### SURFACE

We used the SURFACE method for detecting convergence over similar selective regimes 12. It fits Ornstein-Uhlenbeck stabilizing selection model to first identify regime shifts on branches of the tree, where a proposed regime shift would improve the AIC score of the model.  Employing such information criterion allows for balancing the trade-off between improving the log‐likelihood versus increasing the complexity. During the forward phase, a nonlinear optimization is used to find the maximum likelihood estimate for α in the OU model (α represents the rate of adaptive evolution towards a hypothetic optima – θ) from which the maximum likelihood estimates for σ2 (Brownian rate parameter) and θ (optimal trait value) are obtained. During a second, ‘backward phase’, it collapses regimes at different branches and evaluates the AIC again. This time AIC can improve when reducing the number of the parameters outweighs the potential decrease in the log-likelihood. After collapsing all pairs of identified regime shifts it accepts the set of collapses that reaches optimal AIC, indicating that collapsed regimes are convergent.

The analysis was conducted on natural logged and min-max standardised body size and the natural logged residuals from the phylogenetic regression of brain and body size (residual brain size) due to high collinearity between brain size and body size.

### < Model selection

We performed model selection using the dredge function in the MuMIn package on a set especially imputed for that purposes, using the procedure explained above (see Imputation section). The initial models were based on expectations from previous studies, but included as many interactions as possible, which were subsequently reduced. We ran each full model twice for 250 000 iterations, with burn in of the first 10000 iterations and sampling rate of 101. Convergence was verified visually and tested with Gelman-Rubin criterion (< 1.1) and effective sample size was always above 2000.  
Subsequently, one of the chains was ‘dredged’ and the preferred candidate model was chosen based on the requirement to be within Δ <= 3 of the best candidate model ranked by DIC, and containing the most exhaustive set of variables, and least interactions. >