doi:10.1111/j.1558-5646.2007.00216.x

TESTING THE GENETIC CONSTRAINT HYPOTHESIS IN A PHYLOGENETIC CONTEXT: A SIMULATION STUDY

Liam J. Revell

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

E-mail: lrevell@fas.harvard.edu

Received April 9, 2007

Accepted June 2, 2007

Quantitative genetic theory predicts that when populations diverge by drift the interspecific divergence (D matrix), calculated from species means, will be proportional to the average value of the additive genetic variance–covariance matrix, or G matrix. Most empirical studies in which this hypothesis has been investigated have ignored phylogenetic nonindependence among included taxa. Baker and Wilkinson (2003; also Revell et al. 2007) used a test for constraint in which the D matrix is calculated from phylogenetically independent contrasts (Felsenstein 1985) instead of directly from the species means. I use computer simulations to show that, on average, when the process of evolution is genetic drift, the divergence matrix calculated from independent contrasts (D_{IC}) is more highly correlated with G than is the divergence matrix calculated ignoring phylogenetic nonindependence (D). This effect is more pronounced when speciation is initially slow but increases over time than when speciation decreases over time. Finally, when evolution is primarily by drift but phenotype space is bounded (as if by functional constraint) the average correlation is decreased between both G and D or D_{IC}, however the correlation between G and D_{IC} is much larger than between G and D. Although limited in scope, to my knowledge this is the first study to use individual-based quantitative genetic simulations in a phylogenetic context.

KEY WORDS: Comparative method, evolutionary constraint, genetic constraint, genetic variance–covariance matrix, independent contrasts, quantitative genetics.

A central issue in many evolutionary studies involves the importance played by genetic constraint on phenotypic evolution in diverging populations (Schluter 1996, 2000; Bégin and Roff 2003, 2004). One approach used to evaluate the role of constraint is to compare the alignment of **G**, the matrix of additive genetic variances and covariances, with **D**, a divergence matrix composed of the variances and covariances calculated from population or species means for traits (Lande 1979; e.g., Bégin and Roff 2004).

This test is based on Lande's (1979) prediction that, under genetic drift and constraint, **D** would change as a function of elapsed time and **G**, such that:

$$\mathbf{D} = (t/N_e)\mathbf{G}, \quad (1)$$

where t and N_e are time (in generations) and the effective population size, respectively. Either **G** is assumed to be constant or its time average should be used in this calculation. As equation (1) predicts that **D** will be proportional to **G**, numerous empirical tests of the constraint hypothesis have evaluated the alignment of **G** and **D** and suggested that their proportionality, when found, indicated an important role for genetic drift and constraint in the phenotypic diversification of the species in the study (e.g., Ackermann and Cheverud 2002; Baker and Wilkinson 2003; Bégin and Roff 2003, 2004; Marroig et al. 2004). By extension, uncorrelated or poorly correlated **G** and **D** are sometimes interpreted as indicative of natural selection (e.g., Merilä and Björklund 1999). An implicit assumption of this test is that the taxa have radiated simultaneously from a common ancestor (Bégin and Roff 2004; Revell et al. 2007).

In a recent empirical study, Revell et al. (2007) suggested a phylogenetic test of the constraint hypothesis for the more

common situation in which taxa are related according to some estimable bifurcating history. In this test, a mean squares and mean cross-products (MS-MCP) matrix of independent contrasts (Felsenstein 1985; $\mathbf{D_{IC}}$ in Revell et al. 2007) is substituted for **D**. Revell et al. (2007) are not the first to use independent contrasts to test a hypothesis of quantitative genetic constraint. Baker and Wilkinson (2003) used a similar approach in the analysis of genetic constraint on phenotypic evolution in stalk-eyed flies. They used a correlation matrix of independent contrasts. A phylogenetic approach is also suggested by Bégin and Roff (2004).

Here, I utilize computer simulations to explore the consequences of phylogenetic nonindependence on tests of the genetic constraint hypothesis for population differentiation, in other words, to explore the method of Baker and Wilkinson (2003) adapted in Revell et al. (2007). The simulations performed in this study are individual based and genetically explicit. To my knowledge, this is the first study in which individual-based quantitative genetic simulations are performed in a phylogenetic context, as well as the first in which the genetic architecture of more than two quantitative traits is simulated (but see Revell 2007, Appendix C, for an example of the latter).

In this study I focus in particular on the MS-MCP matrix of independent contrasts ($\mathbf{D_{IC}}$) for testing the genetic constraint hypothesis. I evaluate the performance of $\mathbf{D_{IC}}$ by comparing the alignment of **G** and $\mathbf{D_{IC}}$ to the alignment of **G** and **D** because proportionality of **G** and the divergence matrix (**D** or $\mathbf{D_{IC}}$) is expected under genetic drift and constraint—the evolutionary conditions simulated in this study. I feel that this focus is appropriate as nearly every study involving a test for genetic constraint ignores phylogenetic nonindependence (reviewed in Bégin and Roff 2004; but see Baker and Wilkinson 2003; Revell et al. 2007 for exceptions) and the consequence of ignoring this nonindependence has not been the subject of any prior studies (but see Revell et al. 2007 for an empirical example).

I also investigate conditions under which the performances of phylogenetic and nonphylogenetic tests for constraint seemed likely to differ most and least severely. In particular, I consider the circumstances in which speciation is initially rapid, but slows over time, and vice versa, as well as the circumstance in which evolution is primarily by drift, but in which the morphospace has bounded limits (as when phenotypic evolution is restrained by functional considerations).

These circumstances were chosen for several reasons. In the former case, in which speciation rate is varied, initially rapid speciation tends to produce a star-like phylogenetic tree. A star phylogeny, in which all taxa radiate simultaneously from a common ancestor, is an implicit assumption of using equation (1) as the basis for an empirical test of genetic constraint. Consequently, ignoring phylogenetic nonindependence may be less significant if speciation is initially rapid and decreases over time. Conversely, if speciation is initially slow but accelerates over time, stochasticism in the evolutionary process under genetic drift on the few early branches in the tree may heavily influence the occupation of morphospace by the group and consequently ignoring phylogenetic history could be of greater significance than if speciation were constant over time (Revell et al. 2007).

In the latter case, in which morphospace is bounded as if by functional constraint, the assumption of pure drift is violated. In particular, bounds on morphospace will tend to erode "phylogenetic signal" (the correlation between patristic distance in the phylogeny and phenotypic dissimilarity) over time. Lack of phylogenetic signal is sometimes provided as a rationale for ignoring phylogenetic history (e.g., Bégin and Roff 2004). However, under conditions of bounded phenotype space, most evolution occurs by genetic drift so long as the population mean is not near the morphospace bounds, and many independent contrasts will contain evidence of drift that might be absent if phylogeny is ignored. Consequently, ignoring phylogenetic history may cause a failure to detect the importance of genetic constraint under conditions in which significant functional constraints exist.

To investigate these circumstances (1) I compare phylogenetic and nonphylogenetic tests for constraint; (2) I compare tests for constraint under conditions in which the speciation rate is constant, linearly decreases, or linearly increases with time; and finally (3) I compare tests for constraint when evolution occurs primarily by drift, but when morphospace is bounded, as if by functional constraint.

Methods

PHYLOGENETIC SIMULATION MODEL

I simulated 300 stochastic phylogenies using a discrete time pure birth model. In the first 100 simulations, the birth rate was constant and set equal to $\beta_0 = [\ln(50) - \ln(2)]/10^4 \approx 3.22 \times 10^{-4}$. This rate yields an expected number of lineages equal to 50 after 10^4 generations, which was the total time used in all phylogeny simulations. I henceforward refer to this set of phylogenetic trees as set (I).

In the second and third sets of 100 stochastic phylogeny simulations, sets (II) and (III), I linearly decreased and linearly increased the birth rate over time according to the equation:

$$\beta_t = \beta_0 + \frac{2d\beta_0(t - t_{total}/2)}{t_{total}}, \qquad (2)$$

where β_0 is as specified for (I), above, d is a parameter determining the rate of change of the speciation rate and was set to -1 or 1 for (II) and (III), respectively, t is time, and t_{total} is the total time of the simulation. [Obviously, eq. 2 also applies to (I), but in that case d is set equal to 0]. For both $d = -1$ and $d = 1$, the expected number of lineages is still 50 because the average speciation rate is the same as in (I).

QUANTITATIVE GENETIC SIMULATION MODEL

I used individual-based, genetically explicit numerical simulations to simulate the evolution of quantitative traits in a phylogenetic context according to the following procedure:

I used a Monte Carlo simulation with all individuals modeled. I simulated diploid, hermaphroditic, sexually reproducing populations of size $N_e = 100$.

In all simulations, I simulated the evolution of four traits determined by $m = 20$ unlinked pleiotropic loci. A mutation at a locus, which occurred with an allele generation probability of $\mu = 0.0025$ at all loci, produced a new allele with new effects on all four traits. The new allelic states were determined by adding a set of four mutational effects to the set of prior states for the allele. The mutational effects were drawn from a multivariate normal distribution with means [0, 0, 0, 0], variances $\boldsymbol{\alpha}^{\mathbf{2}} = [0.05, 0.1, 0.15, 0.20]$, and correlational mutation matrix

$$\mathbf{R}_{\mu} = \begin{bmatrix} 1.0 & 0.75 & 0.50 & 0 \\ 0.75 & 1.0 & 0.75 & 0.50 \\ 0.50 & 0.75 & 1.0 & 0.75 \\ 0 & 0.50 & 0.75 & 1.0 \end{bmatrix}.$$

This corresponds to a continuum of alleles mutation model (Crow and Kimura 1964) with correlated effects of pleiotropic mutations (Jones et al. 2003) and results in a mutational variance–covariance matrix, $\mathbf{M}$, with elements $M_{ij} = 2m\mu\sqrt{\alpha_i^2\alpha_j^2}R_{\mu ij}$ (Falconer and Mackay 1996). I chose this combination of correlational mutation matrix and set of mutational variances because it results in a variety of expected genetic variances for and covariances among characters. It also results in a set of expected additive genetic variances equal to [1.0, 2.0, 3.0, 4.0] (see below).

I determined the values for individual phenotypes by summing effects across all loci. In each generation I simulated reproduction by randomly selecting, with replacement, N_e pairs of mates, and then reproducing each pair once by randomly segregating one gene copy at each locus from each parent into the offspring. This corresponds to pure genetic drift in a sexually reproducing hermaphroditic population. The genetic segregation model corresponds to no physical linkage among loci.

For each phylogeny, I first initiated the population at genetic uniformity by setting the value of each allele at each locus in each individual to 0.0. I then simulated evolution by mutation and drift for a 1,000 generation "burn-in" prior to the phylogenetic component of each simulation. This was to allow populations to attain mutation–drift equilibrium. On each phylogeny, I simulated evolution on each branch for a time period specified by the length, in generations, of the internode. At each branching point the population was duplicated and then evolved along each daughter branch. I chose this procedure over the alternative of dividing the population into the daughter branches because the latter would have doubled the rate of drift in the generation of speciation.

At the tips of the tree, I recorded the means and the elements of the additive genetic variance–covariance ($\mathbf{G}$) matrix for each population.

ANALYSIS

For each phylogeny in sets (I), (II), and (III) I calculated the mean value of $\mathbf{G}$, $\bar{\mathbf{G}}$, from all $\mathbf{G}$ matrices from the tips. I also calculated an expected value for $\mathbf{G}$ as $\widehat{\mathbf{G}} = 2N_e\mathbf{M}$ (Falconer and Mackay 1996).

As a check of the simulation results, I compared the overall mean of $\bar{\mathbf{G}}$ from simulations (I), (II), and (III) to the analytic expected value of $\mathbf{G}$ ($\widehat{\mathbf{G}}$). I also computed the "phylogenetic means" for all characters in each phylogeny and compared the mean and variance for the phylogenetic means to those expected under genetic drift. The specific methodology and results are presented in online Supplementary Appendix S1. Agreement between analytic expectations and the results of this study was high for all circumstances in which the former were available or could be derived.

I calculated the variance–covariance matrix of the tip means in a typical fashion. This matrix will henceforward be represented as the nonphylogenetic divergence matrix, $\mathbf{D}$. Finally, for all phylogenies I also calculated the MS-MCP matrix of independent contrasts. This matrix will be represented as $\mathbf{D_{IC}}$ following Revell et al. (2007).

As a further check of the simulation results, I also calculated the mean value of $\mathbf{D_{IC}}$ and compared it to its theoretic value based on equation (1). These results are also presented in online Supplementary Appendix S1.

For all phylogenies I then calculated the vector-correlations between $\bar{\mathbf{G}}$ and $\mathbf{D}$, and $\bar{\mathbf{G}}$ and $\mathbf{D_{IC}}$, as well as between $\widehat{\mathbf{G}}$ and $\mathbf{D}$, and $\widehat{\mathbf{G}}$ and $\mathbf{D_{IC}}$. These will henceforward be referred to as $r(\bar{\mathbf{G}},\mathbf{D})$ and $r(\bar{\mathbf{G}},\mathbf{D_{IC}})$, or $r(\widehat{\mathbf{G}},\mathbf{D})$ and $r(\widehat{\mathbf{G}},\mathbf{D_{IC}})$. Although matrices can be compared in many ways (Steppan et al. 2002), I chose to calculate the element-by-element vector-correlation coefficient between the matrices because it provides a direct measure of the fit of $\mathbf{G}$ and $\mathbf{D}$ or $\mathbf{D_{IC}}$ to Lande's equation (1; above). The vector-correlation was used in lieu of the standard Pearson product-moment correlation coefficient because equation (1) has no intercept term and the vector-correlation is a correlation through the origin.

To ensure that the results obtained herein were not biased by my choice of matrix comparison method, for this and all subsequent simulations, I also compared $\mathbf{G}$ to $\mathbf{D}$ or $\mathbf{D_{IC}}$ using a simple modification of Roff et al.'s (1999) T-method. These analyses and results are presented in online Supplementary Appendix S2. Results using Roff et al.'s (1999) T-method were in close agreement to results obtained using the vector-correlation, but I prefer the vector-correlation because it provides a direct measure of the fit to equation (1) of Lande (1979).

To determine if $\mathbf{D_{IC}}$ provided a better fit to $\mathbf{G}$ than did the nonphylogenetic divergence matrix, $\mathbf{D}$, for a given simulation scenario, I evaluated the difference between the vector-correlations $r(\bar{\mathbf{G}},\mathbf{D_{IC}}) - r(\bar{\mathbf{G}},\mathbf{D})$ for all phylogenies [as well as all $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D})$], calculated the mean difference and its standard error, and compared their ratio to a *t*-distribution with df = 99. This test addresses the specific question: is $\mathbf{G}$ more tightly aligned to $\mathbf{D}$ or $\mathbf{D_{IC}}$?

To ensure that the results were not affected by any failure to satisfy distributional assumptions of the parametric statistical tests used herein, for this and all subsequent parametric statistical tests I compared the results to those obtained with nonparametric statistical methods. These results are presented in online Supplementary Appendix S3. In general, agreement between parametric and nonparametric methods was high.

To determine if there was an effect of speciation rate on the fit of $\mathbf{G}$ and $\mathbf{D}$, for each phylogeny in sets (II) and (III) I calculated $r(\bar{\mathbf{G}},\mathbf{D})$ and $r(\bar{\mathbf{G}},\mathbf{D_{IC}})$, or $r(\hat{\mathbf{G}},\mathbf{D})$ and $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$, and evaluated the differences, $r(\mathbf{G},\mathbf{D_{IC}}) - r(\mathbf{G},\mathbf{D})$ using both $\bar{\mathbf{G}}$ and $\hat{\mathbf{G}}$. I calculated *F* from a univariate ANOVA with the speciation parameter *d* in equation [2; −1, 0, and 1] as treatment and the differences, $r(\mathbf{G},\mathbf{D_{IC}}) - r(\mathbf{G},\mathbf{D})$, as the dependent variable. The groups ($d = -1$, 0, or 1) correspond to linearly decreasing, constant, and linearly increasing speciation rates, respectively. I compared *F*-statistics from this analysis to an *F*-distribution with df = 2,297. This test addresses the specific question: does speciation history influence the importance of a phylogenetic correction in a test of the constraint hypothesis?

BOUNDED PHENOTYPIC EVOLUTION

For each phylogeny in set (I), in which speciation occurred at a constant rate, I performed a fourth and final set of simulations. In these simulations, I simulated evolution primarily by mutation and drift as before, but I also imposed bounds on phenotype space—as if by functional constraint. To simulate bounded morphospace, I set the fitness of each individual to 1.0 if $\sum_{i=1}^{4} z_i^2 \leq b^2$, where *b* was the radius of the bounds on phenotype space, and to 0.0 otherwise. During reproduction, if I randomly selected an individual with a fitness of 0.0 to mate it was discarded and another individual was chosen until one with a fitness of 1.0 was found. I set *b* to 10.0 for all simulations. To determine if there was an effect of bounds on phenotype space on the fit of $\mathbf{G}$ and $\mathbf{D}$, for all pairs of $r(\bar{\mathbf{G}},\mathbf{D})$ and $r(\bar{\mathbf{G}},\mathbf{D_{IC}})$, or $r(\hat{\mathbf{G}},\mathbf{D})$ and $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$, I calculated the difference $r(\mathbf{G},\mathbf{D_{IC}}) - r(\mathbf{G},\mathbf{D})$ and compared those to the differences obtained from set (I) when phenotype space was not bounded using a standard *t*-test. This test addresses the specific question: do bounds on phenotype space influence the consequences of a phylogenetic correction when evolution has otherwise occurred by drift?

Results

$\mathbf{D_{IC}}$ VERSUS D

When speciation occurs at a constant rate the average vector-correlation between $\mathbf{G}$ and $\mathbf{D_{IC}}$ was higher than that between $\mathbf{G}$ and $\mathbf{D}$ regardless of whether the mean value of $\mathbf{G}$ from the tips ($\bar{\mathbf{G}}$) or the theoretic value of $\mathbf{G}$ ($\hat{\mathbf{G}}$) was used, although the mean correlation was high in both cases—as expected under drift [mean $r(\bar{\mathbf{G}},\mathbf{D_{IC}}) = 0.99$, mean $r(\bar{\mathbf{G}},\mathbf{D}) = 0.95$; mean $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) = 0.99$, $r(\hat{\mathbf{G}},\mathbf{D}) = 0.95$]. Nonetheless, $r(\bar{\mathbf{G}},\mathbf{D_{IC}}) - r(\bar{\mathbf{G}},\mathbf{D})$ was highly significantly greater than 0 [mean difference = 0.0372; t (df = 99) = 7.40; P(one-tailed) < 0.001] as was $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D})$ [mean difference = 0.0381; t (df = 99) = 7.74; P(one-tailed) < 0.001; Fig. 1A].

Both $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$ and $r(\hat{\mathbf{G}},\mathbf{D})$ were correlated with the reciprocal of the number of taxa in the phylogenetic tree from which they were calculated (Fig. 2). This indicates an asymptotic relationship between the vector-correlation between the matrices and the sample size. For both $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$ and $r(\hat{\mathbf{G}},\mathbf{D})$, the intercept of the model [which is equivalent to the predicted value for $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$ or $r(\hat{\mathbf{G}},\mathbf{D})$ as the number of taxa goes to ∞] was very close to 1.0 [$r(\hat{\mathbf{G}},\mathbf{D})$: intercept = 0.98; $r(\hat{\mathbf{G}},\mathbf{D_{IC}})$: intercept = 1.00]. Although numerically close, the intercept of the regression with $r(\hat{\mathbf{G}},\mathbf{D})$ was nonetheless significantly less than 1.0 [t (df = 98) = −1.91,

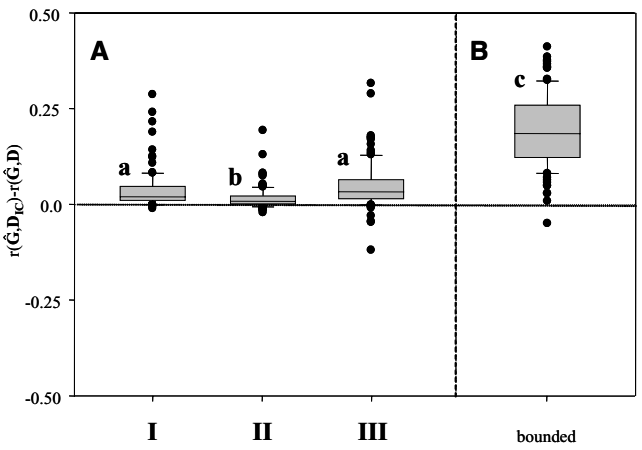

Figure 1. **$r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D})$ from simulation, for various speciation histories and simulation conditions. (A) Speciation histories are: (I) constant rate speciation; (II) linearly decreasing speciation rate over time; and (III) linearly increasing speciation rate over time. (B) Simulations were performed in bounded phenotype space. Differences are from individual-based simulations on the same set of phylogenies as in (I). All sets of differences are significantly greater than 0, indicating that $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) > r(\hat{\mathbf{G}},\mathbf{D})$ regardless of the speciation or simulation model. Groups indicated by a different lowercase letter (i.e., a, b, c) are also significantly different from each other by ANOVA and Tukey's post-hoc HSD. Results are based on four sets of 100 individual-based, genetically explicit computer simulations. Additional details are provided in the text.**

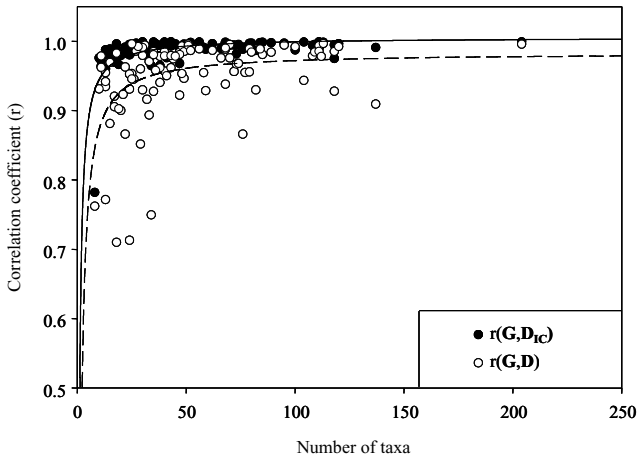

Figure 2. **r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) and r($\widehat{\mathbf{G}}$,D) plotted against the number of taxa in the phylogeny. Both are low when the number of taxa is low, but r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) increases marginally significantly more rapidly than does r($\widehat{\mathbf{G}}$,D) as the number of taxa increases [*t* (df = 198) = 1.61, *P* (one-tailed) = 0.055]. Inverse functions are (A) r($\widehat{\mathbf{G}}$, $\mathbf{D_{IC}}$) = $-0.61n^{-1} + 1.00$ and (B) r($\widehat{\mathbf{G}}$, D) = $-1.11n^{-1} + 0.98$, in which *n* is the number of taxa, for $\mathbf{D_{IC}}$ and D, respectively. The intercept (asymptote) of equation (B) is significantly less than 1.0 [*t* (df = 99) = −1.91, *P* (one-tailed) = 0.030].**

P (one-tailed) = 0.030]. Both r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) and r($\widehat{\mathbf{G}}$,**D**) are frequently low when the number of taxa is low. On average, r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) increases marginally significantly more rapidly than does r($\widehat{\mathbf{G}}$,**D**) as the number of taxa increases [t (df = 198) = 1.61, P = 0.055; Fig. 2]. I obtained virtually identical results using r($\bar{\mathbf{G}}$,$\mathbf{D_{IC}}$) and r($\bar{\mathbf{G}}$,**D**).

INCREASING AND DECREASING SPECIATION RATES

The mean difference between the vector-correlations of **G** and $\mathbf{D_{IC}}$ and of **G** and **D**, r(**G**,$\mathbf{D_{IC}}$) − r(**G**,**D**), was significantly lower when the speciation rate was initially high but decreased over time (III), when compared to the mean difference, r(**G**,$\mathbf{D_{IC}}$) − r(**G**,**D**), obtained when the speciation rate was constant over time (I) [mean r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) = 0.98, mean r($\widehat{\mathbf{G}}$,**D**) = 0.96; results similar for $\bar{\mathbf{G}}$]. Furthermore, the mean difference, r(**G**,$\mathbf{D_{IC}}$) − r(**G**,**D**), was higher when the speciation rate was initially low but increased over time (II), although not significantly higher than when speciation occurred at a constant rate [mean r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) = 0.98, mean r($\widehat{\mathbf{G}}$,**D**) = 0.93; results similar for $\bar{\mathbf{G}}$]. These findings were indicated by a highly significant ANOVA and confirmed by post-hoc tests, and were similar whether $\bar{\mathbf{G}}$ or $\widehat{\mathbf{G}}$ was used [results using $\bar{\mathbf{G}}$: ANOVA F (df = 2,297) = 10.6; $P < 0.001$; results using $\widehat{\mathbf{G}}$: ANOVA F (df = 2,297) = 11.2; $P < 0.001$; Fig. 1A].

BOUNDED PHENOTYPIC EVOLUTION

In general, when compared to the vector-correlation between $\mathbf{D_{IC}}$ or **D** and **G** when evolution occurred without functional constraint, both $\mathbf{D_{IC}}$ and **D** were poorly correlated with **G** when evolution was bounded. For example, as detailed above, when the speciation rate was constant and evolution was unbounded, the average vector-correlation r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) = 0.99, and the average correlation r($\widehat{\mathbf{G}}$,**D**) = 0.95. By contrast, with functional constraints imposed, average r($\widehat{\mathbf{G}}$,$\mathbf{D_{IC}}$) decreased to 0.77 and average r($\widehat{\mathbf{G}}$,**D**) to 0.58.

Nonetheless, substituting $\mathbf{D_{IC}}$ for **D** yielded a significantly larger mean difference, r(**G**,$\mathbf{D_{IC}}$) − r(**G**,**D**), when evolution was bounded by functional constraints than when it was unbounded [as in simulation (I), above; for $\widehat{\mathbf{G}}$: t (df = 150.8) = 14.3, P(two-tailed) < 0.001; Fig. 1B; results similar for $\bar{\mathbf{G}}$].

Discussion

With only a few exceptions (Baker and Wilkinson 2003; Bégin and Roff 2004; Revell et al. 2007) prior studies testing the genetic constraint hypothesis by comparing among- and within-species patterns of variation and covariation have ignored phylogenetic nonindependence among taxa. In this study I simulated multivariate phenotypic evolution by genetic drift and constraint on stochastic phylogenetic trees using individual-based quantitative genetic simulations. I used the results from these simulations to compare the outcome from tests for constraint in which a phylogenetic correction was used to those obtained when phylogeny was ignored. The results indicated strongly that, on average, a higher vector correlation is observed between the among population variance covariance matrix (**D** or $\mathbf{D_{IC}}$ in this study) and the within population genetic variance–covariance matrix (**G**) when the phylogenetic correction is used. Because the equation of Lande (eq. 1, above) on which this type of test is based predicts proportionality between within- and among-population matrices under the circumstances of my numerical simulations, higher correlation between **G** and the phylogenetic divergence matrix ($\mathbf{D_{IC}}$) would seem to indicate that this latter statistic should be preferred by empiricists over the nonphylogenetic among-population matrix, **D**.

Higher correlation between **G** and the phylogenetic divergence matrix, $\mathbf{D_{IC}}$ is due not only to the fact that the mean-squares and cross-products of the independent contrasts provide an unbiased estimate of the evolutionary rate matrix (Garland and Ives 2000), but also because $\mathbf{D_{IC}}$ is estimated with lower error. This is evidenced by the higher variance in **D** than in $\mathbf{D_{IC}}$ (online Supplementary Appendix S4), and agrees well with the findings obtained analytically and by numerical simulation from the phylogenetic generalized least-squares approach by Rohlf (2006).

Under pure drift, the difference between r(**G**,$\mathbf{D_{IC}}$) and r(**G**,**D**) is largest, and thus the consequences of ignoring phylogeny are most severe, when speciation rate is initially low and increases toward the present (Fig. 1A, III), as, perhaps, during a recent adaptive radiation (Schluter 2000). Because rejecting the

proportionality of **G** and **D** is used as evidence for natural selection (e.g., Merilä and Björklund 1999), students of adaptive radiation in particular should be careful to use a phylogenetically corrected divergence matrix ($\mathbf{D_{IC}}$ in this study) to help avoid falsely rejecting a pattern of phenotypic differentiation consistent with drift in favor of adaptive explanations.

The difference between r(**G**,$\mathbf{D_{IC}}$) and r(**G**,**D**) is also large when evolution is bounded by limits on morphospace, perhaps analogous to those imposed by functional constraint (Fig. 1B). This is because—as most evolution is by genetic drift so long as the population is not near the bound of the morphospace—many contrasts are calculated between tree nodes phenotypically differentiated by drift alone. For example, in the hypothetical case illustrated in Figure 3, only the contrast between terminal taxa c and d is directly affected by functional constraint [although our estimation of the contrast between interior nodes (a,b) and (c,d) is also affected because the state at node (c,d) is estimated from the traits values of its daughters]. After sufficient time, the pattern of genetic constraint may no longer be evident in **D**, but will be present to some extent in $\mathbf{D_{IC}}$ so long as some recent contrasts exist that do not involve the morphospace boundary.

Some prior similar studies have provided the lack of phylogenetic signal as a rationale for ignoring phylogenetic nonindependence among the taxa in the study (e.g., Bégin and Roff 2004). Phylogenetic signal tends to be eroded over time by bounded morphospace (online Supplementary Appendix S5). Ignoring phylogenetic nonindependence may be well justified under some circumstances. Nonetheless, my simulations of evolution in a bounded morphospace suggest that an important aspect of the process of differentiation under drift and functional constraint (genetic constraint) could easily be overlooked if phylogenetic nonindependence is ignored.

The difference between r(**G**,$\mathbf{D_{IC}}$) and r(**G**,**D**) is smallest, and thus phylogenetic correction least important, when speciation is initially rapid and decreases over time. This is because phylogenies generated by this process tend to have a star-like shape, which is the tree shape implicitly assumed by the nonphylogenetic test.

Short early branches in a molecular phylogeny can also be produced as an artifact of branch length estimation (Revell et al. 2005), so researchers should consider their results with caution if the phylogenetic tree has many short early branches. This disclaimer raises the important consideration that in most empirical studies the phylogenetic tree used in a given evolutionary inference, such as in the estimation of the phylogenetic divergence matrix, $\mathbf{D_{IC}}$, is not a history known without error (as in this study), but an estimate of that history. For studies in which the phylogenetic history is known with error (which is almost every empirical study), researchers might consider estimating $\mathbf{D_{IC}}$ for a sample of trees from the posterior distribution of trees in a Bayesian analysis to acquire a measure of the variability in $\mathbf{D_{IC}}$ that is likely to be due to phylogenetic error (Revell et al. 2007).

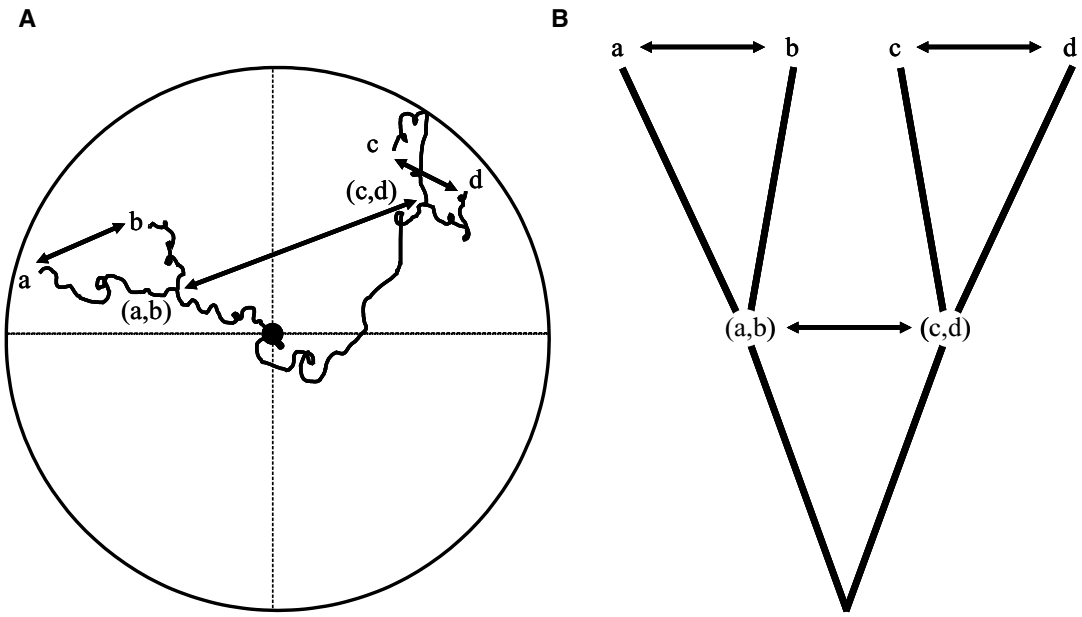

Figure 3. **A hypothetical illustration of genetic drift and speciation in a bounded, bivariate phenotype space. Bounded phenotype space is somewhat analogous to functional constraint. In (A) a hypothetical trajectory of evolution by drift is illustrated for a phylogenetic history with four terminal taxa and three independent contrasts. Only the contrast between tip nodes c and d is directly affected by functional constraint, although the contrast between interior nodes (a,b) and (c,d) is also affected as the state at node (c,d) is not known, but instead estimated from the states at terminal taxa c and d. (B) is shown only to clarify the phylogenetic history of the taxa in (A).**

Although I show that for the conditions explored in this study $\mathbf{D}_{\mathbf{IC}}$ is preferable to **D** in tests of the genetic constraint hypothesis based on equation (1), the severely computationally intensive nature of the numerical simulations performed herein imposed two specific types of limitations.

First, computational constraints limited the number of simulations that could be reasonably performed and thus the scope of the simulation conditions explored. Many circumstances of considerable interest were not explored in this study. For example, speciation rate can change over time in a variety of ways aside from by linearly increasing or decreasing. No conditions aside from strict genetic additivity were explored, nor was the condition of genes of major effect. Both departures from additivity, such as epistasis and dominance, and genes of large effect are common empirical observations and are known to be important in the evolutionary process (e.g., Wolf et al. 2000; Agrawal et al. 2001). In addition, only the processes of pure genetic drift and a very simple functional constraint were simulated in this study. Not explored by this study is the behavior of a phylogenetic divergence matrix when evolution occurs by processes other than those described above, such as by natural selection.

Furthermore, the bounds on phenotype space herein simulated no doubt represent a highly simplified version of functional constraint. Certainly, bounds exist on phenotype space for all lineages, even those evolving primarily by drift. Species cannot evolve to negative size, nor can they evolve to size larger than the Earth. However, as the true limits on phenotype space imposed by functional constraint are poorly known and computational limitations required that the number of simulations be few, I was unfortunately forced to limit my exploration in this area to very simple conditions.

Second, the simulation parameter values used in this study were not especially realistic. In particular, the mutation rate is unrealistically high, the effective population size unrealistically low, and the total time of the simulations unrealistically short. My implied assumption is that increasing one parameter while decreasing another can have compensatory effect. This assumption is not unprecedented. For example, Jones et al. (2003, 2004) use admittedly small population size, high mutation rates, and short simulations, as I do in this study. I also show the existence of compensatory effects in another study (in that case, pertaining to the evolutionary stability of **G**; Revell 2007). Nonetheless, unrealistic parameter values are a limitation imposed on computer intensive simulations both by existing hardware and the programming ability of the author. Future studies performed on more sophisticated computer equipment should attempt to use more realistic parameter values.

In spite of these limitations, I hope that the evidence presented herein is sufficient to convince the reader that the phylogenetic divergence matrix should be preferred in a test for genetic constraint over the divergence matrix calculated ignoring phylogenetic nonindependence—so long as a phylogeny is available for the species in the study.

ACKNOWLEDGMENTS

This work was supported in part by a grant from the National Science Foundation (DEB-0519777). I owe a great deal of thanks to L. Harmon for inspiring much of this study in our discussions of the topic. I owe thanks also to the editor and anonymous reviewers of a previous manuscript whose comments helped motivate this study, and to the Losos Lab, associate editor S. Otto, and two anonymous reviewers for commenting on earlier drafts of this manuscript.

LITERATURE CITED

Ackermann, R. R., and J. M. Cheverud. 2002. Discerning evolutionary processes in patterns of tamarin (genus *Saguinus*) craniofacial variation. Am. J. Phys. Anthropol. 117:260–271.

Agrawal, A. F., E. D. Brodie III, and L. H. Reiseberg. 2001. Possible consequences of genes of major effect: transient changes in the G-matrix. Genetica 112–113:33–43.

Baker, R. H., and G. S. Wilkinson. 2003. Phylogenetic analysis of correlation structure in stalk-eyed flies (*Diasemopsis*, Diopsidae). Evolution 57:87–103.

Bégin, M., and D. A. Roff. 2003. The constancy of the **G** matrix through species divergence and the effects of quantitative genetic constraints on phenotypic evolution: a case study in crickets. Evolution 57:1107–1120.

———. 2004. From micro- to macroevolution through quantitative genetic variation: positive evidence from field crickets. Evolution 58:2287–2304.

Crow, J. F., and M. Kimura. 1964. The theory of genetic loads. Pp. 495–505 *in* S. J. Geerts, ed. Proceedings of the XI international congress of genetics. Pergamon, Oxford, U.K.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Prentice Hall, Essex U.K.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1–15.

Garland, T. Jr., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. Am. Nat. 155:346–364.

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

———. 2004. Evolution and stability of the **G**-matrix on a landscape with a moving optimum. Evolution 58:1639–1654.

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

Marroig, G., M. de Vivo, and J. M. Cheverud. 2004. Cranial evolution in sakis (*Pithecia*, Platyrrhini) II: evolutionary processes and morphological integration. J. Evol. Biol. 17:144–155.

Merilä, J., and M. Björklund. 1999. Population divergence and morphometric integration in the greenfinch (*Carduelis chloris*)—evolution against the trajectory of least resistance? J. Evol. Biol. 12:103–112.

Revell, L. J. 2007. The **G** matrix under fluctuating correlational mutation and selection. Evolution. 61:1857–1872.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. Syst. Biol. 54:973–983.

Revell, L. J., L. J. Harmon, R. B. Langerhans, and J. J. Kolbe. 2007. A phylogenetic approach to determining the importance of constraint on phenotypic evolution in the Neotropical lizard *Anolis cristatellus*. Evol. Ecol. Res. 9:261–282.

Roff, D. A., T. A. Mousseau, and D. J. Howard. 1999. Variation in genetic architecture of calling song among populations of *Allonemobius socius*, *A. fasciatus*, and a hybrid population: drift or selection? Evolution 53:216–224.

Rohlf, F. J. 2006. A comment on phylogenetic correction. Evolution 60:1509–1515.

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

———. 2000. The ecology of adaptive radiation. Oxford Univ. Press, Oxford, U.K.

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

Wolf, J. B., E. D. Brodie III, and M. J. Wade. 2000. Epistasis and the evolutionary process. Oxford Univ. Press, Oxford, U.K.

Associate Editor: S. Otto

Supplementary Material

The following supplementary material is available for this article:

Appendix S1. Comparison of results to analytic predictions.
Appendix S2. Alternative matrix comparison method.
Appendix S3. Comparison of results to non-parametric tests.
Appendix S4. Variance in **D** and $\mathbf{D}_{\mathbf{IC}}$.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2007.00216.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.