Appendix II: Supplementary analyses

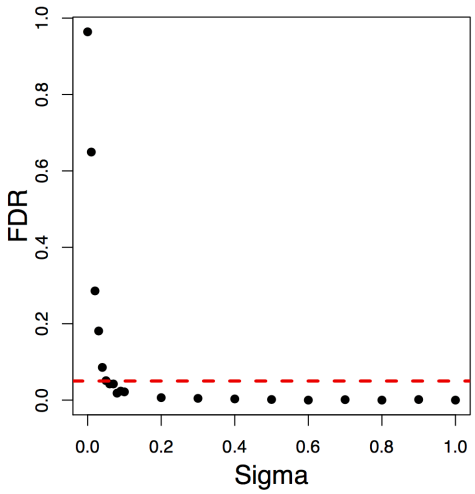
S1 False discovery rate and power of the Monte Carlo simulation method to detect conserved genes

To assess the false-discovery rate (FDR) of the Monte Carlo simulation method when testing for (highly) conserved genes, we perform a simulation study in the following way. We simulated gene expression datasets under a Brownian motion model along the given 5-species *Heliconius* phylogeny. For this simulation, we chose sigma values (i.e. the rate of evolution) of 0.0, 0.01, …, 0.09, 0.1, 0.2, …, 1.0. These values were chosen to resemble our empirically estimated sigma values, which ranged between 0 and 3.5, and to highlight the performance of the method for values of sigma close to 0. For each value of sigma, we simulated 1,000 datasets. Then, we analyzed the simulated datasets as described in the main text. In summary, we computed the posterior probability distribution of sigma for each dataset and assessed if the posterior mean of sigma was larger than 95% of computed sigma values when the data were simulated with sigma=0.

Appendix figure S1 shows the results of the simulation study. Although the FDR for a sigma of 0 is represented as 1.0, this is actually not a false-discovery as the true mode of evolution was highly conserved. The closer the true value of sigma was to 0, the higher the false-discovery rate was. This is not surprising as our approach tested for significant differences of sigma from 0, and if the true sigma was very close to 0 then there was not enough information in the simulated data to reject the null hypothesis (i.e. sigma = 0). When sigma was equal or larger than 0.1 we find a false-discovery rate of below 0.05 (red dashed line), which represents the common nominal false-discovery rate. Thus, the likelihood that our Monte Carlo simulation approach for assessing conserved genes was reporting a false positive was very low for larger values of the rate of evolution. This simulation demonstrates

that our Monte Carlo simulation method only detects genes that are conserved if the true rate of evolution was either 0 or very small.

We omit an additional analysis to assess the power of our Monte Carlo simulation method because the power is directly visible from the false-discovery rate. For the case sigma=0, we observe that our Monte Carlo simulation method correctly classifies all simulated datasets as being conserved (Appendix figure S1).



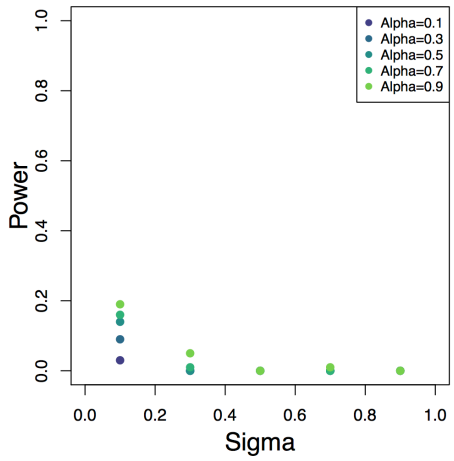
Appendix figure S1. Simulation study to assess the false-discovery rate (FDR) of our Monte Carlo simulation method for assessing if a gene was conserved. Each dot represents the frequency of classifying a dataset as conserved for the 1,000 simulations under a given value for sigma (the rate of evolution). The red dashed line shows the nominal FDR of 0.05. The results show that if the true rate of evolution was 0.1 or larger, then almost no datasets were classified as conserved.

S2 Power of the Monte Carlo simulation method to detect conserved genes when the true mode of evolution was stabilizing selection

Similar to the previous simulation study, we assessed the power of our Monte Carlo simulation method to classify a gene as conserved when the true mode of evolution was stabilizing selection. We simulated datasets as described above with the main difference that we now simulated the datasets using an Ornstein-Uhlenbeck process. We chose values for

sigma between 0.1, 0.3, …, 0.9 and values for alpha (the rate of attraction/selection) between 0.1, 0.3, …, 0.9. For each of the pairwise combinations of sigma and alpha values we simulated 1,000 gene expression datasets.

We observed that the Monte Carlo simulation method has low power to classify a gene as being conserved the higher the rate of evolution of the gene (Figure S2). These results confirm the previous simulation study under the Brownian motion model. However, the power was slightly elevated when the rate of evolution was low (e.g. sigma=0.1) and the rate of attraction was high. This is expected because a process with little evolution and strong attraction is expected to virtually remain constant (i.e. no change). Thus, datasets in these situations are correctly classified as being conserved.



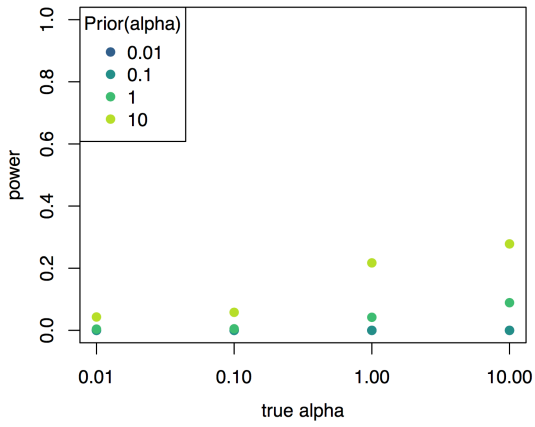
Appendix figure S2. Simulation study to assess the power of our Monte Carlo simulation method for classifying if a gene was conserved. Each dot represents the frequency of classifying a dataset as conserved for the 1,000 simulations under a given value for sigma (the rate of evolution) and alpha (the rate of attraction).

S3 Power of Bayes factors to detect genes evolving under stabilizing selection

To assess the power of the Bayes factor method to detect genes evolving under stabilizing selection, we performed a simulation study as follows. We simulated gene expression datasets for the given 5-species *Heliconius* phylogeny under an Ornstein-

Uhlenbeck process. We chose values for sigma (the rate of evolution) between 0.01, 0.1, 1.0, 10.0 and values for alpha (the rate of attraction/selection) between 0.01, 0.1, 1.0, 10.0. For each combination of sigma and alpha we simulated 1,000 datasets. Then, for each dataset we simulated Bayes factors between a Brownian motion model and an Ornstein-Uhlenbeck process model. Then we assessed if we can reject the simpler model (BM) in favor of the more complex model (OU) with a 2ln(BF) threshold of 6.

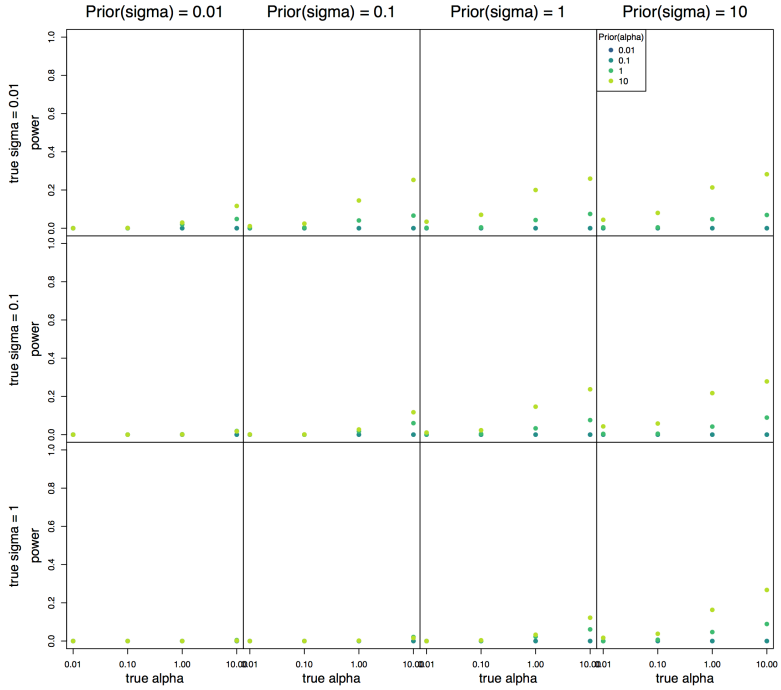
Our results of this simulation study show that Bayes factors seem to have very low power to detect that the true model was in fact an Ornstein-Uhlenbeck process. This lack of power is most likely due to the small dataset size of only 5 species. Bayes factors are calculated as the ratio of the marginal probability of the data given the first model over the marginal probability of the data given the second model. Since the information in the data is comparably low, the marginal probability of the data under the two models is very similar and thus there is little power to distinguish between the models.



Appendix figure S3. Power to detect stabilizing selection using Bayes factors (example when the mean prior on sigma was 10; best case scenario). Here we simulated 1000 datasets on phylogenies of 5 taxa with a rate of evolution sigma=0.1 and varying rate of stabilizing selection, alpha = {0.01, 0.1, 1.0, 10.0}. We also show the impact of the prior distribution on alpha, where we chose different prior means of the exponential distribution similar to the range of true parameter values (prior mean = {0.01, 0.1, 1.0, 10.0}).

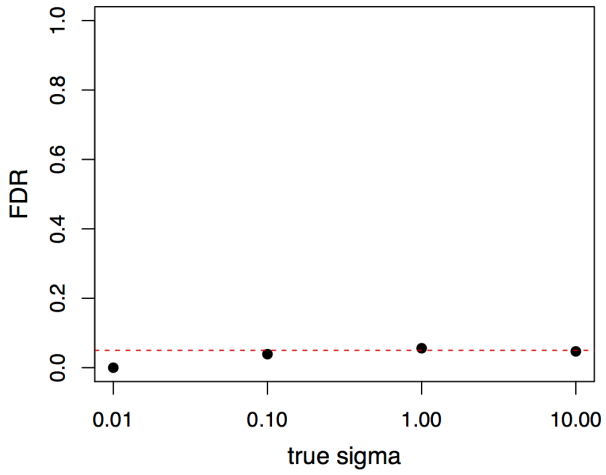
S4 False discovery rate to detect branch-specific shifts in gene-expression evolution

We simulated gene-expression datasets under the null model (Brownian motion without rates shift) to assess the false-discovery rate of our method to identify branch-specific rate shifts. Thus, we simulated a 1000 gene-expression dataset, each using a Brownian motion process. We repeated the simulation for different values of sigma = {0.01, 0.1, 1, 10}. Then, we estimated marginal likelihoods for the Brownian motion without rate shift, and marginal likelihoods for Brownian motion with a shift at any of the five terminal branches. Finally, we recorded the frequency of how often we rejected the null model in favor of one of the five Brownian motion shift models with a Bayes factor larger than 3.



Appendix figure S4. Power to detect stabilizing selection using Bayes factors from the full range of analyses (all combinations of true sigma and sigma priors). Here we simulated 1000 datasets on phylogenies of 5 taxa with a rate of evolution sigma = {0.01, 0.1, 1.0, 10.0} and rate of stabilizing selection, alpha = {0.01, 0.1, 1.0, 10.0}. We also show the impact of the prior distribution on sigma and alpha, where we chose different prior means of the exponential distribution similar to the range of true parameter values (prior mean = {0.01, 0.1, 1.0, 10.0}).

Appendix figure S5 shows that we obtain a nominal false-discovery rate of 5%. Interestingly, we obtain the lowest false-discovery rate in simulations when the true rate of evolution, sigma, was smallest. This is most likely due to the fact that the simulated gene expression levels did not vary much, and thus no rate shift was necessary to explain the data. For larger values of sigma it can happen by chance that the simulated values of gene expression vary a lot at the tips, and thus a Brownian motion model with a branch-specific shift was supported.



Appendix figure S5. Simulation study to assess the false-discovery rate of our branch-specific models. We simulated 1000 gene-expression datasets using a constant rate Brownian motion along the empirical phylogeny for different values of sigma. The false-discovery rate (FDR) shows the frequency of dataset for which we falsely identified a shift using the method described in the main text. The dashed red line shows the nominal FDR of 5%.