Supplementary information

Analyzing disparity and rates of morphological evolution with model-based phylogenetic comparative methods

Thomas F. Hansen¹, Geir H. Bolstad² and Masahito Tsuboi¹,³

¹University of Oslo, Department of Biology, CEES & Evogene, Oslo, Norway
²Norwegian Institute for Nature Research (NINA), Trondheim, Norway
³Lund University, Department of Biology, Lund, Sweden

Contents

Supplementary figures ........................................................................................................................................... 2
Supplementary simulations ................................................................................................................................. 5
  Model 1.......................................................................................................................................................... 5
    Simulations on the bird phylogeny ........................................................................................................... 5
    Simulations on simulated phylogenies ..................................................................................................... 7
  Model 2........................................................................................................................................................ 9
    Simulations on the bird phylogeny ........................................................................................................... 9
    Simulations on simulated phylogenies ..................................................................................................... 11
  Model 3......................................................................................................................................................... 14
    Simulations on the bird phylogeny ........................................................................................................... 14
    Simulations on simulated phylogenies ..................................................................................................... 15
Figure S1: The effect of log brain size on beak-shape evolution. The plot shows the magnitude of the deviance from clade mean of beak-shape variables plotted against log brain mass for 645 bird species. The fitted line and parameter estimates are from Model 1 in the main text assuming that log brain volume evolves like a Brownian motion. The estimated slopes ($b$) of PCs 6, 7 and 8 are statistically significant in the sense that they are two standard errors above zero. The PCs are in units of centroid size multiplied by 100.
Figure S2: The effect of relative brain size on recent beak-shape microevolution. The plot shows the magnitude of the deviance of beak-shape variables from the macroevolutionary prediction plotted against relative brain size for 645 bird species. The fitted line and parameter estimates are from Model 3 in the main text. The slope ($b$) of PC 7 is statistically significant in the sense of being two standard errors above zero, and the slope of PC2 is significantly negative. The PCs are in units of centroid size multiplied by 100.
Figure S3: The effect of body size on recent beak-shape evolution. The plot shows the magnitude of the deviance from clade mean of beak-shape variables from the macroevolutionary prediction plotted against (predicted) log body mass for 645 bird species. The fitted line and parameter estimates are from Model 3 in the main text. The slopes \((b)\) of PCs 2 and 4 to 8 are statistically significant in the sense that they are two standard errors above zero. The PCs are in units of centroid size multiplied by 100.
**Supplementary simulations**

To investigate precision and bias of our statistical methods (models 1-3) we used parametric simulation of the assumed evolutionary process. First, we performed parametric bootstrapping on the estimated relationship between beak-shape variables and absolute brain size using the bird phylogeny described in the main text. The parametric bootstrapping of the fitted models was done using the function `evolvability::rate_gls_boot()`, using 1000 bootstrap samples. Second, we did parametric simulations where we simulate the processes thousand times using simulated phylogenies of either 100 or 400 species. Simulations were done in R using `ape::rtree()` to generate phylogenies, `evolvability::simulate_rate()` to simulate data according to the evolutionary model and the phylogeny, and `evolvability::rate_gls()` to estimate the parameters from the simulated data.

The text below also gives some information on how to analyze and simulate data using the `evolvability` package. More information can be found in the vignette `Analyzing_rates_of_evolution` embedded in the package.

**Model 1**

The first model is a model where the predictor $x$ evolves according to a Brownian-motion process, while $y$ follows a Brownian motion with variance that is linear in $x$. The evolutionary model is

\[ dy = \sqrt{a + bx} dW_1 \]
\[ dx = \sigma dW_2 \]

where $a$ is the evolutionary rate of $y$ at $x = 0$, $b$ is the linear effect of $x$ on the evolutionary rate, and $dW_i$ are two uncorrelated white-noise processes.

**Simulations on the bird phylogeny**

The analysis presented in Table S1 shows that the GLS estimates of the parameters and their associated standard errors were in reasonable concordance with the parametric bootstrap means and standard deviations. The close correspondence between parameter estimates and bootstrap means indicate that there is little bias in the estimation. The $R^2$ of the data ($R^2_{GLS}$) was, however, smaller than the $R^2$ of the simulated data ($R^2_{Boot}$) indicating some misfit of the assumed evolutionary model to the data. For the last three PCs (PC6-8), we got a substantial fraction of negative roots (i.e. negative $a + bx$) during the parametric simulations. This is not a problem for the GLS estimates, but the parametric bootstrap will no longer simulate the assumed evolutionary model, given by the equations above, as we set the diffusion variance of $y$ to zero (i.e. no evolutionary change) in these instances. We would then expect a discrepancy between the GLS estimates and the bootstrap means due to deviation from the model assumptions in the bootstrap.
Table S1. Results from analysis of the bird data presented in the paper. Parameter estimates (a and b) ± one standard error and $R^2$ from the generalized least squares model (subscript GLS) and the corresponding mean ± one standard deviation of the parametric bootstrap distribution (subscript Boot). Response variable (y) for each regression is given by “Trait” and the explanatory variable (x) was log brain size with $\sigma^2 = 1.19 \pm 0.07$. The percent of negative roots during the simulations is given by the last column.

| Trait | $a_{GLS}$ | $a_{Boot}$ | $b_{GLS}$ | $b_{Boot}$ | $R^2_{GLS}$ | $R^2_{Boot}$ | Neg. roots |
|-------|-----------|------------|-----------|------------|-------------|-------------|------------|
| PC1   | 381±98    | 379±88     | -51.6±70.7| -54.4±96.1 | 0.1%        | 0.2±0.3%    | 0.0%       |
| PC2   | 249±99    | 248±68     | 108±67    | 107±64     | 0.4%        | 0.6±0.6%    | 1.1%       |
| PC3   | 51.8±23.9 | 51.7±11.7  | 9.58±16.24| 8.67±12.61 | 0.1%        | 0.2±0.3%    | 0.0%       |
| PC4   | 36.4±16.1 | 36.6±9.6   | 12.4±11.7 | 12.2±9.4   | 0.2%        | 0.4±0.5%    | 0.2%       |
| PC5   | 18.6±8.0  | 18.4±4.2   | 3.30±5.59 | 3.15±4.86  | 0.1%        | 0.2±0.3%    | 0.0%       |
| PC6   | 9.48±3.61 | 10.0±4.4   | 10.5±2.7  | 8.43±3.40  | 2.3%        | 1.7±1.3%    | 17.8%      |
| PC7   | 3.98±1.92 | 4.13±1.65  | 4.31±1.45 | 3.53±1.38  | 1.4%        | 1.7±1.2%    | 17.5%      |
| PC8   | 3.15±1.61 | 3.12±1.08  | 2.53±1.12 | 2.20±0.99  | 0.8%        | 1.3±1.1%    | 10.6%      |

The bootstrap confirms the inference of statistical significance by using ±2x the GLS standard error. For the PCs 6 to 8 the 2.5% quantile of the bootstrap distribution did not overlap zero for the b parameter, and there was only marginal overlap for PC2 (Table S2).

According to the quantiles (Table S2) and the histograms (Figure S4), the bootstrap distributions were relatively symmetric for the parameter b, but had a long upward tail for a. The two parameters show little correlation during the bootstrap (Figure S4).

Table S2. Quantiles from bootstrap distribution of parameters a and b from analysis in Table S1.

| Trait | $a_{Boot}$ | $b_{Boot}$ |
|-------|------------|------------|
|       | 2.5% | 50% | 97.5% | 2.5% | 50% | 97.5% |
| PC1   | 254  | 364 | 585  | -255 | -54 | 128   |
| PC2   | 148  | 237 | 413  | -11  | 104 | 242   |
| PC3   | 34.5 | 49.5| 81.3 | -17.1| 7.9 | 35.5  |
| PC4   | 22.6 | 34.9| 59.9 | -4.4 | 11.5| 32.7  |
| PC5   | 12.6 | 17.6| 28.5 | -6.3 | 3.2 | 12.1  |
| PC6   | 3.0  | 9.4 | 19.8 | 3.0  | 8.0 | 16.4  |
| PC7   | 1.45 | 3.98| 7.86 | 1.24 | 3.33| 6.67  |
| PC8   | 1.43 | 3.01| 5.47 | 0.47 | 2.11| 4.40  |
Figure S4. Histograms and bivariate scatterplots for the bootstrap distribution of parameters $a$ and $b$. 
Simulations on simulated phylogenies
To further investigate the accuracy of parameter estimates in Model 1, we used simulated phylogenies.

Table S3. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate       | Bias            | Median |
|-----------|------------|----------------|-----------------|--------|
| Run 1:    |            |                |                 |        |
| a         | 1.00       | 1.01±0.22      | 0.01±0.01       | 0.99   |
| b         | 1.00       | 1.04±1.50      | 0.04±0.05       | 0.97   |
| σ²        | 0.0625     | 0.0627±0.009   | 0.0002±0.0003   | 0.0621 |
| Run 2:    |            |                |                 |        |
| a         | 1.00       | 1.00±0.22      | 0.00±0.01       | 0.97   |
| b         | 1.00       | 0.95±1.45      | -0.05±0.05      | 0.82   |
| σ²        | 0.0625     | 0.0624±0.009   | -0.0001±0.0003  | 0.0621 |

Table S4. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 400 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate       | Bias            | Median |
|-----------|------------|----------------|-----------------|--------|
| Run 1:    |            |                |                 |        |
| a         | 1.00       | 0.99±0.25      | -0.01±0.01      | 0.96   |
| b         | 1.00       | 0.99±1.46      | -0.01±0.05      | 0.92   |
| σ²        | 0.0625     | 0.0626±0.0042  | 0.0001±0.0001   | 0.0625 |
| Run 2:    |            |                |                 |        |
| a         | 1.00       | 1.00±0.33      | 0.00±0.01       | 0.93   |
| b         | 1.00       | 0.85±2.20      | -0.15±0.07      | 0.77   |
| σ²        | 0.0625     | 0.0625±0.0046  | -0.0000±0.0001  | 0.0623 |

The simulations (table S3 and S4) showed no indication of systematic bias in the parameters a, b or σ². Note, however, that the standard errors of the estimates do not seem to decrease much with sample size. This is because the uncertainty in the estimates are influenced by the phylogenetic structure. In all the simulations, the precision of the focal parameter b was low (i.e. the standard error of the estimate was high). The reason for this is that we chose to have little variation in x (i.e. low σ²) in the simulation to avoid the problem of negative roots. Increasing the variation in x increases the precision (table S5 and S6). Note that the apparent downward bias in b (Table S5 and S6) comes from the change in model assumptions when negative a + bx (negative roots) are set to zero in the simulation.
Table S5. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision. The number of negative roots in the simulations was about 10%.

| Parameter | True value | Estimate | Bias   | Median |
|-----------|------------|----------|--------|--------|
| Run 1:    |            |          |        |        |
| $a$       | 1.00       | 1.04±0.27| 0.04±0.01| 1.01   |
| $b$       | 1.00       | 0.87±0.47| -0.13±0.02| 0.83   |
| $\sigma^2$| 1.00       | 1.01±0.14| 0.01±0.00| 1.00   |
| Run 2:    |            |          |        |        |
| $a$       | 1.00       | 1.02±0.29| 0.02±0.01| 0.99   |
| $b$       | 1.00       | 0.85±0.47| -0.15±0.01| 0.80   |
| $\sigma^2$| 1.00       | 1.00±0.14| 0.00±0.00| 0.99   |

Table S6. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 400 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision. The number of negative roots in the simulations was about 13%.

| Parameter | True value | Estimate | Bias   | Median |
|-----------|------------|----------|--------|--------|
| Run 1:    |            |          |        |        |
| $a$       | 1.00       | 1.04±0.38| 0.04±0.01| 1.00   |
| $b$       | 1.00       | 0.87±0.45| -0.13±0.01| 0.81   |
| $\sigma^2$| 1.00       | 1.00±0.07| 0.00±0.00| 1.00   |
| Run 2:    |            |          |        |        |
| $a$       | 1.00       | 1.02±0.24| 0.03±0.01| 1.00   |
| $b$       | 1.00       | 0.86±0.30| -0.14±0.01| 0.83   |
| $\sigma^2$| 1.00       | 1.00±0.07| 0.00±0.00| 1.00   |

**Model 2**

The second model is a model where the predictor $x$ evolves according to a geometric Brownian motion process, which is equivalent to Brownian motion of the natural logarithm of $x$, while $y$ is following a Brownian motion with a variance that is linear in $x$. The evolutionary model is

\[
dy = \sqrt{a + bx} dW_1
\]

\[
dx = \frac{1}{2} \sigma^2 x dt + xdW_2
\]

where $a$ is the evolutionary rate of $y$ at $x = 0$, $b$ is the linear effect of $x$ on the evolutionary rate, and the $dW_i$ are two uncorrelated white-noise processes.

**Simulations on the bird phylogeny**

For Model 2, the GLS estimates were in concordance with the bootstrap estimates, indicating low degree of bias in the parameter estimation. The GLS standard errors, however, were not always in concordance with the bootstrap standard deviation (Table S6). In general, the bootstrap standard deviations were higher than the GLS standard error estimates. This is
expected as the GLS standard error for \(a\) and \(b\) are approximations and does not take the error of \(\sigma^2\) into account. Compared to Model 1, the \(R^2\) of the data (\(R^2_{\text{GLS}}\)) was closer to the \(R^2\) of the simulated data (\(R^2_{\text{Boot}}\)). Combined with higher \(R^2\), this suggest that Model 2 fits the data better than Model 1. For PC1, 6, 7 and 8 there were some negative roots (i.e. negative \(a + bx\)) during the parametric simulations. This stems from negative values of \(a\) or negative values of \(b\), since \(x\) cannot take negative values in Model 2. As for Model 1, the negative roots are not a problem for the GLS estimates, but the parametric bootstrap will no longer simulate the assumed evolutionary model, as given by the equations above. In instances of negative roots in the simulation, we set the diffusion variance of \(y\) to zero (i.e. no evolutionary change). We would then expect a discrepancy between the GLS estimates and the bootstrap means due to deviation from the model assumptions in the bootstrap. Judging from the simulation, however, negative roots coming from a negative \(a\) parameter (PC6-8) did not generate much discrepancy in the estimate, while for PC1, where the negative roots come from a negative \(b\) parameter, there was a larger discrepancy between the GLS and the bootstrap estimate of \(b\).

Table S6. Results from analysis of the bird data presented in the paper. Parameter estimates (\(a\) and \(b\)) ± one standard error and \(R^2\) from the generalized least squares model (subscript GLS) and the corresponding mean ± one standard deviation of the parametric bootstrap distribution (subscript Boot). Response variable (\(y\)) for each regression is given by “Trait” and the explanatory variable (\(x\)) was brain size with \(\sigma^2 = 1.19±0.07\). The percent of negative roots during the simulations is given by the last column.

| Trait | \(a_{\text{GLS}}\) | \(a_{\text{Boot}}\) | \(b_{\text{GLS}}\) | \(b_{\text{Boot}}\) | \(R^2_{\text{GLS}}\) | \(R^2_{\text{Boot}}\) | Neg. roots |
|-------|------------------|------------------|------------------|------------------|------------------|------------------|------------|
| PC1   | 506±132          | 479±468          | -91.5±64.8       | -69.7±64.3       | 0.3%             | 0.3±0.2%         | 3.9%       |
| PC2   | 80±104           | 64.0±77.7        | 125±54           | 132±110          | 0.8%             | 2.0±1.1%         | 0.0%       |
| PC3   | 18.4±24.7        | 14.3±16.7        | 24.8±12.9        | 26.9±22.3        | 0.6%             | 2.0±1.1%         | 0.0%       |
| PC4   | 22.3±19.5        | 21.0±21.4        | 9.7±10.0         | 10.1±8.3         | 0.1%             | 0.9±0.3%         | 0.0%       |
| PC5   | 11.7±9.28        | 11.3±11.4        | 4.99±4.72        | 5.24±4.29        | 0.2%             | 0.9±0.4%         | 0.0%       |
| PC6   | -7.54±3.67       | -8.03±6.26       | 12.4±2.1         | 12.9±10.8        | 5.1%             | 6.4±4.9%         | 30.6%      |
| PC7   | -1.73±2.05       | -2.26±1.68       | 4.18±1.17        | 4.44±3.63        | 2.0%             | 5.8±4.2%         | 18.4%      |
| PC8   | -1.14±1.54       | -1.53±1.17       | 3.15±0.86        | 3.32±2.81        | 2.1%             | 5.1±3.9%         | 15.3%      |

The bootstrap distribution of the parameter \(b\) did not overlap zero for PCs 3, 6, 7, and 8, and only marginally overlapped zero for PC2 (Table S7). There is, however, some discrepancy between inference based the bootstrap quantiles and the inference one would make by using \(\pm 2x\) the standard error of the GLS estimates. For example, for PCs 6-8 the bootstrap interval is much wider than one would expect from the GLS standard error, but this might be due to negative roots influencing bootstrap distribution.

The bootstrap distributions were not symmetric for \(a\) nor \(b\) (Table S7, Fig. S5). The bootstrap distribution of the two parameters were negatively correlated generating a tail in the opposite direction of the individual bootstrap distribution (Fig. S5).
Table S7. Quantiles from bootstrap distribution of parameters $a$ and $b$ from analysis in Table S6.

| Trait | $a_{Boot}$ | $b_{Boot}$ |
|-------|------------|------------|
|       | 2.5%  | 50%  | 97.5% | 2.5%  | 50%  | 97.5% |
| PC1   | 290   | 468   | 782   | -190  | -64  | 35    |
| PC2   | -244  | 78    | 324   | -0.4  | 110  | 393   |
| PC3   | -50.9 | 16.7  | 64.1  | 0.9   | 22.3 | 78.2  |
| PC4   | -10.8 | 21.4  | 49.5  | -4.9  | 8.3  | 36.6  |
| PC5   | -4.8  | 11.4  | 27.1  | -3.1  | 4.3  | 18.2  |
| PC6   | -32.5 | -6.3  | 5.1   | 4.0   | 10.8 | 36.3  |
| PC7   | -10.8 | -1.7  | 3.3   | 0.9   | 3.6  | 12.4  |
| PC8   | -7.65 | -1.17 | 2.66  | 0.62  | 2.81 | 9.25  |

Figure S5. Histograms and bivariate scatterplots for the bootstrap distribution of parameters $a$ and $b$. 
Simulations on simulated phylogenies

The simulations revealed that the parameter estimates can be biased for this model (Table S8 and S9). The $a$-parameter was downwardly biased, and the $b$-parameter was upwardly biased. The bias increased with increasing $\sigma^2$, when $\sigma^2 = 0.5$ up to around 1, as in the bird data (Table S6), the bias was negligible. When $\sigma^2 = 2.0$ the bias was up to 18% for $b$. Sample size, 100 species (Table S8) vs 400 species (Table S9), did not much influence the precision (the standard error) of the $a$ and $b$ estimates. However, the precision of $b$ increased when the true value of $a$ was small compared to the true value of $b$ (Table S10).

**Table S8.** True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate | Bias | Median |
|-----------|------------|----------|------|--------|
| Run 1:    |            |          |      |        |
| $a$       | 1.00       | 0.66±1.81| -0.34±0.06 | 0.97 |
| $b$       | 1.00       | 1.18±1.18| 0.18±0.04  | 0.89 |
| $\sigma^2$| 2.00       | 2.01±0.28| 0.01±0.01  | 2.00 |
| Run 2:    |            |          |      |        |
| $a$       | 1.00       | 0.67±1.82| -0.33±0.06 | 1.04 |
| $b$       | 1.00       | 1.17±1.23| 0.17±0.04  | 0.82 |
| $\sigma^2$| 2.00       | 2.00±0.29| 0.00±0.01  | 2.00 |
| Run 3:    |            |          |      |        |
| $a$       | 1.00       | 0.98±1.33| -0.02±0.04 | 1.15 |
| $b$       | 1.00       | 1.00±1.09| 0.00±0.03  | 0.79 |
| $\sigma^2$| 1.00       | 0.99±0.14| -0.01±0.00 | 0.99 |
| Run 4:    |            |          |      |        |
| $a$       | 1.00       | 0.98±1.37| -0.02±0.04 | 1.12 |
| $b$       | 1.00       | 0.99±1.13| -0.01±0.04 | 0.75 |
| $\sigma^2$| 1.00       | 1.00±0.14| 0.00±0.00  | 1.00 |
| Run 5:    |            |          |      |        |
| $a$       | 1.00       | 1.08±1.32| 0.08±0.04  | 1.13 |
| $b$       | 1.00       | 0.94±1.18| -0.06±0.04 | 0.75 |
| $\sigma^2$| 0.50       | 0.50±0.07| 0.00±0.00  | 0.50 |
| Run 6:    |            |          |      |        |
| $a$       | 1.00       | 1.04±1.34| 0.04±0.04  | 1.06 |
| $b$       | 1.00       | 0.97±1.22| -0.03±0.04 | 0.81 |
| $\sigma^2$| 0.50       | 0.50±0.07| -0.00±0.00 | 0.49 |
Table 9. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 400 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate | Bias     | Median |
|-----------|------------|----------|----------|--------|
| Run 1:    |            |          |          |        |
| $a$       | 1.00       | 0.81±1.31| -0.19±0.04| 0.95   |
| $b$       | 1.00       | 1.08±0.83| 0.08±0.03| 0.90   |
| $\sigma^2$| 2.00       | 2.01±0.15| 0.01±0.00| 2.00   |
| Run 2:    |            |          |          |        |
| $a$       | 1.00       | 0.68±1.68| -0.33±0.05| 0.85   |
| $b$       | 1.00       | 1.15±1.06| 0.16±0.03| 0.91   |
| $\sigma^2$| 2.00       | 2.00±0.14| 0.00±0.00| 2.00   |
| Run 3:    |            |          |          |        |
| $a$       | 1.00       | 0.93±1.65| -0.07±0.05| 0.95   |
| $b$       | 1.00       | 1.00±1.27| 0.00±0.04| 0.82   |
| $\sigma^2$| 1.00       | 1.00±0.07| 0.00±0.00| 1.00   |
| Run 4:    |            |          |          |        |
| $a$       | 1.00       | 0.98±0.80| -0.02±0.03| 1.07   |
| $b$       | 1.00       | 1.01±0.67| 0.01±0.02| 0.89   |
| $\sigma^2$| 1.00       | 1.00±0.07| 0.00±0.00| 1.00   |
| Run 5:    |            |          |          |        |
| $a$       | 1.00       | 1.03±1.67| 0.03±0.05| 1.09   |
| $b$       | 1.00       | 0.96±1.45| -0.04±0.04| 0.79   |
| $\sigma^2$| 1.00       | 1.01±0.14| 0.01±0.00| 1.00   |
| Run 6:    |            |          |          |        |
| $a$       | 0.01       | 0.09±0.96| -0.10±0.03| 0.07   |
| $b$       | 1.00       | 1.07±0.84| 0.07±0.03| 0.87   |
| $\sigma^2$| 1.00       | 1.01±0.14| 0.01±0.00| 1.00   |

Table S10. True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate | Bias     | Median |
|-----------|------------|----------|----------|--------|
| Run 1:    |            |          |          |        |
| $a$       | 0.01       | -0.09±0.96| -0.10±0.03| 0.07   |
| $b$       | 1.00       | 1.07±0.84| 0.07±0.03| 0.87   |
| $\sigma^2$| 1.00       | 1.01±0.14| 0.01±0.00| 1.00   |
| Run 2:    |            |          |          |        |
| $a$       | 0.01       | -0.10±0.95| -0.11±0.03| 0.07   |
| $b$       | 1.00       | 1.07±0.85| 0.07±0.03| 0.85   |
| $\sigma^2$| 1.00       | 1.01±0.15| 0.01±0.00| 1.01   |
Model 3

This model differs from the two above in that the focus is on the recent evolutionary process. The species mean trait vector is modeled as \( \mathbf{y} = \mathbf{y}_{macro} + \mathbf{y}_{micro} \) where \( \mathbf{y}_{macro} \) is the result of a Brownian motion process with rate equal to \( \sigma_y \), and \( \mathbf{y}_{micro} \) is a microevolutionary deviation. The variance of \( \mathbf{y}_{micro} \) depend on a predictor variable according to:

\[
\mathbf{y}_{micro} \sim N(0, a \mathbf{I} + \text{diag}(b \mathbf{x}))
\]

where \( a \) and \( b \) are parameter, \( \mathbf{I} \) is the identity matrix and \( \mathbf{x} \) is a vector of species-specific predictor variables. The \( \text{diag} \)-function applied to a vector yields a diagonal matrix with the vector along the diagonal.

Simulations on the bird phylogeny

For Model 3 both the GLS estimates and their standard errors were in reasonable accordance with the bootstrap mean of the bootstrap distribution and its standard deviation (Table S12). The \( R^2 \) of the data (\( R_{GLS}^2 \)) were systematically lower than that for the simulated data (\( R_{Boot}^2 \)), suggesting that there is some discrepancy between the assumed model and the data.

Table S12. Results from analysis of the bird data presented in the paper. Parameter estimates \((a \text{ and } b) \pm \) one standard error and \( R^2 \) from the generalized least squares model (subscript GLS) and the corresponding mean \( \pm \) one standard deviation of the parametric bootstrap distribution (subscript Boot). Response variable \((y) \) for each regression is given by “Trait” and the explanatory variable \((x) \) was log brain size with \( \text{Var}(x) = 1.14 \pm 0.0636 \).

| Trait | \( a_{GLS} \) | \( a_{Boot} \) | \( b_{GLS} \) | \( b_{Boot} \) | \( R_{GLS}^2 \) (%) | \( R_{Boot}^2 \) (%) |
|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| PC1   | 15.0±4.4       | 15.3±5.5       | 1.59±4.20      | 1.63±2.91      | 0.0%           | 0.2±0.3%       |
| PC2   | 48.2±6.4       | 47.7±6.9       | 12.4±5.8       | 12.5±4.0       | 0.7%           | 1.8±1.1%       |
| PC3   | 6.10±0.92      | 6.05±1.21      | 1.49±0.86      | 1.49±0.68      | 0.5%           | 0.9±0.8%       |
| PC4   | 2.78±0.58      | 2.80±0.59      | 2.27±0.48      | 1.69±0.28      | 3.4%           | 5.4±1.5%       |
| PC5   | 1.78±0.31      | 1.76±0.36      | 0.831±0.270    | 0.809±0.177    | 1.5%           | 3.2±1.3%       |
| PC6   | 1.78±0.23      | 1.68±0.23      | 0.933±0.177    | 0.847±0.113    | 4.2%           | 8.3±2.0%       |
| PC7   | 0.593±0.080    | 0.573±0.099    | 0.314±0.064    | 0.292±0.048    | 3.6%           | 5.9±1.8%       |
| PC8   | 0.777±0.087    | 0.751±0.096    | 0.322±0.068    | 0.300±0.047    | 3.4%           | 6.1±1.9%       |

In accordance with the inference made from the GLS standard errors, the 95% confidence interval of the \( b \) parameter did not overlap zero for any PC but PC1 (Table S13).

Table S13. Quantiles from bootstrap distribution of parameters \( a \) and \( b \) from analysis in Table S12.

| Trait | \( a_{Boot} \) | \( b_{Boot} \) |
|-------|----------------|----------------|
|       | 2.5% | 50% | 97.5% | 2.5% | 50% | 97.5% |
| PC1   | 5.3  | 15.1 | 27.1  | -4.09 | 1.69 | 7.39  |
| PC2   | 34.6 | 47.7 | 61.6  | 4.8  | 12.4 | 20.2  |
| PC3   | 3.76 | 6.07 | 8.41  | 0.18 | 1.48 | 2.82  |
| PC4   | 1.67 | 2.80 | 3.94  | 1.11 | 1.69 | 2.24  |
| PC5   | 1.04 | 1.76 | 2.44  | 0.45 | 0.82 | 1.14  |
| PC6   | 1.24 | 1.67 | 2.13  | 0.63 | 0.85 | 1.07  |
| PC7   | 0.383| 0.570| 0.771 | 0.196| 0.293| 0.383 |
| PC8   | 0.571| 0.751| 0.949 | 0.211| 0.298| 0.389 |
The bootstrap distribution was symmetric for both the $a$ and the $b$ parameter, and there was little correlation between the two parameters (Fig. S6).

Figure S6. Histograms and bivariate scatterplots for the bootstrap distribution of parameters $a$ and $b$. 

15
Simulations on simulated phylogenies

In the simulations, the standard deviation of the elements of $x$ was set to 0.25 and the same $x$ was used throughout each simulation run.

Table S14. True and estimated values of parameters with associated biases ($\pm$ one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate   | Bias      | Median |
|-----------|------------|------------|-----------|--------|
| Run 1:    |            |            |           |        |
| $a$       | 1.00       | 1.04±0.38  | 0.03±0.01 | 1.01   |
| $b$       | 1.00       | 0.96±0.85  | -0.04±0.03| 1.02   |
| $\sigma^2_y$ | 1.00     | 0.99±0.50  | -0.01±0.02| 0.96   |
| Run 2:    |            |            |           |        |
| $a$       | 1.00       | 1.01±0.36  | 0.01±0.01 | 1.02   |
| $b$       | 1.00       | 0.96±0.92  | -0.04±0.03| 1.00   |
| $\sigma^2_y$ | 1.00     | 1.01±0.50  | 0.01±0.02 | 0.98   |

Table S15. True and estimated values of parameters with associated biases ($\pm$ one standard error) using simulated phylogenies of 400 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate   | Bias      | Median |
|-----------|------------|------------|-----------|--------|
| Run 1:    |            |            |           |        |
| $a$       | 1.00       | 0.99±0.14  | -0.01±0.00| 0.99   |
| $b$       | 1.00       | 1.00±0.41  | -0.00±0.01| 0.99   |
| $\sigma^2_y$ | 1.00     | 1.01±0.27  | 0.01±0.01 | 0.99   |
| Run 2:    |            |            |           |        |
| $a$       | 1.00       | 0.98±0.14  | -0.02±0.00| 0.98   |
| $b$       | 1.00       | 0.99±0.39  | -0.01±0.01| 0.99   |
| $\sigma^2_y$ | 1.00     | 1.02±0.25  | 0.02±0.01 | 1.01   |

Across the simulations, there did not seem to be any evidence for systematic bias in $a$ or $b$ (Table S14 and S15). Within simulation, however, the bias was larger than 2x its standard error in one instance. This suggest that the bias may depend on the particularities of the phylogeny. The maximum estimated bias was 4%, however. There was a large effect of sample size on the bootstrap standard error of the estimates (i.e. the precision), and the power to test whether $b$ was different from zero was adequate in the simulations with 400 species.

In the above simulations for Model 3 we estimated the parameters in a model where the mean vector of the species was calculated separately for each species by not including that species. This should be more important for small compared to large phylogenies for getting better estimates of microevolutionary deviations. Below, we estimated the parameters using the overall mean of all species in the species vector. This is implemented by setting `evolvability::rate_gls(useLFO = FALSE)`. This seem to induce a downward bias in $b$
when the sample size is low (i.e. 100 species; Table S16), but not when the sample size was higher (i.e. 400 species; Table S17).

**Table S16.** True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 100 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate | Bias | Median |
|-----------|------------|----------|------|--------|
| Run 1:    |            |          |      |        |
| $a$       | 1.00       | 1.01±0.36| 0.01±0.01| 1.00   |
| $b$       | 1.00       | 0.90±0.91| -0.10±0.04| 0.93   |
| $\sigma^2$| 1.00       | 1.01±0.52| 0.01±0.02| 0.96   |
| Run 2:    |            |          |      |        |
| $a$       | 1.00       | 1.03±0.39| 0.03±0.01| 1.04   |
| $b$       | 1.00       | 0.95±1.10| -0.05±0.03| 0.99   |
| $\sigma^2$| 1.00       | 0.98±0.54| -0.02±0.02| 0.93   |

**Table S17.** True and estimated values of parameters with associated biases (± one standard error) using simulated phylogenies of 400 species. The estimate is the mean of the bootstrap distribution, the median is shown for comparison. The standard error of the estimate is the standard deviation of the bootstrap distribution and is therefore an estimate of the precision.

| Parameter | True value | Estimate | Bias | Median |
|-----------|------------|----------|------|--------|
| Run 1:    |            |          |      |        |
| $a$       | 1.00       | 1.00±0.12| 0.00±0.00| 1.00   |
| $b$       | 1.00       | 0.96±0.36| -0.04±0.01| 0.98   |
| $\sigma^2$| 1.00       | 1.02±0.27| 0.02±0.01| 1.00   |
| Run 2:    |            |          |      |        |
| $a$       | 1.00       | 1.00±0.13| 0.00±0.00| 0.99   |
| $b$       | 1.00       | 1.00±0.39| 0.00±0.01| 1.01   |
| $\sigma^2$| 1.00       | 1.00±0.25| 0.00±0.01| 1.00   |

The algorithm of Model 3, as implemented in `evolvability::rate_gls(model = "recent_evol")`, is sometimes not converging. If this happens, the options are:

1. increase the maximum number of iterations through the argument `maxiter`;
2. provide more sensible starting values through the argument `startv`;
3. increase the convergence tolerance through the argument `tol`;
4. remove potential outliers from the data.

Note that non-convergence during the bootstrapping is not a major problem if the fraction of non-converged iterations are only a few percent. The non-converged iterations are deleted from the bootstrap distribution.