SOFTWARE TOOL ARTICLE

EvolQG - An R package for evolutionary quantitative genetics

[version 1; peer review: 1 approved, 1 approved with reservations]

Diogo Melo¹, Guilherme Garcia¹, Alex Hubbe², Ana Paula Assis¹, Gabriel Marroig¹

¹Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

²Departamento de Oceanografia, Instituto de Geociências, Universidade Federal da Bahia, Salvador, Brazil

V1 First published: 30 Sep 2015, 4:925
<https://doi.org/10.12688/f1000research.7082.1>
 Second version: 27 Jun 2016, 4:925
<https://doi.org/10.12688/f1000research.7082.2>
 Latest published: 08 Nov 2016, 4:925
<https://doi.org/10.12688/f1000research.7082.3>

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. Given this mathematical representation of available variation, the EvolQG package provides functions for calculation of relevant evolutionary statistics, estimation of sampling error, corrections for this error, matrix comparison via correlations and distances, and functions for testing evolutionary hypotheses on taxa diversification.

Keywords

P-matrix , G-matrix , multivariate evolution , drift , morphological evolution , directional selection , matrix comparison , covariance matrix

This article is included in the **RPackage** gateway.

Open Peer Review

Approval Status **✓✓?**

	1	2	3
version 3			
(revision)		✓	
08 Nov 2016		view	
version 2			
(revision)		?	?
27 Jun 2016		view	view
version 1			
30 Sep 2015	✓	?	
	view	view	

1. **Mark Grabowsky**, University of Oslo, Oslo, Norway

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

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

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the **Phylogenetics** collection.

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

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by funding from FAPESP. D.M. was funded by grants 2012/20180-0 and 2014/01694- 9; G.G. was funded by grant 2011/52469-4; A.H. was funded by grants 2008/56886-6 and 2012/24937-9; A.P.A was funded by grant 2010/52369-0; G.M. was funded by grant 2011/14295-7.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2015 Melo D *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Melo D, Garcia G, Hubbe A *et al.* **EvolQG - An R package for evolutionary quantitative genetics [version 1; peer review: 1 approved, 1 approved with reservations]** F1000Research 2015, 4:925 <https://doi.org/10.12688/f1000research.7082.1>

First published: 30 Sep 2015, 4:925 <https://doi.org/10.12688/f1000research.7082.1>

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 variation²⁵. 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^{22,46}.

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, $\Delta z = G\beta$, 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., 2,28). Thus, describing and understanding changes in standing variation among populations^{5,24,27} as well as understanding constraints imposed by populations standing variation (e.g., 15,26,42,49) are major elements in evolutionary quantitative genetics.

In this article we describe the **EvoIQG** package, developed to deal with the evolutionary quantitative genetics questions addressed above in the R environment for statistical computing⁴⁴. 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 analyses⁹. 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 not associated with the individuals is called the repeatability²³. A repeatability of 1 means that no variation is associated with measurement error. The function **CalcRepeatability()** performs the calculation described in 23 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 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 analysis²⁷. The **EvoIQG** package 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. Of course more complicated methods may be used to obtain **G**-matrices, such as an animal model or a mixed model^{25,48}.

Accurate **G**-matrix estimation can be hard to achieve, requiring large sample sizes, many families and known genealogies⁵⁰. One alternative that is usually more feasible is to use the phenotypic covariance matrix (the **P**-matrix) as a proxy of the population's **G**-matrix^{6,47}. 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., 11). 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^{16,30,33}, 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 identical⁶. 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. A number of methods for repeatability estimation are provided, and their results can be passed on to the functions that calculate matrix correlations (section **Matrix Comparison**):

AlphaRep(): Cheverud⁶ 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, since 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 (or correlation) 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 repeatability²⁷. This method has the advantage of being easy to apply to matrices coming from linear models with many controlled effects, and not requiring the original data.

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. This 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. 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**).

ExtendMatrix(): Marroig *et al.*³⁰ 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 equation²². 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 causes minimal changes in the distribution of phenotypes, but improves dramatically the estimates of net selection gradients. See 30 for a thorough examination of the performance and consequences of the extension method on simulated and real data sets.

Evolutionary statistics

Hansen and Houle¹⁵ provide a suite of statistics that have fairly good biological interpretations for a given **G**- or **P**-matrix. Marroig *et al.*³¹ 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.

In the following, $E[\cdot]_{\beta}$ represents the expected value over many random β vectors with unit norm, $\langle \cdot, \cdot \rangle$ represents the dot product between two vectors, $\cos(\cdot, \cdot)$ is the cosine between two vectors, **G** is an arbitrary covariance matrix, **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 40).
- 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 very 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 λ_1^G is the leading eigenvalue of **G**, we calculate this percentage as:

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

- Evolvability (Figure 1): The mean projection of the response to random selection gradients with unit norm onto the selection gradient. This is a measure of a population's available variation in the direction of a particular selection gradient, averaged across all directions¹⁵.

$$\bar{e} = E[\langle \mathbf{G}\beta, \beta \rangle]_{\beta}$$

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

responses. Flexibility measures on average how the response to selection aligns with the selection gradient³¹.

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

- **Responsability (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^{15,31}.

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

- **Conditional Evolvability:** Measures the mean response to selection in the direction of a given β when other directions are under stabilizing selection¹⁵.

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

- **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 ($\bar{e}$)¹⁵.

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

- **Constraints:** The mean correlation between the response vector to random selection gradients and the matrix's first principal component³¹. If Λ_1^G is the first principal component of $\mathbf{G}$, constraints are measured as:

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

Figure 1. Graphical representation of evolvability ($\bar{e}$), responsibility ($\bar{r}$) 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 .

Matrix comparison

A $\mathbf{G}$ -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. Since 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 macroevolution, since this requires some additional assumptions, such as a relative stability of the $\mathbf{G}$ -matrix over generations. Comparing extant covariance matrices is a test of this required stability (e.g. 27). For a thoughtful discussion on the biological relevance of statistical significance in matrix comparisons, see the discussion in 13.

Matrix correlations

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

RandomSkewers(): The Random Skewers (RS) method makes use of the Lande equation²², $\Delta z = \mathbf{G}\beta$, where Δz represents the vector of response to selection, $\mathbf{G}$ the $\mathbf{G}$ -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:

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

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. This 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. The RS values range between -1 (the matrices have the opposite structure) and 1 (the matrices share the same structure), and zero means the matrices have distinct structures.

MantelCor(): Correlation matrices can be compared 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. The observed correlation is significant when it is larger than the 95% quantile of the permuted distribution. This method can not 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 (the matrices have the opposite structure) and 1 (the matrices share the same structure), and zero means the matrices have distinct structures.

KrzCor(): The Krzanowski shared space, or Krzanowski correlation, measures the degree to which the first principal components (eigenvectors) span the same subspace^{3,20}, 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):

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

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 space³². In order to take the variation into account, we can add the eigenvalue associated with each principal component into the calculation, effectively pondering each correlation by the variance in the associated directions. If Λ_i^A is the i th eigenvalue of matrix A , we have:

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

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).

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 populations²⁹. 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:

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

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

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

Each of these row vectors $r_i^{AB} = (A_{ij}\beta_j)_{j=1 \dots n}$ 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:

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

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

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

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 29 for details and examples.

Matrix distances

Another approach to estimate the similarity or dissimilarity between matrices is to calculate the distance between a pair of 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.*³⁹ 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:

$$q(f, g) = \int_{\mathbb{R}^n} \frac{g(x)}{f(x) + g(x)} f(x) dx \quad (8)$$

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:

$$d(f, g) = \sqrt{1 - 2q(f, g)} \quad (9)$$

Since $q(f, g)$ is symmetrical, $d(f, g)$ is also symmetrical, and the square root guarantees that $d(f, g)$ satisfies the triangle inequality³⁹. Calculation is straight forward and can be done with a simple sampling Monte Carlo scheme, see 39 for details.

- Riemann distance: Mitteroecker and Bookstein³⁴ use a Riemannian metric in the space of positive definite matrices (either covariance or correlation matrices), based on exponential mapping³⁵ 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:

$$\|A, B\|_{cov} = \|B, A\|_{cov} = \sqrt{\sum_{i=1}^p [\log(\lambda_i)]^2} \quad (10)$$

This distance has the advantage of being invariable under changes in the base used to represent the matrices. See 34 for a discussion on the biological relevance of this distance.

Phylogenetic comparisons

AncestralStates(): Ancestral state reconstruction of continuous traits using maximum likelihood can be performed in R (for example, using ape or phytools), but most packages don't provide a simple interface for multivariate data. **AncestralStates()** is a wrapper for **fastAnc()** in phytools⁴⁵ that works on multivariate data, reconstructing each character independently.

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 by taking a tree and a set of measurements (covariance matrices) and returns means 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.

While using the within-group covariance matrix is a reasonable alternative as the estimator of an ancestral covariance matrix, this 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 phylogenetically 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 modules⁵¹. In the context of morphological traits, variational modularity is associated with the concept of integration³⁸, that is, the tendency of morphological systems to exhibit correlations due to common developmental factors and functional demands^{7,14}.

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 another³¹. At the same time, selection can alter existing patterns of integration and modularity, leading to traits becoming more or less correlated^{18,32}. The pattern of correlation 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 by creating a modularity hypothesis matrix and comparing it via Mantel correlation with 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., 41). **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 19 for other alternatives). In these cases, the Modularity Hypothesis Index is useful and allows comparing results between raw and residual matrices⁴².

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 index³⁶. If A is a correlation matrix we define L modularity as:

$$L = \sum_{i \neq j} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(g_i, g_j) \quad (11)$$

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:

$$\delta(x, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases} \quad (12)$$

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 :

$$k_i = \sum_{j \neq i} A_{ij} \quad (13)$$

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 symmetrical³⁶. 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 a Markov Chain Monte Carlo annealing method 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* package⁸ provides a number of similar community detection algorithms that can also be used in correlation matrices.

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 26), it is useful to look at the structure of modularity after removing this dominant integrating factor. This is done using the method described in 4. Porto *et al.*⁴² 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 described in 1, or the correlation of principal component scores²⁸. 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.

DriftTest(): Under drift, 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. Conditions for the validity of these assumptions are reviewed in 43. 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:

$$B \propto (t / N_e) W \quad (14)$$

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 :

$$\log(B_i) = \log(t / N_e) + \beta \log(\lambda_i^W) \quad (15)$$

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^{17,21,22}. 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, we expect the mean scores of different populations in the principal components of the within-group covariance matrix to be uncorrelated²⁸. Significant correlations between the scores of the means on any two principal components is an indication of correlated directional selection¹⁰.

Random matrices

RandomMatrix(): Generating realistic random covariance matrices for null hypothesis testing is a challenging task, since 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 12).

The method described in 37 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. We also lack Bayesian implementations that would be possible for some functions (i.e. 3). Our reasons for this are twofold: the difficulty in transposing these methods efficiently to multiple traits, especially with respect to Bayesian implementations of existing functions, 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: <https://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.3091053>
4. License: The MIT License (<https://opensource.org/licenses/MIT>)

Author contributions

D.M. compiled existing implementations, re-factored existing functions and contributed new code, documentation and unit tests. G.G. provided the initial set of functions and first implementations. A.H. tested and revised code and documentation. A.P.A. contributed code and documentation, G.M. developed methods and devised necessary elements for package. All authors wrote the paper.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by funding from FAPESP. D.M. was funded by grants 2012/20180-0 and 2014/01694-9; G.G. was funded by grants 2008/56886-6 and 2011/52469-4; A.H. was funded by grant 2012/24937-9; A.P.A. was funded by grant 2010/52369-0; G.M. was funded by grant 2011/14295-7.

I confirm that the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We would like to thank Barbara Costa, Daniela Rossoni, Edgar Zanella and Fabio Machado for contributing code and revising documentation and results.

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. 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](#)
5. Cheverud JM: **Phenotypic, Genetic, and Environmental Morphological Integration in the Cranium.** *Evolution.* 1982; **36**(3): 499–516.
[Publisher Full Text](#)
6. Cheverud JM: **A Comparison of Genetic and Phenotypic Correlations.** *Evolution.* 1988; **42**(5): 958–968.
[Publisher Full Text](#)
7. Cheverud JM: **Developmental Integration and the Evolution of Pleiotropy.** *Integr Comp Biol.* 1996; **36**(1): 44–50.
[Publisher Full Text](#)
8. Csardi G, Nepusz T: **The igraph software package for complex network research.** *Inter Journal. Complex Systems:* 1695, 2006.
[Reference Source](#)
9. Falconer DS, Mackay TF: **Introduction to Quantitative Genetics.** Benjamin Cummings, London, 4th edition, 1996.
[Reference Source](#)
10. Felsenstein J: **Phylogenies And Quantitative Characters.** *Annu Rev Ecol Syst.* 1988; **19**(1): 445–471.
[Publisher Full Text](#)
11. 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](#)
12. Haber A: **A Comparative Analysis of Integration Indices.** *Evol Biol.* 2011; **38**(4): 476–488.
[Publisher Full Text](#)
13. Haber A: **The Evolution of Morphological Integration in the Ruminant Skull.** *Evol Biol.* 2015; **42**(1): 99–114.
[Publisher Full Text](#)
14. 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](#)
15. 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](#)
16. Hill WG, Thompson R: **Probabilities of Non-Positive Definite between-Group or Genetic Covariance Matrices.** *Biometrics.* 1978; **34**(3): 429–439.
[Publisher Full Text](#)
17. 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](#)
18. 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](#)
19. Jungers WL, Falsetti AB, Wall CE: **Shape, relative size, and size-adjustments in morphometrics.** *Am J Phys Anthropol.* 1995; **38**(S2): 137–161.
[Publisher Full Text](#)
20. Krzanowski WJ: **Between-Groups Comparison of Principal Components.** *J Am Stat Assoc.* 1979; **74**(367): 703–707.
[Publisher Full Text](#)
21. Lande R: **Natural Selection and Random Genetic Drift in Phenotypic Evolution.** *Evolution.* 1976; **30**(2): 314–334.
[Publisher Full Text](#)
22. Lande R: **Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry.** *Evolution.* 1979; **33**(1): 402–416.
[Publisher Full Text](#)
23. Lessells CM, Boag PT: **Unrepeatable repeatabilities: a common mistake.** *Auk.* 1987; **104**(1): 116–121.
[Publisher Full Text](#)
24. 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](#)
25. Lynch M, Walsh B: **Genetics and analysis of quantitative traits.** Sinauer, Sunderland, MA., 1998.
[Reference Source](#)
26. 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](#)

27. 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](#)
28. 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](#)
29. 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](#)
30. Marroig G, Melo DA, Garcia G: **Modularity, noise, and natural selection.** *Evolution*. 2012; **66**(5): 1506–1524.
[PubMed Abstract](#) | [Publisher Full Text](#)
31. Marroig G, Shirai LT, Porto A, *et al.*: **The Evolution of Modularity in the Mammalian Skull II: Evolutionary Consequences.** *Evol Biol*. 2009; **36**(1): 136–148.
[Publisher Full Text](#)
32. 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](#)
33. 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](#)
34. 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](#)
35. Moakher M: **On the Averaging of Symmetric Positive-Definite Tensors.** *J Elast*. 2006; **82**(3): 273–296.
[Publisher Full Text](#)
36. 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](#)
37. 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](#)
38. Olson R, Miller E: **Morphological integration.** University of Chicago Press, Chicago, 1958.
[Reference Source](#)
39. 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](#)
40. Pavlicev M, Cheverud JM, Wagner GP: **Measuring Morphological Integration Using Eigenvalue Variance.** *Evol Biol*. 2009; **36**(1): 157–170.
[Publisher Full Text](#)
41. 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](#)
42. 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](#)
43. 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](#)
44. R Core Team. **R: A Language and Environment for Statistical Computing.** R Foundation for Statistical Computing, Vienna, Austria, 2014.
[Reference Source](#)
45. Revell LJ: **phytools: An R package for phylogenetic comparative biology (and other things).** *Methods Ecol Evol*. 2012; **3**(2): 217–223.
[Publisher Full Text](#)
46. 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](#)
47. 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](#)
48. 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](#)
49. Schluter D: **Adaptive Radiation Along Genetic Lines of Least Resistance.** *Evolution*. 1996; **50**(5): 1766–1774.
[Publisher Full Text](#)
50. 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](#)
51. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Rev Genet*. 2007; **8**(12): 921–931.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. 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](#)
53. Melo D, Alvarenga EZ, Hubbe A, *et al.*: **EvoIQG: F1000Research.** *Zenodo*. 2015.
[Data Source](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 16 November 2015

<https://doi.org/10.5256/f1000research.7623.r10610>

© 2015 Houle D. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

David Houle

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

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.

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 “evolability” 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](#)

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and 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.

Author Response 13 Jun 2016

Diogo Melo, Universidade de São Paulo, São Paulo, Brazil

- This paper describes the functions implemented in the R package "evolgg." 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 online 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 "evolability" 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 cotton-top (*Saguinus oedipus*) and saddle-back (*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.

Reviewer Report 16 October 2015

<https://doi.org/10.5256/f1000research.7623.r10607>

© 2015 Grabowsky M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mark Grabowsky

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

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard.

Author Response 13 Jun 2016

Diogo Melo, Universidade de São Paulo, 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 felt 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.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research