These are electronic supplementary appendices to: Revell, L. J. 2007. Testing the genetic constraint hypothesis in a phylogenetic context: A simulation study. *Evolution* 61: 2720-2727.

Supplementary Appendix A. Comparison of results to analytic predictions.

As a check of the numerical simulations, I compared some of the results obtained in this study to those expected based on analytic predictions. Under genetic drift, analytic predictions for **G**, the 'phylogenetic mean' (i.e., the expected value at the tips), the variance on the phylogenetic mean, and $\mathbf{D_{IC}}$ are all relatively straightforward to obtain.

To determine whether the observed values for **G** conform to analytic predictions, I first computed a grand mean value for the **G** matrix across simulations (I), (II), and (III). I excluded bounded simulations from this analysis because constrained morphospace acts as a selective force and thus makes an analytic prediction of **G** impossible to obtain exactly (Falconer and Mackay 1996). I calculated the overall mean of $\overline{\mathbf{G}}_{\mathrm{GM}}$ by first averaging **G** across all tips in each phylogeny to compute $\overline{\mathbf{G}}$ for each tree and dataset, and then by averaging $\overline{\mathbf{G}}$ across the total of 300 phylogenies in sets (I), (II), and (III). I then compared this grand mean, $\overline{\mathbf{G}}_{\mathrm{GM}}$ to the predicted value for $\widehat{\mathbf{G}}$ which is provided in the text as $\widehat{\mathbf{G}} = 2\mathrm{N_e}\mathbf{M}$.

$\overline{\mathbf{G}}_{\mathrm{GM}}$ and $\widehat{\mathbf{G}}$ were highly similar :

$$\overline{\mathbf{G}}_{\mathrm{GM}} = \begin{bmatrix} 0.99 & 1.05 & 0.86 & 0.00 \\ 1.05 & 1.98 & 1.83 & 1.41 \\ 0.86 & 1.83 & 2.97 & 2.57 \\ 0.00 & 1.41 & 2.57 & 3.95 \end{bmatrix} \quad \text{and} \quad \widehat{\mathbf{G}} = \begin{bmatrix} 1.00 & 1.06 & 0.87 & 0.00 \\ 1.06 & 2.00 & 1.84 & 1.41 \\ 0.87 & 1.84 & 3.00 & 2.60 \\ 0.00 & 1.41 & 2.60 & 4.00 \end{bmatrix}.$$

The element by element matrix correlation is $\mathrm{r}(\overline{\mathbf{G}}_{\mathrm{GM}}, \widehat{\mathbf{G}}) = 1.00$ and the regression coefficient from ordinary least-squares regression is $\beta(\overline{\mathbf{G}}_{\mathrm{GM}}, \widehat{\mathbf{G}}) = 1.00$.

To determine if the phylogenetic mean was equivalent to the starting condition (0.0), as predicted under drift, I calculated the phylogenetic means for each simulation and averaged across simulations in sets (I), (II), and (III). Once again, I excluded bounded simulations for the reasons discussed in the preceding text.

The phylogenetic mean is equivalent to the state at the ancestral node in the tree and can be computed following Rohlf (2001) as:

$$a = (\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1})^{-1}(\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{X}).$$

Here a is the phylogenetic mean for the trait of interest, **1** is column vector of length n and composed of ones (where n = the number of taxa in the tree), superscript $^{\mathrm{T}}$ indicates that a transpose is calculated, **C** is the coancestry matrix – a matrix proportional to the expected covariances among the tips (following Rohlf 2001), and **X** is column vector of length n composed of the character states at each tip.

The observed mean value for the phylogenetic means of the four traits, $\bar{\mathbf{a}}$ agreed quite well with analytic predictions:

$$\text{expected}(\hat{\mathbf{a}}) = [0.00,\ \ 0.00,\ \ 0.00,\ \ 0.00],$$

$$\text{observed}(\bar{\mathbf{a}}) = [0.29\ (0.32),\ 0.30\ (0.43),\ 0.78\ (0.57),\ 0.35\ (0.61)].$$

Standard errors for each value of $\bar{\mathbf{a}}$ (from variances, below) are shown in parentheses behind

each mean. The correlation between expected($\hat{\mathbf{a}}$) and observed($\bar{\mathbf{a}}$) is meaningless to evaluate, since there is no variance in the former – however no value of $\bar{\mathbf{a}}$ was significantly different from its expected value of 0.0.

Variance among runs in the values of the phylogenetic means for a given trait comes from two sources. Firstly, variance comes from divergence among simulation runs during the 1,000 generation 'burn-in' preceding the phylogenetic component of the simulation. Secondly, error variance is associated with the estimation of the phylogenetic mean (Rohlf 2001). These variance components are independent and thus to compute the expected total variance in *a*, these sources of variance can just be added together.

Normally, the expected variance among runs after 1,000 generations of drift could easily be computed as $V_D = t \cdot V_A / \mathrm{N_e}$, which is the univariate version of equation [1] in the main text. However, because I initiated the simulations with conditions of genetic uniformity, genetic variance is accumulated by mutation at the same time as the populations differentiate. Consequently, computation of the expected value of V_D is much more complicated. In discrete time, this first component of variance (due to drift among simulation runs) can be estimated as follows:

$$V_D(t) = \sum_{i=1}^{t} \left(\frac{\sum_{j=1}^{i} V_M (1 - 1/2\mathrm{N_e})^{j-1}}{\mathrm{N_e}} \right).$$

In this equation, V_D is just the variance among populations, t is the total time of the simulation (1,000 generation burn-in for this example), V_M is the mutational variance, and $\mathrm{N_e}$ is the effective population size (100). This equation is based on the recursive function for the accumulation of additive genetic variance within a single lineage under mutation and drift (numerator and inner summation), and the rate of variance accumulation among lineages under drift (denominator and outer summation).

The second component of variance is due to error in the estimation of *a*. This error variance is given by Rohlf (2001) as:

$$V_{err} = (V_A / \mathrm{N_e}) \cdot (\mathbf{1}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{1})^{-1}.$$

I used the expected values of V_A, the additive genetic variance within each lineage, from the diagonal of $\widehat{\mathbf{G}} = 2\mathrm{N_e}\mathbf{M}$, as above and in the text. Here I can assume a constant value for V_A because only the branches after the burn-in are significant in estimation error in $\hat{a}$, and an equilibrium value for V_A is expected by the end of the burn-in. **1** and **C** are as described in preceding text.

Because the stochastic topology of each tree was different, I computed the total expected variance in $\hat{a}$ as the sum of V_D and the average value of V_{err}, i.e.:

$$V(\hat{a}) = \sum_{i=1}^{p} V_{err}(i) \Big/ p + V_D$$

for p = 300 trees in simulation sets (I), (II), and (III). This is possible because the expected variance among a set of observations drawn from distributions with different variances is equivalent to the mean value of those variances.

$V(\hat{a})$ agreed quite well with analytic predictions:

$$\text{expected}[\mathbf{V}(\hat{\mathbf{a}})] = [29.8,\quad 59.1,\quad 88.6,\quad 118.1],$$

$$\text{observed}[\mathbf{V}(\mathbf{a})] = [29.5,\quad 55.1,\quad 96.6,\quad 112.2].$$

The correlation between expected$[\mathbf{V}(\hat{\mathbf{a}})]$ and observed$[\mathbf{V}(\mathbf{a})]$ was r = 0.99, and

$\beta(\text{observed}[\mathbf{V}(\mathbf{a})], \text{expected}[\mathbf{V}(\hat{\mathbf{a}})]) = 0.98$.

To determine if $\mathbf{D_{IC}}$ had a mean value equivalent to analytic prediction, I calculated $\mathbf{D_{IC}}$ by scaling standardized independent contrasts to have an expected variance equal to the total tree length, which was held standard across simulations at 10^4 generations. This was accomplished by multiplying each contrast by the square root of the total tree length.

This yields an expected value of $\widehat{\mathbf{D}}_{\mathbf{IC}} = (\text{total tree length}) \times \widehat{\mathbf{G}} / \mathrm{N_e}$. Under the simulation conditions of this study, this expectation evaluates to:

$$\widehat{\mathbf{D}}_{\mathbf{IC}} = \begin{bmatrix} 100.0 & 106.1 & 86.6 & 0.0 \\ 106.1 & 200.0 & 183.7 & 141.4 \\ 86.6 & 183.7 & 300.0 & 259.8 \\ 0.0 & 141.4 & 259.8 & 400.0 \end{bmatrix}.$$

For comparison, the observed mean value of $\mathbf{D_{IC}}$ for simulations (I), (II), and (III) was:

$$\overline{\mathbf{D}}_{\mathbf{IC}} = \begin{bmatrix} 100.7 & 107.6 & 88.8 & 1.0 \\ 107.6 & 201.3 & 187.4 & 142.0 \\ 88.8 & 187.4 & 303.9 & 261.5 \\ 1.0 & 142.0 & 261.5 & 400.4 \end{bmatrix}.$$

Bounded simulations were again excluded, for reasons detailed above. $\widehat{\mathbf{D}}_{\mathbf{IC}}$ and $\overline{\mathbf{D}}_{\mathbf{IC}}$ seem to be in very close agreement.

Thus, by every measure possible, for circumstances in which such predictions are available (or derived herein) the results of this study conform well to analytic predictions.

Supplementary Appendix B. Alternative matrix comparison method.

In order to test for the generality of the results of this study, I also compared matrices using an alternative matrix comparison based on the T-method of Roff et al. (1999). In this method, matrices are compared by evaluating the summed absolute values of the difference between corresponding elements.

In the present study I compare within and among population matrices (**G** and **D** or $\mathbf{D_{IC}}$). These matrices are on different scales. Thus, before subtraction I first rescaled the matrices so that their elements had the same mean-square (by dividing each element by the square root of the mean-square of the elements). Results were highly similar for $\overline{\mathbf{G}}$ and $\hat{\mathbf{G}}$, so only the results for $\hat{\mathbf{G}}$ are shown.

Figure B1 shows $T(\widehat{\mathbf{G}}, \mathbf{D_{IC}}) - T(\widehat{\mathbf{G}}, \mathbf{D})$ for various speciation and evolutionary models. $T(\widehat{\mathbf{G}}, \mathbf{D_{IC}})$ is the summed absolute values of the differences between the elements of the theoretic expected value of the additive genetic variance-covariance matrix, **G**, and the mean squares, mean cross-products matrix of independent contrasts (phylogenetic divergence matrix), $\mathbf{D_{IC}}$. $T(\widehat{\mathbf{G}}, \mathbf{D})$ is the summed absolute values of the differences between the expected value of **G** and the among species variance-covariance matrix calculated ignoring phylogenetic non-independence (non-phylogenetic divergence matrix), **D**.

For constant rate speciation, $T(\widehat{\mathbf{G}}, \mathbf{D_{IC}}) - T(\widehat{\mathbf{G}}, \mathbf{D})$ was significantly less than 0.0 (mean

difference -1.18, *t* (df=99) = -11.1, P < 0.001), indicating that (scaled appropriately) $\mathbf{D_{IC}}$ is more similar to **G** than is **D** (Fig. B1A).

The model of speciation affected the difference, $T(\widehat{\mathbf{G}},\mathbf{D_{IC}}) - T(\widehat{\mathbf{G}},\mathbf{D})$, with the difference being significantly less negative than constant rate speciation if the speciation rate was initially high but decreased over time (speciation model II), and more negative, but not significantly so, when the speciation rate was initially low but increased over time (ANOVA F (df=2,297) = 16.7; P < 0.001; Fig. B1A).

Results for $T(\widehat{\mathbf{G}},\mathbf{D_{IC}}) - T(\widehat{\mathbf{G}},\mathbf{D})$ were also qualitatively similar to those obtained for the vector-correlation when evolution was bounded. Both $T(\widehat{\mathbf{G}},\mathbf{D_{IC}})$ and $T(\widehat{\mathbf{G}},\mathbf{D})$ were greater than in the unbounded case, and their difference was significantly more negative [*t* (df=185.6) = -4.84, P(two-tailed) < 0.001; Fig. B1B].

Thus, overall results were highly qualitatively highly similar whether matrix similarity was measured using vector-correlation, or via a modification of Roff et al.'s (1999) T-method approach.

Fig. B1

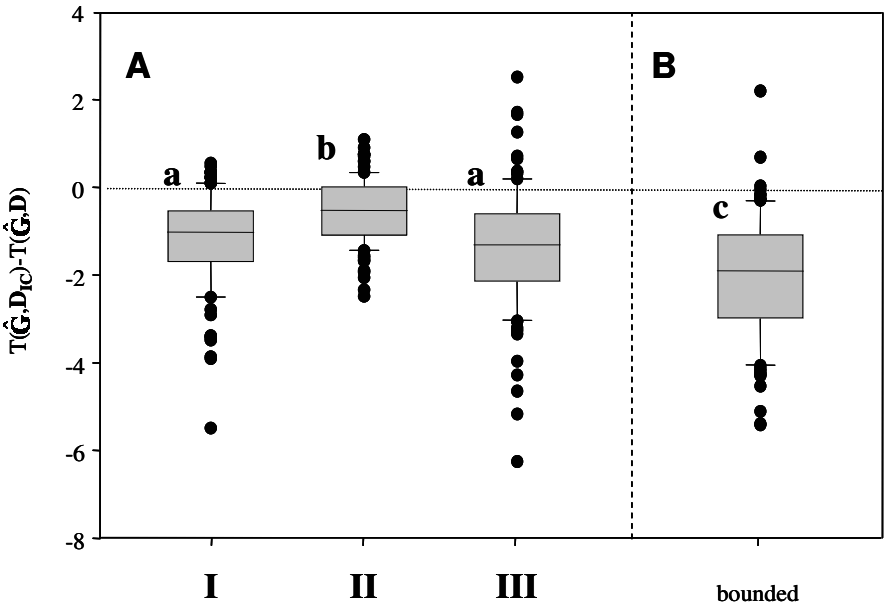

Figure B1. $T(\widehat{\mathbf{G}},\mathbf{D_{IC}})$ - $T(\widehat{\mathbf{G}},\mathbf{D})$ from simulation, for various speciation histories and simulation conditions. All matrices were rescaled to have equivalent mean-square prior to analysis. A. Speciation histories are as in Figure 1 and text. B. Simulation was performed in bounded phenotype space as described in Figure 1 and text. Differences are from individual-based simulations on the same set of phylogenies as in (I). Groups indicated by a different lowercase letter (i.e., **a**, **b**, **c**) are significantly different from each

Supplementary Appendix C. Comparison of results to non-parametric tests.

In order to determine if the results of this study were robust to violations in the assumptions of parametric statistics, I also performed analogous non-parametric statistical tests and compared the results. In this appendix, I present the results and significances of both parametric and non-parametric tests for comparison.

Table C1 shows the results from parametric and non-parametric tests of the hypothesis that $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D}) = 0.0$ against that alternative hypothesis that $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D}) > 0.0$, which would be the case if, as hypothesized, $\mathbf{D_{IC}}$ provides a better fit to Lande's (1979) equation (equation [1], in text) than does **D**. Qualitatively, parametric and non-parametric results are in close agreement.

Table C2 shows the results from parametric and non-parametric tests of an effect of speciation model (comparison: I, II, III) or bounded morphospace (comparison: I, bounded) on the difference $r(\hat{\mathbf{G}},\mathbf{D_{IC}})- r(\hat{\mathbf{G}},\mathbf{D})$. As in Table C1, parametric and non-parametric results of Table C2 are in close agreement.

Table C1. Parametric and non-parametric results from tests of the hypothesis that $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D}) = 0.0$ against that alternative hypothesis that $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D}) > 0.0$ for several simulation conditions. Conditions are: (I) constant rate speciation; (II) linearly decreasing speciation rate over time; (III) linearly increasing speciation rate over time, and (bounded) bounded phenotype space.

Simulation	Mean $[r(\hat{\mathbf{G}},\mathbf{D_{IC}})- r(\hat{\mathbf{G}},\mathbf{D})]$	*t* (df=99)	P (>t)	Wilcoxon W	P (>W)
I	0.037	7.34	<0.001	2397.5	<0.001
II	0.015	5.35	<0.001	1790.5	<0.001
III	0.047	7.69	<0.001	2214.0	<0.001
bounded	0.191	20.2	<0.001	2519.0	<0.001

Table C2. Parametric and non-parametric results from hypothesis tests for an effect on speciation model (I, II, III – see Table C1 caption) or simulation condition (I, bounded) on the difference, $r(\hat{\mathbf{G}},\mathbf{D_{IC}}) - r(\hat{\mathbf{G}},\mathbf{D})$.

Comparison	ANOVA F	ANOVA df	P (>F)	Kruskal-Wallis χ^2	P ($>\chi^2$)
I, II, III	11.2	2, 297	<0.001	36.6	<0.001
I, bounded	205.7	1, 198	<0.001	112.5	<0.001

Supplementary Appendix D. Variance in D and $\mathbf{D_{IC}}$.

For simulations (I), (II), and (III) the following variances among simulation runs were calculated for the elements composing **D** and $\mathbf{D_{IC}}$:

$$\text{var}(\mathbf{D}) = \begin{bmatrix} 1.40 & 2.10 & 1.98 & 2.11 \\ 2.10 & 5.23 & 5.47 & 5.78 \\ 1.98 & 5.47 & 11.9 & 13.5 \\ 2.11 & 5.78 & 13.5 & 22.3 \end{bmatrix} \times 10^3, \quad \text{var}(\mathbf{D_{IC}}) = \begin{bmatrix} 1.27 & 1.66 & 2.45 & 1.54 \\ 1.66 & 3.12 & 4.10 & 3.24 \\ 2.45 & 4.10 & 7.16 & 6.64 \\ 1.54 & 3.24 & 6.64 & 12.1 \end{bmatrix} \times 10^3 .$$

On average the elements of **D** are 1.47 times more variable than those of $\mathbf{D_{IC}}$ under genetic drift.

Supplementary Appendix E. Testing for phylogenetic signal in unbounded and bounded simulations.

Some of the simulations performed in this study involved simulating quantitative trait evolution in a bounded morphospace. The morphospace was strictly bounded in my individual based simulations by setting the fitness of individuals outside of the bounds to 0.0, and to 1.0 otherwise.

Bounded morphospace was simulated because this type of constraint is expected to erode so-called 'phylogenetic signal' (or the correlation between patristic distance and phenotypic similarity) over time. This expectation can easily be confirmed by comparing the mean value of the K-statistic of Blomberg et al. (2003) from unbounded and bounded simulations [$\overline{K}$ (unbounded) = 1.00; $\overline{K}$ (bounded) = 0.28]. Figure E1 shows the distribution of the natural logarithm of K for unbounded (black bars) and bounded (grey bars) simulations. The expected value of K under genetic drift and in the absence of constraint is 1.00 (the logarithm of which is 0.0, see Figure E1).

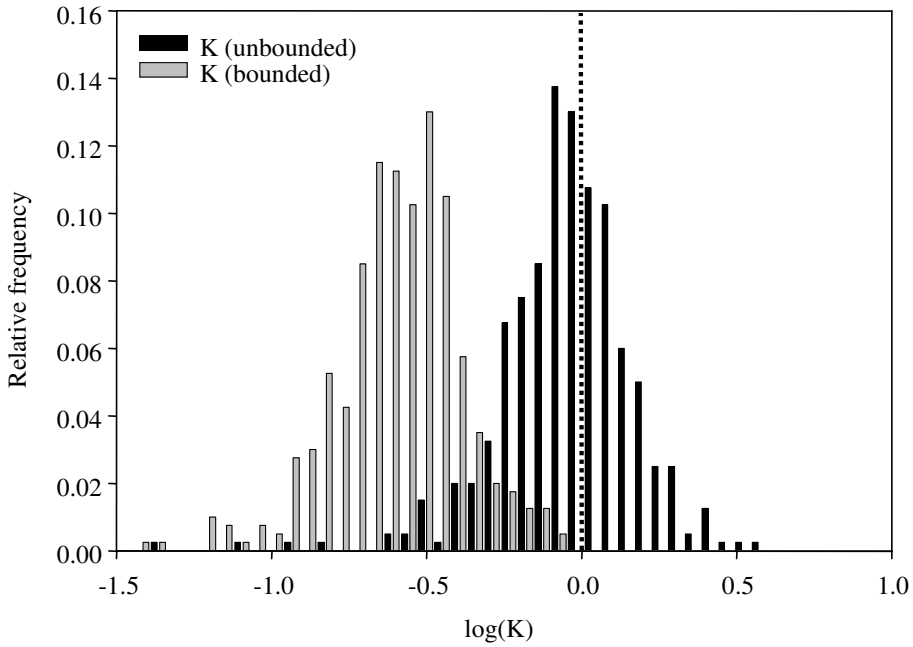

Figure E1. Histogram indicating the distribution of the natural logarithm of the K-statistic of Blomberg et al. (2003) for unbounded (black bars) and bounded (grey bars) quantitative genetic simulations. K provides a measure of the 'phylogenetic signal' in the data, and has an expected value of log(K) = 0.0 under genetic drift, which is indicated by the vertical dashed line. Phylogenetic signal is decreased when evolution is bounded.

Supplementary Appendix Literature Cited.

Blomberg, S. P., T. Garland Jr., and A. R. Ives. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57:717-745.
Falconer, D. S., and Mackay, T. F. C. 1996. Introduction to Quantitative Genetics. Prentice Hall, Essex UK.
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. 2001. Comparative methods for the analysis of continuous variables: Geometric interpretations. *Evolution* 55:2143-2160.