1 “Exact” method to calculate clade-level significance

As discussed in the text, we suggest the most useful method of test for clade-level significance is through null permutation. It is possible, however, to calculate an exact p-value for a clade's variance. This is by testing whether there has been a significant change in the variance of that clade's composition given the variances of the species that constitute it. The confidence interval for a variance is well-known, and for a normal distribution it is defined as $\pm\frac{(n-1)^2 s^2}{\chi^2_\alpha}$, where n is the sample size, s^2 the variance, and χ^2_α the appropriate α quantile of the χ^2 distribution for $n-1$ degrees of freedom. There is statistical evidence for a clade having higher or lower variance (phylogenetic clustering and overdispersion respectively) than expected if the sum of the variances of a clade's constituent species falls outside this confidence interval.

2 Permutation tests of evolutionary model fitting

To test our evolutionary model-fitting, we permuted species' body mass values and, 100 times, fitted the same Brownian and OU-models to the permuted data. This procedure is extremely computationally intensive, and not strictly necessary as these methods have already been extensively explored (Beaulieu *et al.* 2012), hence we performed only 100 replicate tests. By comparing the support for each model-type in the null permutations with our observed results, we can assess whether our phylogenetic data or modelling approach may be biased towards supporting certain evolutionary hypotheses. The model that best fitted the empirical data (a three-rate Brownian motion model) had, on average, a δAIC of 273.92 (table 1), and the best-supported model in the null permutations (an OU-model with a mean δAIC of 0.51; table 1) contained no across-clade variation. Thus there is strong support for variation in the rate of body mass evolution in the squirrel and cavi clades identified using independent ecological information. Note that two of the model-types failed to converge on the permuted data (table 1); for our purposes we consider a failure to converge an indication of poor model fit.

θ_0	θ_c	θ_s	σ_0	σ_c	σ_s	α_0	α_c	α_s	$\widehat{\delta AIC} \pm SE$
—	—	—	✓	✓	✓	—	—	—	273.92 ± 6.56
✓	✓	✓	✓	✓	✓	✓			3.81 ± 0.24
✓	✓	✓	✓			✓	✓	✓	n/a
✓	✓	✓	✓	✓	✓	✓	✓	✓	n/a
—	—	—	✓			—	—	—	282.72 ± 6.87
✓			✓			✓			0.51 ± 0.12
✓	✓	✓	✓			✓			2.27 ± 0.15

Table 1: **Null results of log(body mass) evolutionary modelling**. Results of 100 models fit to the same phylogenetic and clade data, but permuted trait (body mass) data, as the observed evolutionary models (table 2). Each row represents a different model; '—' is used to indicate when a parameter is not fit in a model. Rows one and four represent Brownian motion (they have σ 'rate' parameters only), and all other rows are variants of Ornstein-Uhlenbeck models [with at least one optimum and rate of return to that optimum, θ and α respectively). In subscripts of parameters, 'c' refers to the 'capi' clade, 's' to the 'squirrel' clade, and '0' to the remainder of the phylogeny. Thus the presentation is the same as in table 2, with two exceptions. First, no parameter estimates are given (they are uninformative, as the models were fit to permuted data), and instead a '✓' indicates when a parameter was fit. Secondly, δAIC and its standard error across the 100 bootstrap models is reported. As described in the text, two model-types could not be successfully fit to the data and have 'n/a' estimates of δAIC. These show that, by chance, we would expect only very weak support for the most-supported model in our analysis of the observed data (compare the first rows of this and table 2).

Bibliography

Beaulieu, Jeremy M *et al.* (2012). Modeling stabilizing selection: expanding the Ornstein–Uhlenbeck model of adaptive evolution. *Evolution* 66 (8), 2369–2383.