

Supplementary Material for “Estimating the effect of competition on trait evolution using maximum likelihood inference”

Appendix S1 & S2

Figure S1-S8

1 Supplementary Appendix 1

Considering that n lineages are interacting at time t , each trait i evolves following the equation :

$$dz_i(t) = z_i(t + dt) - z_i(t) = S \left(\left(\frac{1}{n_i} \sum_{l=1}^n A_{i,l} z_l(t) \right) - z_i(t) \right) dt + \sigma dB_i(t)$$

Where $A_{i,l}$ is equal to 1 if lineages i and l are sympatric, and to 0 otherwise, $n_i = \sum_{l=1}^n A_{i,l}$ is the total number of lineages in sympatry with lineage i , and $B_i(t)$ represents standard Brownian motion.

Here, we present the derivation of Equations 3a and 3b from the main text. To make the derivation easier to follow, we drop the dependence on time t , replacing $z_i(t)$ with z_i and $B_i(t)$ with B_i .

First, applying the Itô formula to these stochastic processes gives us :

$$\begin{aligned} d(z_i z_j) &= z_i dz_j + z_j dz_i + d \langle z_i, z_j \rangle \\ &= S \left(\left(\frac{1}{n_j} \sum_{l=1}^n A_{j,l} z_l z_i \right) - z_j z_i \right) dt + \sigma z_i dB_j \\ &\quad + S \left(\left(\frac{1}{n_i} \sum_{k=1}^n A_{i,k} z_k z_j \right) - z_j z_i \right) dt + \sigma z_j dB_i \\ &\quad + \sigma^2 1_{i=j} dt \end{aligned}$$

where $1_{i=j}$ equals one if $i = j$ and zero otherwise.

Taking this expectation, it follows that :

$$\begin{aligned} \frac{d}{dt} \mathbb{E}(z_i z_j) &= S \left(\left(\frac{1}{n_j} \sum_{l=1}^n A_{j,l} \mathbb{E}(z_l z_i) \right) - \mathbb{E}(z_j z_i) \right) dt \\ &\quad + S \left(\left(\frac{1}{n_i} \sum_{k=1}^n A_{i,k} \mathbb{E}(z_k z_j) \right) - \mathbb{E}(z_j z_i) \right) dt \\ &\quad + \sigma^2 1_{i=j} dt \end{aligned}$$

Moreover, we get :

$$\frac{d}{dt} \mathbb{E}(z_i) = S \left(\left(\frac{1}{n_i} \sum_{l=1}^n A_{i,l} \mathbb{E}(z_l) \right) - \mathbb{E}(z_i) \right) dt$$

which leads to :

$$\begin{aligned} \frac{d}{dt} (\mathbb{E}(z_i) \mathbb{E}(z_j)) &= \mathbb{E}(z_j) \frac{d}{dt} \mathbb{E}(z_i) + \mathbb{E}(z_i) \frac{d}{dt} \mathbb{E}(z_j) \\ &= S \left(\left(\frac{1}{n_i} \sum_{k=1}^n A_{i,k} \mathbb{E}(z_k) \mathbb{E}(z_j) \right) - \mathbb{E}(z_i) \mathbb{E}(z_j) \right) dt \\ &\quad + S \left(\left(\frac{1}{n_j} \sum_{l=1}^n A_{j,l} \mathbb{E}(z_l) \mathbb{E}(z_i) \right) - \mathbb{E}(z_j) \mathbb{E}(z_i) \right) dt \end{aligned}$$

Taking together these different parts gives us the ODE satisfied by all covariances (denoted $v_{i,j} = \text{Cov}(z_i, z_j)$) :

$$\begin{aligned} \frac{d}{dt} v_{i,j} &= \frac{d}{dt} (\mathbb{E}(z_i z_j) - \mathbb{E}(z_i) \mathbb{E}(z_j)) \\ &= -2S v_{i,j} + \frac{S}{n_i} \sum_{k=1}^n A_{i,k} v_{k,j} + \frac{S}{n_j} \sum_{l=1}^n A_{j,l} v_{l,i} + \sigma^2 1_{i=j} \end{aligned} \tag{1}$$

Using these derivations, the variance terms ($i = j$) are calculated using:

$$\frac{d}{dt}v_{i,i} = -\frac{2S(n_i - 1)}{n_i}v_{i,i} + \frac{2S}{n_i} \sum_{\substack{l=1 \\ l \neq i}}^n A_{i,l}v_{l,i} + \sigma^2 \quad (2)$$

The covariance terms ($i \neq j$) are calculated using:

$$\frac{d}{dt}v_{i,j} = -S \left(\frac{n_i - 1}{n_i} + \frac{n_j - 1}{n_j} \right) v_{i,j} + \frac{S}{n_i} \sum_{\substack{k=1 \\ k \neq i}}^n A_{i,k}v_{k,j} + \frac{S}{n_j} \sum_{\substack{l=1 \\ l \neq j}}^n A_{j,l}v_{l,i} \quad (3)$$

In the case where lineages i and j are in sympatry, this formula simplifies to:

$$\frac{d}{dt}v_{i,j} = -\frac{2S(n_i - 1)}{n_i}v_{i,j} + \frac{S}{n_i} \left(\sum_{\substack{k=1 \\ k \neq i}}^n A_{i,k}v_{k,j} + \sum_{\substack{l=1 \\ l \neq j}}^n A_{j,l}v_{l,i} \right) \quad (4)$$

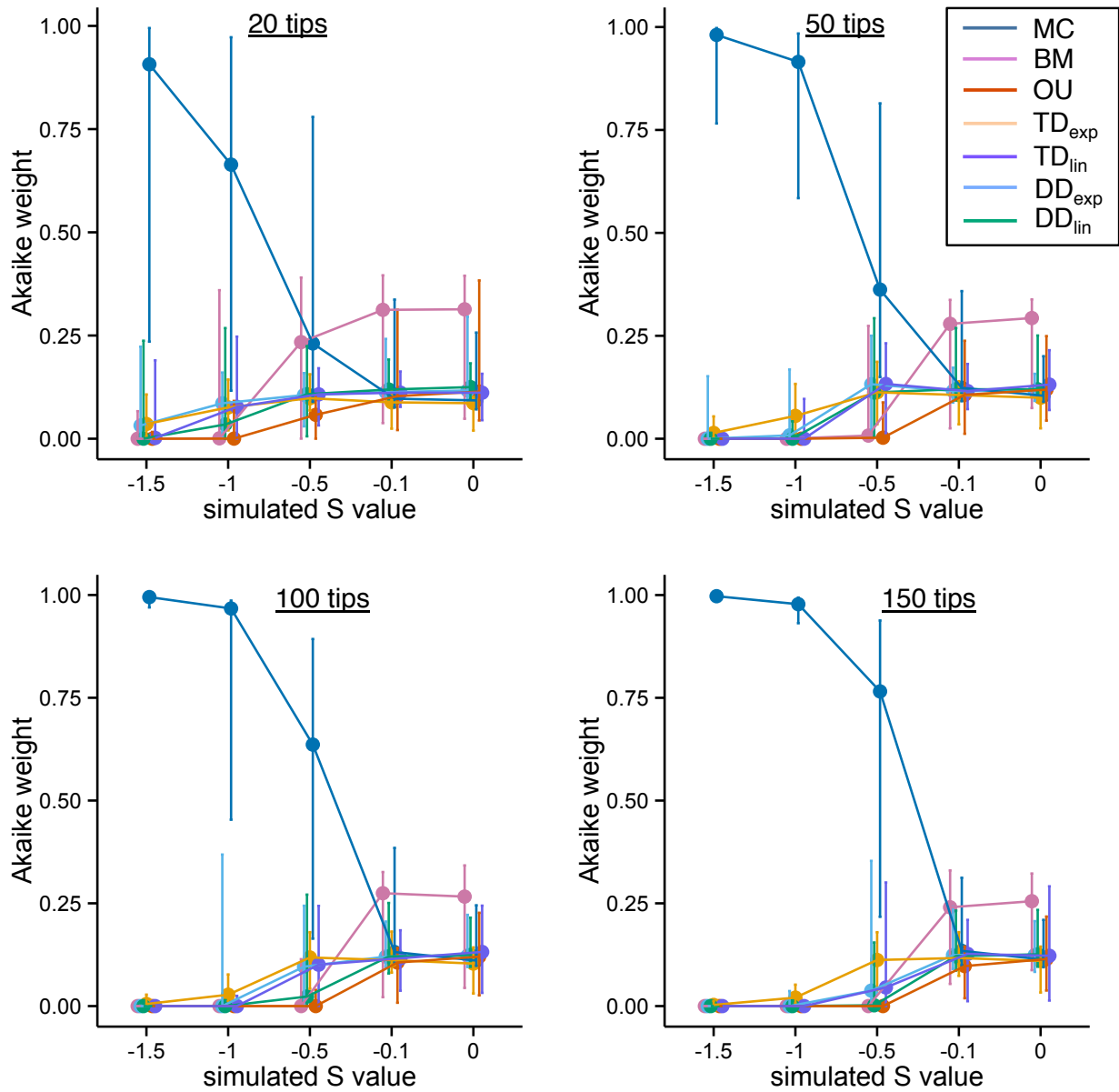
To solve the ODEs for the variance and covariance terms from the root to the tip, we begin by fixing the variance v_0 for the process at the root to 0. At each speciation event, the starting value for both the variance of each of the new lineages and the covariance between the two new lineages is the variance of the immediate ancestor at the time of the speciation event, and the starting value for the covariance between the new lineages and any other persisting lineage is set to the value of the covariance between the persisting lineage and the ancestor of the new lineages at the time of speciation.

Supplementary Appendix 2: Estimating the effect of extinction on parameter estimation for the matching competition and density-dependent models.

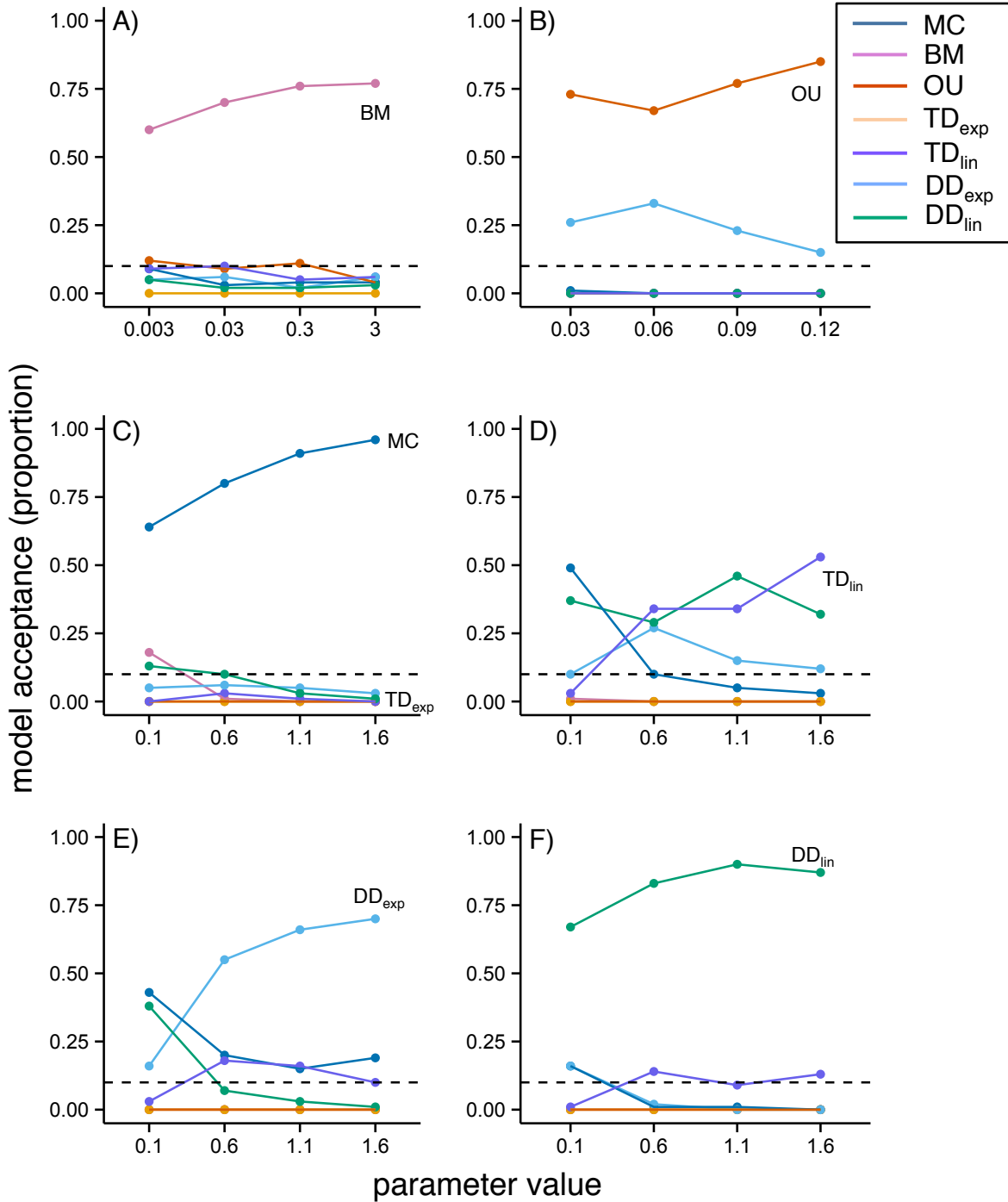
Given that the matching competition and diversity-dependent models take into account the number of interacting lineages, extinction may affect our ability to recover true parameter values. To estimate the impact of extinction, we simulated 100 trees with 100 extant species, varying the extinction fraction (μ : $\lambda = 0.2, 0.4, 0.6$, and 0.8). As above, we recursively simulated traits using the matching competition model with $\sigma^2 = 0.05$ and $S = -1.5, -1, -0.5, -0.1$, or 0 , and the linear and exponential diversity-dependent models with starting rates of $\sigma^2 = 0.6$ and ending rates of $\sigma^2 = 0.01$. We then estimated the maximum likelihood parameter estimates for the generating models by fitting the models to the trait values for extant species and the tree with extinct lineages removed. In the case of the matching competition model, because many simulated birth-death trees with high extinction rates have substantially older root ages, the simulated trait datasets for some trees had very large variances. For these biologically unrealistic trait datasets (i.e., variance in trait values $\geq 1 \times 10^8$), ML does not yield reliable parameter estimates, so we removed them from further analyses (the sample size of included simulations is reported in Fig. S5, S6).

Parameter estimates are quite robust to extinction under the matching competition model (Fig. S5, S6), and much more so than under both diversity-dependent models (Fig. S7, S8). Under the matching competition model, the maximum likelihood optimization returns reliable estimates of S and σ^2 values used to simulate datasets on trees with extinct lineages (Fig. S5, S6), although the estimates become much less reliable with larger extinction fractions, likely because simulations under the matching competition model were unbounded, resulting in trait datasets with biologically unrealistic variances. Under both diversity-dependent models, the magnitude of both the slope and σ^2 parameter values are increasingly underestimated with increasing extinction fractions (Fig. S7, S8).

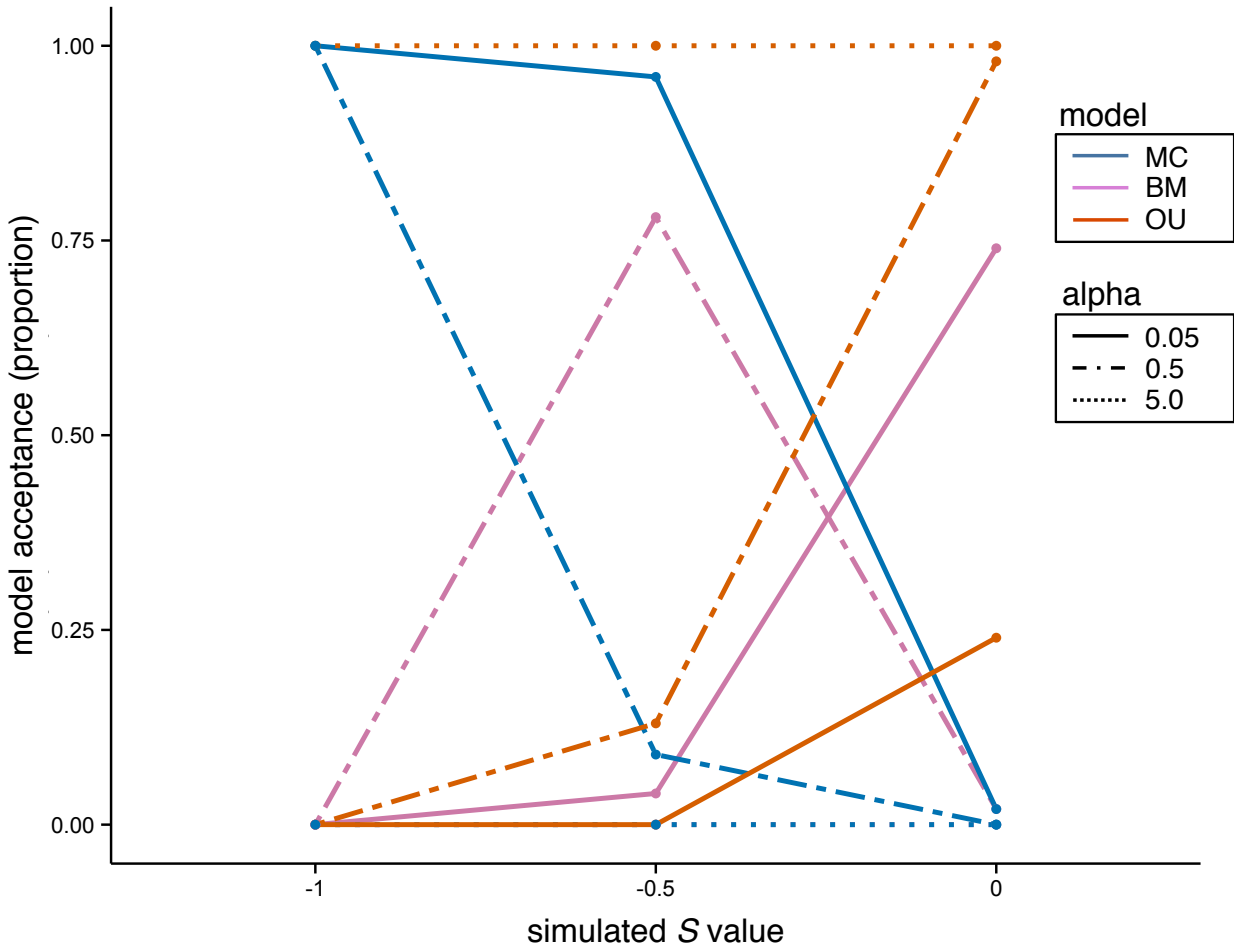
Supplementary Figure 1. As tree size and/or the degree of competition (S) increases, model selection becomes more reliable. Comparison of Akaike weights (median & 90% CIs) for NH, BM, OU, and EB models when simulated under various levels of competition ($S = -1.5, -1, -0.5, -0.1, \text{ and } 0$) for trees with 20, 50, 100, and 150 tips.



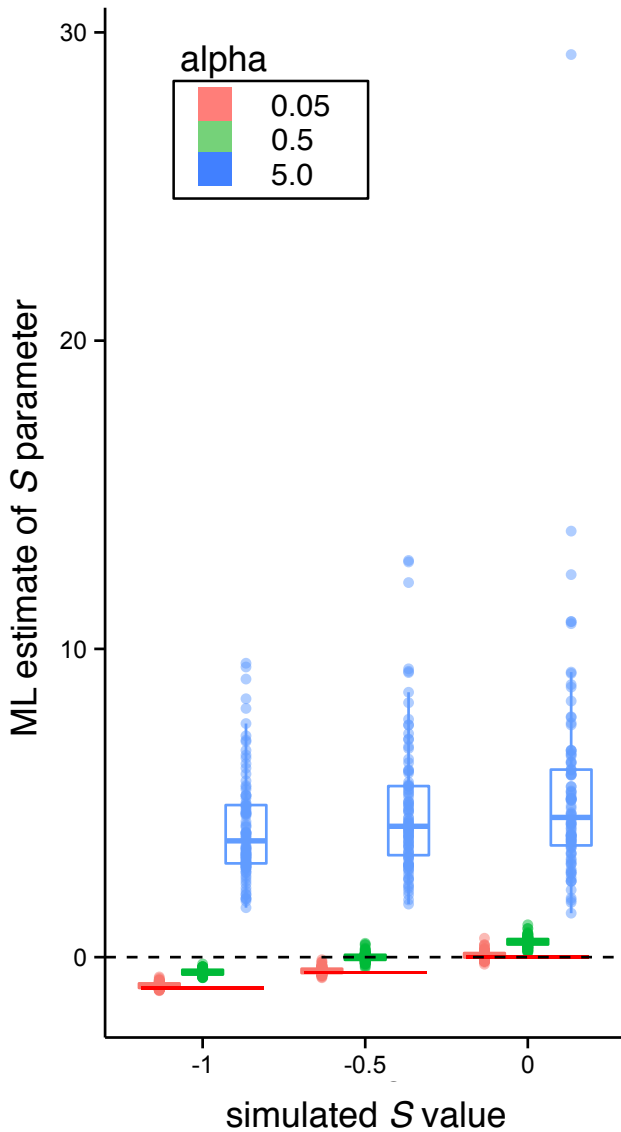
Supplementary Figure 2. Identifiability simulation results for the matching competition model as a function of varying parameter values of the generating models. Parameter values are (A) σ^2 for BM, (B) α for OU (σ^2 was fixed at 0.3), and the σ^2 value at the root for (C) TD_{exp} , (D) TD_{lin} , (E) DD_{exp} , and (F) DD_{lin} (for C-F, σ^2 at the present was fixed at 0.01).



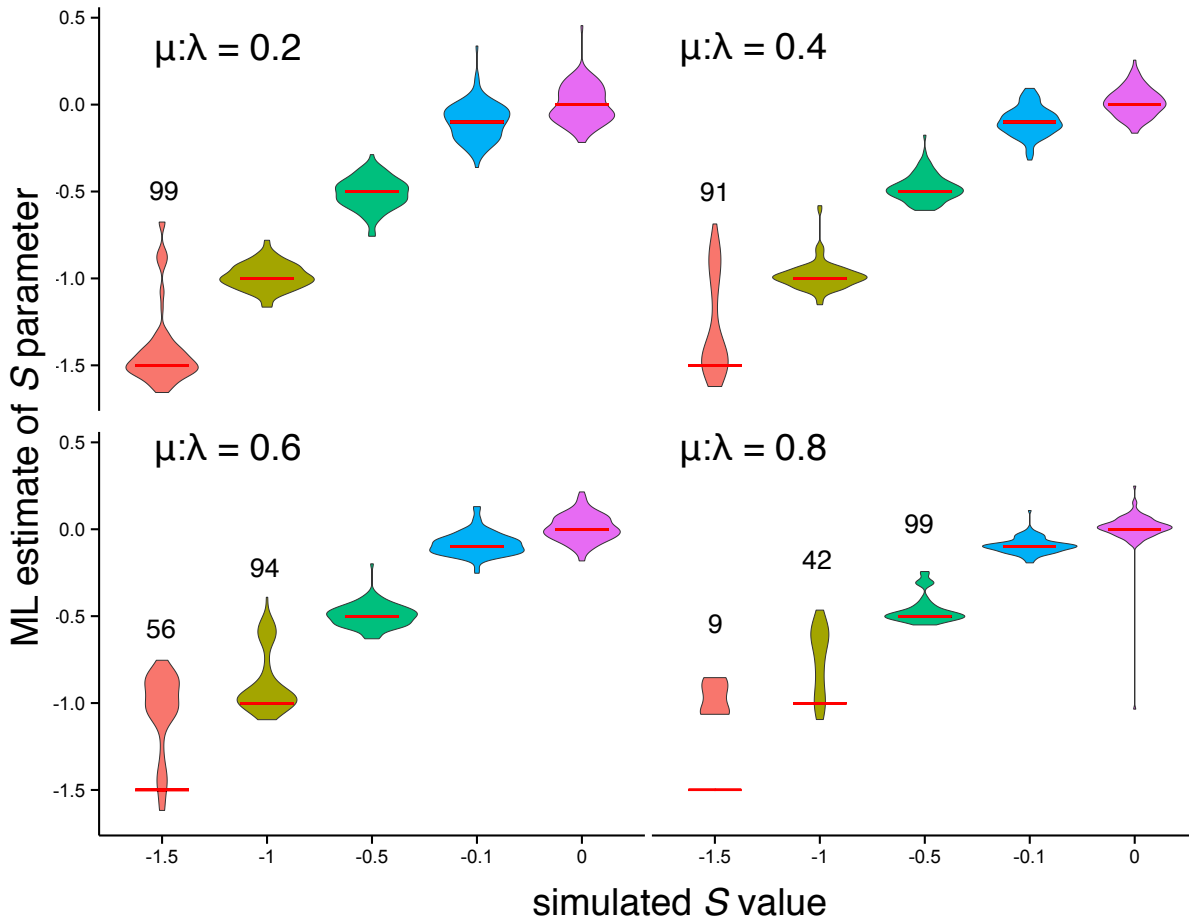
Supplementary Figure 3. The effect of incorporating stabilizing selection into trait evolution on model selection. For datasets generated under the matching competition model with stabilizing selection included, as the ratio of the strength competition (S) to the strength of selection toward an optimum (α) varies, so does the model preferred by model selection.



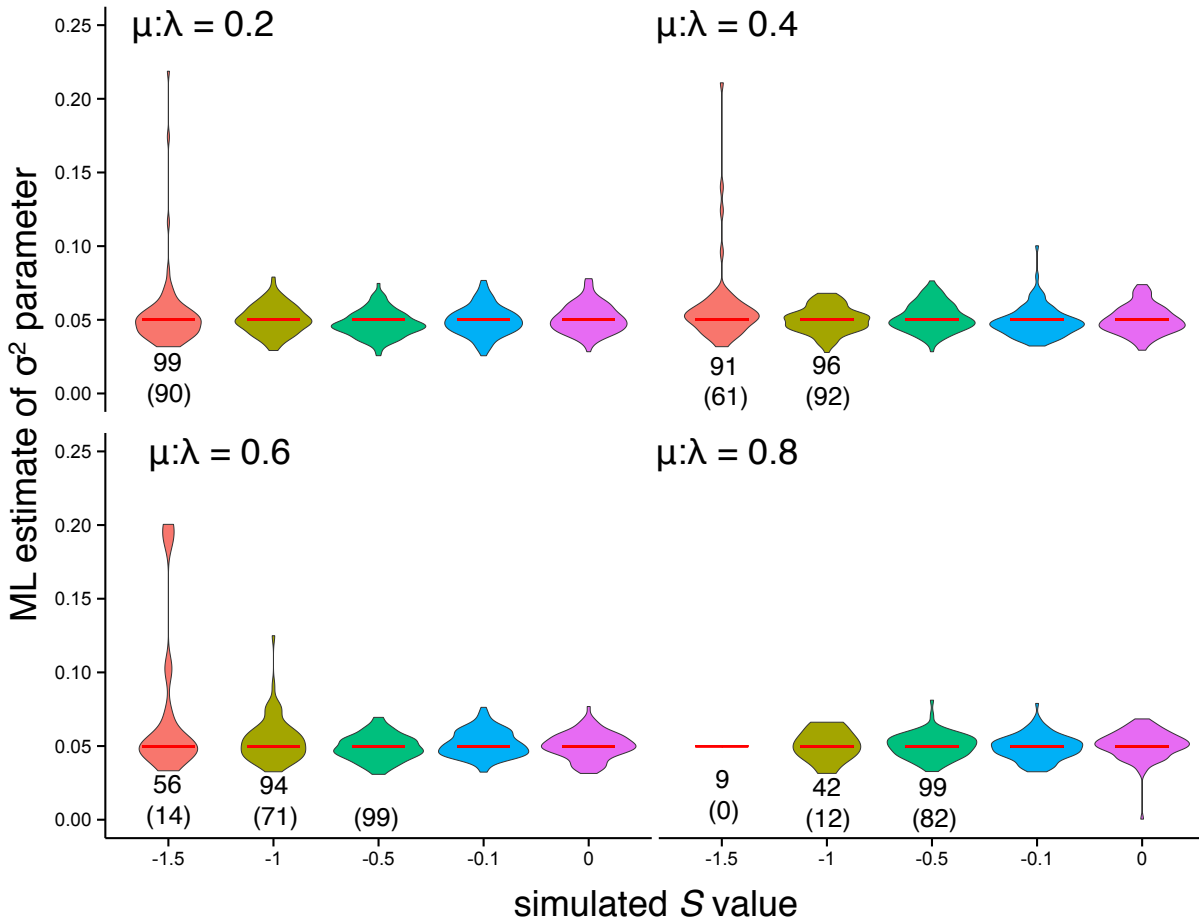
Supplementary Figure 4. The effect of incorporating stabilizing selection into trait evolution on parameter estimation. As the strength of stabilizing selection increases (i.e., as α increases), maximum likelihood under the matching competition model underestimates the true S value used to simulate datasets. Positive S values represent selection toward, rather than away, from the clade mean and are thus expected when the ratio of α to S is large. The horizontal red line represents the simulated S value, and the dashed horizontal line represents $S = 0$.



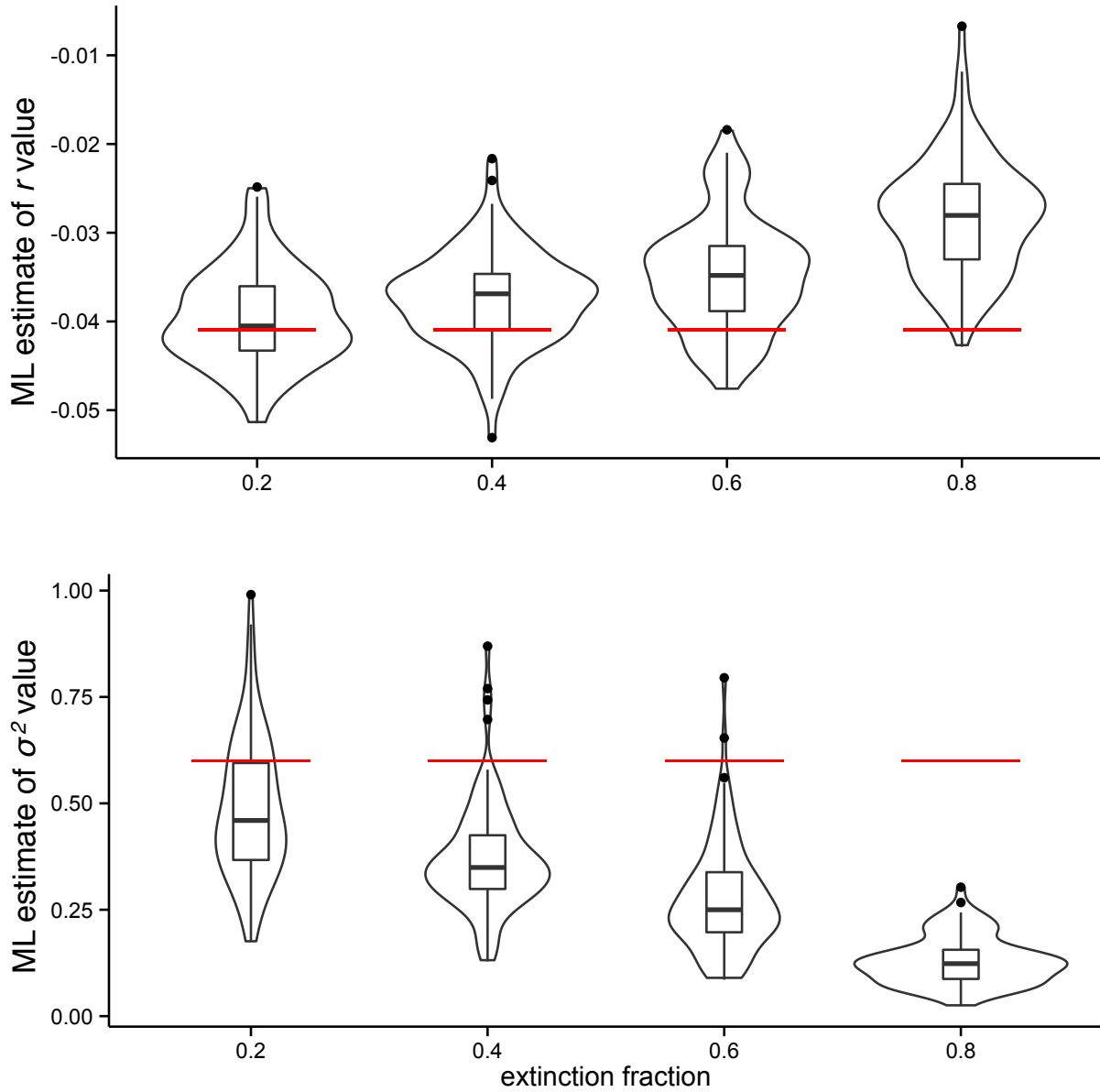
Supplementary Figure 5. Simulation results showing the effect of varying the extinction fraction on estimation of the S parameter for the matching competition model. Red horizontal lines indicate the simulated S values, and numbers above sets of simulations indicate the sample size of included simulations under those scenarios (see main text for more details).



Supplementary Figure 6. Simulation results showing the effect of varying the extinction fraction on estimation of the σ^2 parameter for the matching competition model. Red horizontal lines indicate the simulated σ^2 value (0.05), the numbers below sets of simulations indicate the sample size of included simulations under those scenarios (see main text for more details), and the number in parentheses indicate sample size after σ^2 values > 0.25 were removed.



Supplementary Figure 7. Simulation results showing the effect of varying the extinction fraction on slope (top) and σ^2 (bottom) parameters for the exponential diversity-dependent model. Increasing extinction levels result in increasingly underestimated slope values and σ^2 parameters. Red horizontal lines indicate the simulated parameter values.



Supplementary Figure 8. Simulation results showing the effect of varying the extinction fraction on slope (top) and sigma-squared (bottom) parameters for the linear diversity-dependent model. Increasing extinction levels result in increasingly underestimated slope values and σ^2 parameters. Red horizontal lines indicate the simulated parameter values.

