Abstract | **Publisher Full Text**

38. Marroig G, Shirai LT, Porto A, *et al.*: **The Evolution of Modularity in the**

39. Melo D, Marroig G: **Directional selection can drive the evolution of modularity in complex traits.** *Proc Natl Acad Sci U S A.* 2015; **112**(2): 470–475.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

40. Meyer K: **WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML).** *J Zhejiang Univ Sci B.* 2007; **8**(11): 815–821.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

41. Meyer K, Kirkpatrick M: **Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices.** *Genetics.* 2008; **180**(2): 1153–66.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

42. Mitteroecker P, Bookstein F: **The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans.** *Evolution.* 2009; **63**(3): 727–737.
PubMed Abstract | **Publisher Full Text**

43. Moakher M: **On the Averaging of Symmetric Positive-Definite Tensors.** *J Elast.* 2006; **82**(3): 273–296.
Publisher Full Text

44. Murphy KP: **Machine learning: a probabilistic perspective.** MIT press, 2012.
Reference Source

45. Newman ME: **Modularity and community structure in networks.** *Proc Natl Acad Sci U S A.* 2006; **103**(23): 8577–8582.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

46. Numpacharoen K, Atsawarungruangkit A: **Generating correlation matrices based on the boundaries of their coefficients.** *PLoS One.* 2012; **7**(11): e48902.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

47. Olson R, Miller E: **Morphological integration.** University of Chicago Press, Chicago, 1958.
Reference Source

48. Ovaskainen O, Cano JM, Merilä J: **A Bayesian framework for comparative quantitative genetics.** *Proc Biol Sci.* 2008; **275**(1635): 669–678.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

49. Pavlicev M, Cheverud JM, Wagner GP: **Measuring Morphological Integration Using Eigenvalue Variance.** *Evol Biol.* 2009; **36**(1): 157–170.
Publisher Full Text

50. Porto A, de Oliveira FB, Shirai LT, *et al.*: **The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes.** *Evol Biol.* 2009; **36**(1): 118–135.
Publisher Full Text

51. Porto A, Shirai LT, de Oliveira FB, *et al.*: **Size variation, growth strategies, and the evolution of modularity in the mammalian skull.** *Evolution.* 2013; **67**(11): 3305–3322.
PubMed Abstract | **Publisher Full Text**

52. Prôa M, O'Higgins P, Monteiro LR: **Type I error rates for testing genetic drift with phenotypic covariance matrices: a simulation study.** *Evolution.* 2013; **67**(1): 185–195.
PubMed Abstract | **Publisher Full Text**

53. R Core Team: **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2014.
Reference Source

54. Revell LJ: **phytools: An R package for phylogenetic comparative biology (and other things).** *Methods Ecol Evol.* 2012; **3**(2): 217–223.
Publisher Full Text

55. Roff D: **Evolutionary quantitative genetics: Are we in danger of throwing out the baby with the bathwater?** *Ann Zool Fennici.* 2003; **40**(4): 315–320.
Reference Source

56. Roff DA: **The estimation of genetic correlations from phenotypic correlations: a test of Cheverud's conjecture.** *Heredity.* 1995; **74**(5): 481–490.
Publisher Full Text

57. Runcie DE, Mukherjee S: **Dissecting high-dimensional phenotypes with Bayesian sparse factor analysis of genetic covariance matrices.** *Genetics.* 2013; **194**(3): 753–767.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

58. Schluter D: **Adaptive Radiation Along Genetic Lines of Least Resistance.** *Evolution.* 1996; **50**(5): 1766–1774.
Publisher Full Text

59. Schäfer J, Strimmer K: **A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics.** *Stat Appl Genet Mol Biol.* 2005; **4**: Article32.
PubMed Abstract | **Publisher Full Text**

60. Steppan SJ, Phillips PC, Houle D: **Comparative quantitative genetics: evolution of the G matrix.** *Trends Ecol Evol.* 2002; **17**(7): 320–327.
Publisher Full Text

61. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Rev Genet.* 2007; **8**(12): 921–931.
PubMed Abstract | **Publisher Full Text**

62. Yang K, Shahabi C: **A PCA-based similarity measure for multivariate time series**. In *Proceedings of the 2nd ACM international workshop on Multimedia databases.* ACM, 2004; 65–74.
Publisher Full Text

63. Melo D, Alvarenga EZ, Hubbe A, *et al.*: **EvolQG: F1000Research v2.0.** *Zenodo.* 2016.
Data Source

Mammalian Skull II: Evolutionary Consequences. *Evol Biol.* 2009; **36**(1): 136–148.
Publisher Full Text

Open Peer Review

Current Referee Status: ✔ ❓ ❓

Version 2

❓ **Charles Roseman**

Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Thank you for the opportunity to review the article "EvolQG - An R package for evolutionary quantitative genetics." I am reviewing the version of 27 June 2016 (V.2). My overall impression of the paper and the associated package, which I have gone through in some detail, is that it makes a useful contribution to the study of the evolution of complex traits. Having many of these techniques available in an easily useable form will be a a boon to many researchers. I have some reservations that echo those of reviewers of previous and current versions of this manuscript.

I think that a useful revision to the paper might be to clearly distinguish between applications and treatments of the P and G matrices. For morphological characteristics, I concur that there are good reasons to believe that P and G will be roughly proportional to one another (there are issues with what that might mean for evolutionary analyses), but handling errors in either case are rather different. The BootStrapRep() and MonteCarloRep() functions will work as described for a phenotypic correlation matrix, but not for a genetic correlation matrix, which would require some kind of resampling or simulation procedure that considered relatedness among individuals. This is not easy even in simple circumstances like those found in experimental studies (full-sib designs and the like) and very taxing in the kinds of pedigree based studies of wild organisms that are becoming the norm in the field.

A work-around that I would accept only for analysis of legacy results from the literature would be to include a way to include effective sample sizes (Cheverud 1996, 1988). This would give an estimate of the number of effectively independently estimated breeding values worth of information in a sample and would be a better reflection of the errors in the genetic quantities. In the absence of this, adding explicit ways of working with the posterior distributions or estimated matrix of errors from the estimate of the G matrix would be welcome in your package. At the very least, you should make the distinction between these cases clearer in the text.

I think the random skewers procedure should be revisited in one respect. The Marroig and Cheverud (2007) procedure describes drawing random selection gradient values from a uniform distribution and then standardizing the resulting vector to unit length. This is meant to provide random directions in the phenotypic space. This procedure, however, does not achieve this as the distribution of random points before standardization is from a hypercube of the same dimensionality as the number of traits. This will lead to an uneven distribution of points around the hypersphere on which the unit length vectors of standardized simulated selection gradients would occupy. In effect, the extra space in the corners of the hypersphere cause patches of excess vectors. This could cause considerable mischief if, for example, the first principal component of a covariance matrix was oriented directly at corner of the hypercube. It might

be worth modifying the procedure to draw random Gaussian variables, which are then standardized to unit length following Muller (1959) and Marsaglia (1972). This would ensure truly random simulated directional selection.

I would like to second Prof. Houle's point that there are many aspects of the different techniques that do not have much to do with evolution in an explicit way. Many of the matrix comparisons don't have much to do with the degree to which responses to evolution might differ given a difference in the matrices. We could easily extend this critique to the parts of the package linking covariation to development. Phenotypic covariance/correlation matrices underdetermine the developmental processes through which covariation is generated and without some kind of independent assessment of developmental process. LModularity() arbitrarily assumes a lot about how organisms work. Without some extra information about the perturbations to developmental process, we have no idea what the biological meaning resulting clusters of traits might be or whether they relate to development at all.

That said, I do think most of the techniques contained in the package will be very useful for many people. On balance, I regard EvolQG as a welcome contribution to the toolkit available to scholars interested in the evolution of complex traits. I would like to see the issues dealing with the errors in G and P to be clarified and the correction to the sampling in the random skewers to be implemented.

Some smaller comments:

"The Krzanowski shared space, or Krzanowski correlation, measures the degree to which the first principal components (eigenvectors) span the same subspace"

This sentence and the section in which it is nested is a little opaque. Please reword so that it is clear how many principal components we need to consider.

"transposing these methods efficiently to multiple traits, and the many different robust packages for performing some of these analyses.."

"Transposing" does not strike me as the right word here. Please reword.

References
1. Cheverud J: A Comparison of Genetic and Phenotypic Correlations. *Evolution*. 1988; **42** (5). Publisher Full Text
2. Cheverud J: Quantitative genetic analysis of cranial morphology in the cotton-top (Saguinus oedipus) and saddle-back (S. fuscicollis) tamarins. *Journal of Evolutionary Biology*. 1996; **9** (1): 5-42 Publisher Full Text
3. Marsaglia G: Choosing a Point from the Surface of a Sphere. *The Annals of Mathematical Statistics*. 1972; **43** (2): 645-646 Publisher Full Text
4. Muller M: A note on a method for generating points uniformly on n-dimensional spheres. *Communications of the ACM*. 1959; **2** (4): 19-20 Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 31 Oct 2016

Diogo Melo, Universidade de São Paulo, Brazil

Thank you for the careful review! We respond to some of the specific comments below:

I think that a useful revision to the paper might be to clearly distinguish between applications and treatments of the P and G matrices. For morphological characteristics, I concur that there are good reasons to believe that P and G will be roughly proportional to one another (there are issues with what that might mean for evolutionary analyses), but handling errors in either case are rather different. The BootStrapRep() and MonteCarloRep() functions will work as described for a phenotypic correlation matrix, but not for a genetic correlation matrix, which would require some kind of resampling or simulation procedure that considered relatedness among individuals. This is not easy even in simple circumstances like those found in experimental studies (full-sib designs and the like) and very taxing in the kinds of pedigree based studies of wild organisms that are becoming the norm in the field.

This is correct, and we have modified the text to make this clear.

A work-around that I would accept only for analysis of legacy results from the literature would be to include a way to include effective sample sizes (Cheverud 1996, 1988). This would give an estimate of the number of effectively independently estimated breeding values worth of information in a sample and would be a better reflection of the errors in the genetic quantities.

I believe this is already possible in the package, if I understood correctly. Both BootstrapRep and MonteCarloRep have sample size arguments.

In the absence of this, adding explicit ways of working with the posterior distributions or estimated matrix of errors from the estimate of the G matrix would be welcome in your package. At the very least, you should make the distinction between these cases clearer in the text.

As much as we would like to provide general methods for G matrix error estimation, the current available mixed model packages in R for ML estimation make this difficult. We recommend using MCMCglmm or some other Bayesian framework to incorporate uncertainty, and the posterior samples can be trivially used with the functions in the package.

I think the random skewers procedure should be revisited in one respect. The Marroig and Cheverud (2007) procedure describes drawing random selection gradient values from a uniform distribution and then standardizing the resulting vector to unit length. This is meant to provide random directions in the phenotypic space. This procedure, however, does not achieve this as the distribution of random points before standardization is from a hypercube of the same dimensionality as the number of traits. This will lead to an uneven distribution of points around the hypersphere on which the unit length vectors of standardized simulated selection gradients would occupy. In effect, the extra space in the corners of the hypersphere cause patches of excess vectors. This could cause considerable mischief if, for example, the first principal component of a covariance matrix was oriented directly at corner of the hypercube. It might be worth modifying the

> **procedure to draw random Gaussian variables, which are then standardized to unit length following Muller (1959) and Marsaglia (1972). This would ensure truly random simulated directional selection.**
>
> We generate random vectors using normal distributions for the elements to avoid this problem. Since the multivariate normal is spherically symmetrical, there are no privileged directions.
>
> ***Competing Interests:*** No competing interests were disclosed.

Referee Report 20 July 2016

doi:10.5256/f1000research.9735.r14607

? **David Houle**

Department of Biological Science, Florida State University, Tallahassee, FL, USA

This article has been improved in response to the previous comments of the reviewers. The authors have added several important and difficult-to-implement new analyses to the package, especially those discussed in Aguirre et al. 2014. The text has been altered to make clearer the limitations of some of the procedures.

On the other hand, many of my more general comments about the shortcomings of the package still hold, and I refer the reader to my first set of comments.

One problem has, however, gotten worse, and that is how the package discusses the estimation of errors in G matrices and other variance component matrices. The text and code does not properly distinguish between sampling error of estimates of G matrices and sampling error of simple statistics based directly on the data. Consequently at many points the package could be used to produce misleading estimates of error in evolutionary statistics.

The section on matrix error and repeatability is the locus of this problem. The opening paragraph in this section is about error in the G matrices, but none of the methods in that section actually apply to G matrices. In particular the text equates the REML-MVN method (Houle and Meyer, 2015) with this package's MonteCarloRep, but this is incorrect. REML-MVN requires you to have an estimate of the sampling variance-covariance of matrix elements, but the software implements the assumption that the only parameter that is needed to construct an estimate of error is the sample size. This works for a P matrix of multivariate normal data, but is NOT the REML-MVN method.

Similarly, the bootstrap routine seems only to bootstrap at the level of individuals. Bootstrapping over individuals will always severely underestimate the error in G. For a designed experiment like a full and half-sib estimation of G, one could get proper estimates of error in G by bootstrapping over sires, which are the exchangeable units in that design. A pedigree cannot be effectively bootstrapped as individuals are not exchangeable. For a G matrix, you need additional estimates of error from either the posterior distribution of matrices, the variance-covariance matrix of estimates from a ML procedure, or samples drawn from such a distribution. Consequently the statement at the start of the Evolutionary Statistics section that **"MonteCarloStat()** and **BootstrapStat()** can be used to generate confidence intervals for these statistics" is incorrect. To properly deal with this problem all of the routines should allow the user to furnish samples of matrices from the sampling distributions (e.g. samples from the posterior distribution),

and then produce measures of uncertainty from those.

It would be ideal for the package to implement the ability to utilize such samples in every procedure.

References

1. Houle D, Meyer K: Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood.*J Evol Biol*. 2015; **28** (8): 1542-9 PubMed Abstract | Publisher Full Text
2. Aguirre JD, Hine E, McGuigan K, Blows MW: Comparing G: multivariate analysis of genetic variation in multiple populations.*Heredity (Edinb)*. 2014; **112** (1): 21-9 PubMed Abstract | Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 31 Oct 2016

Diogo Melo, Universidade de São Paulo, Brazil

Thank you for the second careful review! We agree completely with the problems regarding G-matrix error and have modified the text to make these limitations clear. As much as we would like to provide general methods for G matrix error estimation, the current available mixed model packages in R for ML estimation make this difficult. We recommend using MCMCglmm or some other Bayesian framework to incorporate uncertainty.

Competing Interests: No competing interests were disclosed.

Version 1

Referee Report 16 November 2015

?

David Houle
Department of Biological Science, Florida State University, Tallahassee, FL, USA

This paper describes the functions implemented in the R package "evolvqg." This package implements a large set of methods for characterizing and comparing variance-covariance matrices. A compendium of such methods is useful, but the user should be aware of shortcomings in this package.

First, the authors justify grouping these methods together under the umbrella of "evolutionary quantitative genetics." Under this label, the very diversity of methods is a bit misleading - the majority of implemented methods have no explicit relationship to evolution, and this is not clear in the paper. To a large degree this confusion reflects the state of the field. A great many ad hoc methods are proposed and widely applied, such as the popular Mantel test. For example, this package implements methods for quantifying

repeatability of the phenotypic covariance and correlation matrices, which is a fine thing to do, but has little relevance to the genetic matrices that underlie evolution. For those matrices, any procedure for estimating them will partition out the non-genetic factors as part of the hierarchical model fitting. Repeatability is usually the least of our estimation problems. Other methods that are in this not-really-evolutionary category include, matrix correlations, the 'modularity' analyses.

A related issue is that this package does not include state-of-the-art techniques. In particular, the advent of MCMC methods (plus other methods for getting sampling variation of matrices: Houle and Meyer 2015) provides proper measures of uncertainty for the evolutionarily relevant G matrices. The previously available R package "evolvability" uses the posterior distributions for G matrices as well as selection gradients, when available, to place confidence intervals on the Hansen and Houle measures of evolvabilities. The "evolqg" package under review here primarily implements methods developed and used by the Marroig group, lacking some of the most promising approaches and metrics developed by others (e.g. Hine *et al*. 2009, Houle and Fierst 2013; Aguirre *et al*. 2014). Programming time is of course a limitation, and no one is under obligation to implement everything. However, it is deficient scholarship that this paper makes no mention of these approaches, or of the availability of software that does these analyses.

Based on their descriptions here, the phylogenetic comparison methods implemented in this package seem to be quite deficient. AncestralStates is contradictorily described as dealing with multivariate data, but reconstructs each character independently. It appears to implement an interface to a univariate method, and is not actually a multivariate approach. PhyloW and PhloCompare compute "weighted" estimates of matrices at internal nodes, but the proper weight to apply to a G matrix is not clear at all. This appears to only deal with sampling variance at the individual level, neglecting the more important sources of matrix variation. This disconnect appears to rest on the author's assumption that "As a general rule, high similarity between populations' P-matrices is a good indicator of high similarity between P and G". This is certainly not a general rule, as P and G matrices for traits with low heritability can be very different from each other. The classic examples are life history traits.

For drift models, this implementation shares a major deficiency with previous work, in assuming that the expectation is that matrices will remain proportional. This is indeed the large-sample expectation, but any actual population will deviate from that expectation in ways that depend on the unknown parameters of the underlying system – the number of loci, their relative mutability, the underlying M matrix, and of course the effective population size of each part of the genome. For example, Griswold *et al*. 2007 show that even when the underlying M matrix is spherical, realizations based on this will have substantial deviations from sphericity. There is no general treatment of this problem, so all results from such analyses need to be treated with caution. A failure to reject a departure from proportionality is meaningful, but rejections of proportionality do not necessarily indicate that drift is not responsible.

Finally, like much of the software being made freely available, there is no description of what the authors did to validate their implementation of these techniques, and no comparisons with previous analyses to indirectly validate them. One or both of these should really be standard with new software. I know that it is not standard, and do not want to single these authors out on that account. The user, however, should be aware of all unvalidated software, and should perform their own checks. Unfortunately, this is only easy for simple procedures for which a package is not really necessary.

Some very specific issues:
1. Errors: "The proportion of variance not associated with the individuals is called the repeatability."

2. The description of Mantel tests is misleading, as a high correlation does not mean that matrices are the "same," nor does a negative correlation mean that matrices are "opposite."

3. The description of the PCA similarity algorithm is opaque. What is "pondering"?

References

1. Griswold CK, Logsdon B, Gomulkiewicz R: Neutral evolution of multiple quantitative characters: a genealogical approach.*Genetics*. 2007; **176** (1): 455-66 PubMed Abstract | Publisher Full Text

2. Hine E, Chenoweth SF, Rundle HD, Blows MW: Characterizing the evolution of genetic variance using genetic covariance tensors.*Philos Trans R Soc Lond B Biol Sci*. 2009; **364** (1523): 1567-78 PubMed Abstract | Publisher Full Text

3. Houle D, Meyer K: Estimating sampling error of evolutionary statistics based on genetic covariance matrices using maximum likelihood.*J Evol Biol*. 2015; **28** (8): 1542-9 PubMed Abstract | Publisher Full Text

4. Houle D, Fierst J: Properties of spontaneous mutational variance and covariance for wing size and shape in Drosophila melanogaster.*Evolution*. 2013; **67** (4): 1116-30 PubMed Abstract | Publisher Full Text

5. Aguirre JD, Hine E, McGuigan K, Blows MW: Comparing G: multivariate analysis of genetic variation in multiple populations.*Heredity (Edinb)*. 2014; **112** (1): 21-9 PubMed Abstract | Publisher Full Text

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 13 Jun 2016

Diogo Melo, Universidade de São Paulo, Brazil

- This paper describes the functions implemented in the R package "evolqg." This package implements a large set of methods for characterizing and comparing variance-covariance matrices. A compendium of such methods is useful, but the user should be aware of shortcomings in this package.

Thank you for your insightful comments. One of the reasons we chose F1000research for this article is our perception that something is amiss in the current system of peer-review, where good reviewers are burdened and swamped by manuscripts and authors tend to often ignore their criticisms and take the short road to the next journal inline instead of dealing with the criticisms and polish their manuscripts. The fully open system implemented by F1000 seems like a good way out of this trap we express our gratitude to Drs. Houle and Grabowsky for taking their time to criticize this package.

We respond to comments individually after each of your considerations.

- First, the authors justify grouping these methods together under the umbrella of "evolutionary quantitative genetics." Under this label, the very diversity of methods is a bit misleading - the majority of implemented methods have no explicit relationship to evolution, and this is not clear in the paper. To a large degree this confusion reflects the

state of the field. A great many ad hoc methods are proposed and widely applied, such as the popular Mantel test. For example, this package implements methods for quantifying repeatability of the phenotypic covariance and correlation matrices, which is a fine thing to do, but has little relevance to the genetic matrices that underlie evolution. For those matrices, any procedure for estimating them will partition out the non-genetic factors as part of the hierarchical model fitting. Repeatability is usually the least of our estimation problems. Other methods that are in this not-really-evolutionary category include, matrix correlations, the 'modularity' analyses.

This is a fair comment, and indeed an argument could be made to separate the different aspects of evolutionary research that are presented here into different packages. We chose to keep these different methods in a single source since we feel a consistent workflow is very beneficial to research on evolution and covariation. For example, many of the latter hypothesis-testing methods presented in our paper should only be used if some level of matrix similarity is detected, and the matrix similarities, in turn, should not be interpreted without matrices' repeatabilities. Furthermore, while the development of multivariate evolutionary theory and the theory of integration and modularity were separate, Lande's, Cheverud's and Wagner's work since the 80s have linked these fields very intimately, to a point that it is hard for us to think of these fields as separate. From our point of view, the influence of covariation in evolution, and the genetic and developmental origin of these variational associations make integration and modularity central to modern evolutionary theory, even though admittedly there are a number of researchers who consider morphological integration and comparative quantitative genetics as two separate fields. We feel this integrative approach links the developmental and intra-populational causes of genetic covariation to their evolutionary consequences, leading to a more complete and robust understanding of micro- and macro-evolution.

Finally, the inclusion of methods that are mostly "statistical" and not "evolutionary" should make it easier for researchers to check the quality of their data and its appropriateness for further analyses.

- A related issue is that this package does not include state-of-the-art techniques. In particular, the advent of MCMC methods (plus other methods for getting sampling variation of matrices: Houle and Meyer 2015) provides proper measures of uncertainty for the evolutionarily relevant G matrices. The previously available R package "evolvability" uses the posterior distributions for G matrices as well as selection gradients, when available, to place confidence intervals on the Hansen and Houle measures of evolvabilities. The "evolqg" package under review here primarily implements methods developed and used by the Marroig group, lacking some of the most promising approaches and metrics developed by others (e.g. Hine et al. 2009, Houle and Fierst 2013; Aguirre et al. 2014). Programming time is of course a limitation, and no one is under obligation to implement everything. However, it is deficient scholarship that this paper makes no mention of these approaches, or of the availability of software that does these analyses.

While we challenge the reviewer's assertion that we fail to mention these methods (see the second paragraph of our Summary section; and the method described in Houle and Meyer is implemented by the function MonteCarloStat()), the point is well taken. Another reason for choosing F1000Research was the ease of updating the manuscript as we add new functionality to the EvolQG package. In this spirit of continuous development, we added in the revised version of the package three of the methods described in Aguirre et al. 2014, including the eigentensor decomposition described by Hine et al. 2009, using fast and flexible implementations in R.

- Based on their descriptions here, the phylogenetic comparison methods implemented in this package seem to be quite deficient.

We would argue that the available methods in the literature are deficient in dealing with multivariate correlated traits, and we provide the rather simple available methods.

- AncestralStates is contradictorily described as dealing with multivariate data, but reconstructs each character independently. It appears to implement an interface to a univariate method, and is not actually a multivariate approach.

We clearly stated that AncestralStates is not a multivariate approach. Indeed AncestralStates is just a wrapper to facilitate reconstructing multiple traits independently. If the reviewer has a suggestion on how to implement this taking the multivariate covariance structure into account for many traits we would be very interested, but this remains an active research topic, and we have not found a satisfactory solution for this problem. Since this seems to be more misleading than helpful, we have removed this function from the package and manuscript.

- PhyloW and PhyloCompare compute "weighted" estimates of matrices at internal nodes, but the proper weight to apply to a G matrix is not clear at all. This appears to only deal with sampling variance at the individual level, neglecting the more important sources of matrix variation.

This is another unambitious method, and was intended only to calculate within-group phenotypic covariance matrices in a phylogenetically structured way. The use with G-matrices coming from more complex linear models would indeed be non-trivial, and we now make this explicit in the description.

- This disconnect appears to rest on the author's assumption that "As a general rule, high similarity between populations' P-matrices is a good indicator of high similarity between P and G". This is certainly not a general rule, as P and G matrices for traits with low heritability can be very different from each other. The classic examples are life history traits.

We would argue that in the case of low heritability the P-matrices between populations would also be dissimilar. The point here is not that P and G are always similar, but that similar Ps between populations are a fair indication of similar Gs and Ps, at least in the groups we have worked with. In mammals, this conclusion is supported empirically by the comparisons of 5 different G-matrices with P-matrices of several groups, which indicate similar responses to random selection on average (Porto, 2009). In any event, we add a caveat on the function description that structurally similar matrices are a key component of the methods implemented here.

- For drift models, this implementation shares a major deficiency with previous work, in assuming that the expectation is that matrices will remain proportional. This is indeed the large-sample expectation, but any actual population will deviate from that expectation in ways that depend on the unknown parameters of the underlying system – the number of loci, their relative mutability, the underlying M matrix, and of course the effective population size of each part of the genome. For example, Griswold et al. 2007 show that even when the underlying M matrix is spherical, realizations based on this will have

substantial deviations from sphericity. There is no general treatment of this problem, so all results from such analyses need to be treated with caution. A failure to reject a departure from proportionality is meaningful, but rejections of proportionality do not necessarily indicate that drift is not responsible.

This is indeed a problem, and we try to remedy this by ensuring the matrices share some minimum level of similarity before using these drift models. Proa et al. 2013 analysed the type I error rate in the DriftTest method, and found that if matrices are similar the test is well behaved. This similarity must be tested on a case by case basis. Error rate analysis of the other tests is an open problem. Another option is to repeat the analysis using different matrices from the terminal taxa that represent extremes of variability, and check if the results are robust to this. We make these caveats and problems clear in the revised manuscript.

This paragraph now reads:

"Since both these tests use drift as a null hypothesis, failure to reject the null hypothesis is not evidence that selection was not involved in the observed pattern of diversification, only that the observed pattern is compatible with drift. Also, these methods assume that the matrices involved share some degree of similarity, and should ideally be proportional to each other. We would be very weary of using these methods if the matrices are too dissimilar, or if the results change radically if different matrices are used as the ancestral matrix. Also, these tests rely on two levels of replication, taxa and traits. As a general guideline, at least 20 traits and at least 8 taxa should be sampled for using these methods with any confidence, and results should be analyzed in conjunction with other lines of evidence."

With regards to Griswold et al. 2007, we believe that verifying the extant matrices are similar somewhat sidesteps these problems, and in any event their simulations do not include stabilizing selection on covariance patterns, which is very likely to exist if matrices are stable in evolutionary timescales. It's important to realize that the methods we describe are for identifying drift on species means, not covariances. Evolution of covariance patterns is a different matter altogether.

- Finally, like much of the software being made freely available, there is no description of what the authors did to validate their implementation of these techniques, and no comparisons with previous analyses to indirectly validate them. One or both of these should really be standard with new software. I know that it is not standard, and do not want to single these authors out on that account. The user, however, should be aware of all unvalidated software, and should perform their own checks. Unfortunately, this is only easy for simple procedures for which a package is not really necessary.

We agree entirely, and feel that the bar for scientific software should be high, and so we took additional steps in this direction. While no implementation is bug free, we compared all results from our initial set of functions between different implementations done by members of our lab and available implementations in the literature. Also, all development for the package was done in a test driven development framework, and all functions have unit tests for the most or all of their functionality, that is run every time the package is built. This insures modifications do not alter previous results. The implementations in the package follow a modular design for most functionality, minimizing code duplication and reducing the chance of bugs. We also have a fast and constantly maintained issue and bug tracker in github, where users can ask questions, request new functionality, and report bugs.

- Some very specific issues:

- Errors: "The proportion of variance not associated with the individuals is called the repeatability."

Changed to something clearer. Now reads: "The proportion of variance associated with among individual variation, and not within individual variation, is called the repeatability"

- The description of Mantel tests is misleading, as a high correlation does not mean that matrices are the "same," nor does a negative correlation mean that matrices are "opposite."

We chose a more cautious wording of these general guidelines. Now reads: "The correlation between matrices range between -1 and 1, and higher correlations indicate matrices have more similar structures, while null correlations indicate the matrices have very distinct correlation structures. Correlations near zero can also occur if the elements of the matrices have nonlinear relationships between them, as in all Pearson correlations. Negative correlations indicate the pattern of association between traits is reversed in the two matrices."

- The description of the PCA similarity algorithm is opaque. What is "pondering"?

Sorry, this was a rather hard false cognate with portuguese for us to catch, and now reads: "In order to take the variation into account, we can add the eigenvalue associated with each principal component into the calculation, effectively weighting each correlation by the variance in the associated directions"

Literature cited
1. Cheverud, J. M. (1996). Quantitative genetic analysis of cranial morphology in the cottontop (Saguinus oedipus) and saddleback (S. fuscicollis) tamarins. Journal of Evolutionary Biology, 9(1), 5-42.
2. Marroig, G., & Cheverud, J. M. (2001). A comparison of phenotypic variation and covariation patterns and the role of phylogeny, ecology, and ontogeny during cranial evolution of New World monkeys. Evolution, 55(12), 2576-2600.
3. Prôa, M., O'Higgins, P., & Monteiro, L. R. (2013). Type I error rates for testing genetic drift with phenotypic covariance matrices: a simulation study. Evolution, 67(1), 185-195.
4. Porto, A., de Oliveira, F. B., Shirai, L. T., De Conto, V., Marroig, G., (2009). The Evolution of Modularity in the Mammalian Skull I: Morphological Integration Patterns and Magnitudes. Evolutionary Biology, 36(1), 118–135. doi:10.1007/s11692-008-9038-3

Competing Interests: No competing interests were disclosed.No competing interests were disclosed.

Referee Report 16 October 2015

doi:10.5256/f1000research.7623.r10607

Mark Grabowsky

Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, The Faculty of Mathematics and Natural Sciences, University of Olso, Oslo, Norway

This Software Tool is an open source R package for performing a wide variety of evolutionary genetic analyses. It comes at a particularly useful time when large data sets required to use these analytical techniques are more and more available. I predict this package will be used to a greater and greater extent over time.

The team, with lead author Diogo Melo, has made a number of major innovations in the field of evolutionary quantitative genetics and this software tool is well within their purview.

The group describes the potential application of the new software well, cites the sources for the development of the different methods included, and briefly explains each function with an adequate level of discussion. I found the section on Phylogenetic Comparisons most interesting, and would have liked more discussion of these points.

Issues:
1. It would be nice to include data for which the tools could be tried out on and examples of the output included in the text.

2. I also worry about the power of some of these tests, and would appreciate if the authors were to think about adding in caveats where available. There are a few papers out currently that address these issues (e.g. Haber, 2011). For example, I have both heard and performed some analyses using another version of the DriftTest() function and it appears about 20 traits are required to ever reject the null hypothesis. As this package is sure to be used by researchers who are less acquainted with these issues they may be unaware of issues with sample sizes, etc.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 13 Jun 2016
Diogo Melo, Universidade de São Paulo, Brazil

Thank you for your kind comments. We have included some additional caveats in relation to the power and level of replication required for some of the tests. With regards to examples, we fell a package vignette tutorial is much more suited to this, and we are working on more documentation to be distributed with the package.

Competing Interests: No competing interests were disclosed.