Supplementary material of Inferring the effect of species interactions on trait evolution

Liang Xu[1], Sander van Doorn[1], Hanno Hildenbrandt[1]& Rampal S. Etienne[1]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, PO Box 11103, Groningen 9700 CC, The Netherlands

Model description and analysis

We considered the coevolution of quantitative trait and abundance under stabilizing selection to an environmental optimum and interspecific competition among n species in a community. These species are assumed to belong to a single trophic level such that they are competing for one type of natural resource. The distribution of natural resources or environmental factor is assumed to be an unimodal distribution for simplicity. Thus the trait optimum that is favored by the environment is fixed. We begin by defining an individual-based fitness function associated with population growth and for a given trait value. We then derive a model describing the dynamics of trait evolution and population dynamics.

Trait change in frequency distribution due to selection

The fitness of an individual with trait z is related to the change of abundance and frequency of the trait

$$N_{t+1}f_{t+1}(z,\mu+\Delta\mu,V_P+\Delta V_p)=N_tf_t(z,\mu,V_P)\omega(z) \tag{1}$$

where N_t denotes the population size of the focal species at the t-th generation and $f_t(z)$ is the probability densities of the trait at the t-th generation. We assume that the trait distribution follows a normal distribution but with a small change in mean μ and variance V_P in the next generation. Thus, we define the distribution at the t-th generation as follows

$$f_t(z,\mu,V_P) = \frac{1}{\sqrt{2\pi V_P}} e^{-\frac{1}{2}\frac{(z-\mu)^2}{V_P}} \tag{2}$$

and at the $t+1$-th generation with the slight difference in mean $\Delta\mu$ and variance ΔV_P

$$f_{t+1}(z,\mu+\Delta\mu,V_P+\Delta V_p) = \frac{1}{\sqrt{2\pi(V_P+\Delta V_P)}} e^{-\frac{1}{2}\frac{(z-\mu-\Delta\mu)^2}{V_P+\Delta V_P}} \tag{3}$$

The increase of the population size at the $t+1$th generation N_{t+1} thus can be simply expressed in terms of the mean fitness of the population (Lande 1976):

$$N_{t+1} = N_t \int f_t(z,\mu,V_P)\omega(z)dz = N_t\bar{\omega}_t \tag{4}$$

where the integration $\bar{\omega} = \int f_t(z,\mu,V_P)\omega(z)dz$ measures the mean fitness of the population at the t-th generation.

Substituting eq. 4 into eq. 1 yields the following dynamics of the frequency of trait z:

$$f_{t+1}(z,\mu+\Delta\mu,V_P+\Delta V_p) = f_t(z,\mu,V_P)\frac{\omega(z)}{\bar{\omega}_t}. \tag{5}$$

Expanding $\omega(z)$ around the mean trait μ up to the first order gives

$$\omega(z) \approx \omega(\mu) + \omega'(\mu)(z-\mu). \tag{6}$$

The mean fitness $\bar{\omega}_t$ can then be approximated by (see Fig. S1)

$$\begin{aligned}\bar{\omega}_t &= \int \omega(z) f_t(z)dz \\ &\approx \int \left(\omega(\mu) + \omega'(\mu)(z-\mu)\right) f_t(z)dz \\ &\approx \omega(\mu)\end{aligned} \tag{7}$$

Applying Taylor expansion to the left hand side of 5 with respect $\Delta\mu, \Delta V_p$ yields:

$$\Delta\mu = V_P\frac{\omega'(\mu)}{\bar{\omega}_t} \tag{8}$$

$$\Delta V_P = V_P^2\frac{\omega''(\mu)}{\omega(\mu)}. \tag{9}$$

Considering heritability $h^2 = \frac{V_G}{V_P}$ where V_G is the genetic variance, the change of the trait mean $\Delta\mu$ must be

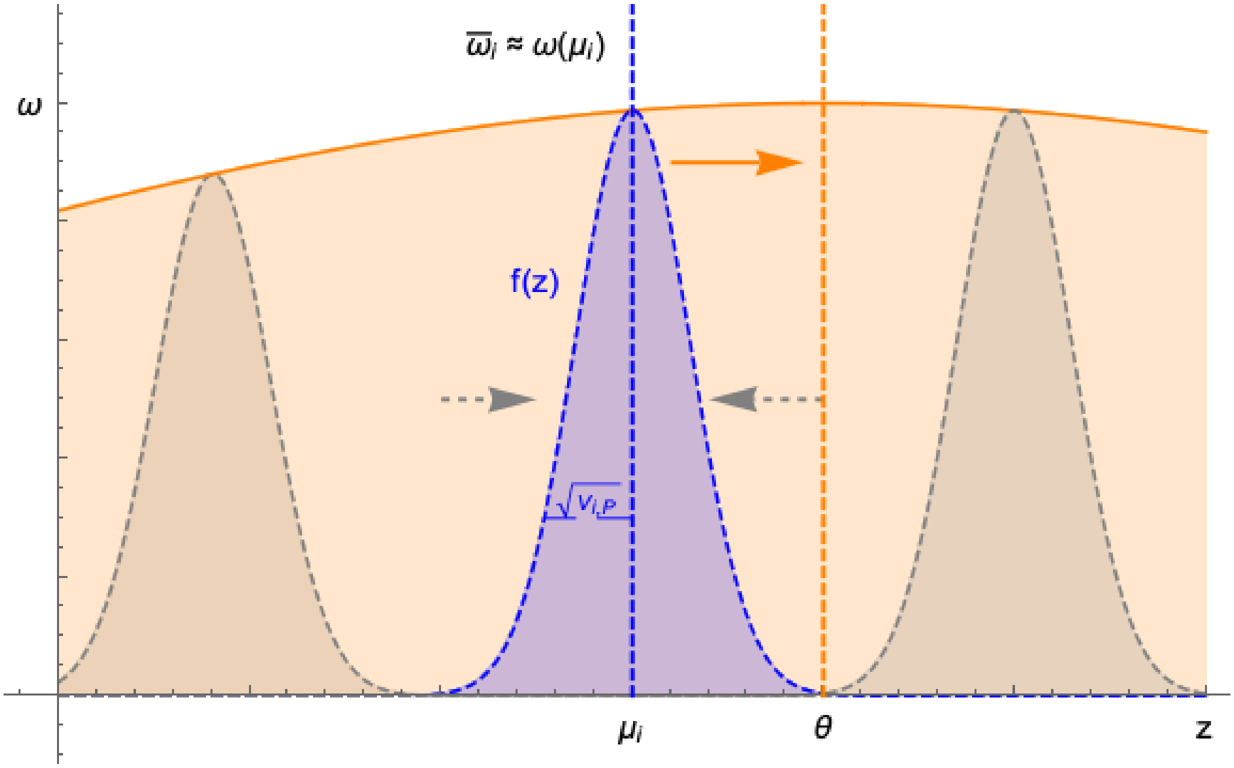

Fig. S1 – Illustration of the model derivation. The orange curve denotes the fitness landscape with a wide and flat shape due to the assumption of weak selection. The dashed curves are the trait distribution of species in which the blue dashed curve is for the focal species i with the density function $f(z)$ and the mean trait μ_i and the variance $V_{i,P}$. The trait evolution strategy of species i depends on the attraction of the stabilizing selection and the interspecific competition. The orange arrow accounts for the attraction toward the optimum θ favored by the environment while the dashed gray arrows are the competition forces from the competitors.

multiplied by h^2, which yields:

$$\Delta\mu = h^2 V_P \frac{\omega'(\mu)}{\omega(\mu)}. \tag{10}$$

Change in variance in the $t+1$-th generation

The phenotypic variance results from two factors, i.e. genetic variance (V_G) and environmental sources (V_E). The genetic variancechanges due to selection, which is formulated in the above section eq.9, due to reproduction, and due to mutation.

Change in variance due to selection

At the t-th generation, according to the dynamics Eq.9 after selection the variance changes to

$$V_{P,t,\text{after selection}} = V_{P,t,\text{before selection}} + V^2_{P,t,\text{before selection}} \frac{\omega''(\mu)}{\omega(\mu)} \tag{11}$$

$$= V_{P,t,\text{before selection}} \cdot (1 + \frac{\omega''(\mu)}{\omega(\mu)} V_{P,t,\text{before selection}}) \tag{12}$$

where $V_{P,t,\text{before selection}}$ is just $V_{P,t}$. Thus, the genetic variance $V_{G,t,\text{after selection}}$ yields:

$$V_{G,t,\text{after selection}} = h^2 V_{P,t,\text{before selection}} \cdot (1 + \frac{\omega''(\mu)}{\omega(\mu)} V_{P,t,\text{before selection}}). \tag{13}$$

Change in variance due to sexual reproduction

We can compute the trait distribution of offspring with trait value z according to its parents' traits $z - \frac{\delta}{2}, z + \frac{\delta}{2}$ that independently follow the trait distribution of the previous generation, where δ is the deviation of the parents from the offspring's trait z. For clarity, we define the trait distribution by

$$H(z, \mu, V) \sim \mathcal{N}(\mu, V)$$

where $\mathcal{N}(\mu, V)$ is a normal distribution with mean μ and variance V. The density function of the offspring's trait is therefore

$$\begin{aligned} f(z) &= \int_{-\infty}^{+\infty} H(z - \frac{\delta}{2}, \mu, V) \cdot H(z + \frac{\delta}{2}, \mu, V) d\delta \\ &= H(z, \mu, \frac{V}{2}). \end{aligned} \tag{14}$$

Eq.14 tells us that the reproduction will lose **half of the variance** in the trait distribution of offspring.

Change in variance due to mutation

Another factor that affects the trait distribution is mutation. The change of variance by mutation is given by (Kimura and Crow 1964)

$$M = \frac{2n\nu V_{G,max}}{1 + 4n\nu} \tag{15}$$

where ν is the mutation rate and $V_{G,max}$ is the maximum additive genetic variance. The expected change in the trait by mutation is 0, thus it has no contribution to the mean of the trait distribution.

Overall, due to reproduction and mutation, the additive genetic variance at the $t+1$th yields:

$$V_{G,t+1} = \frac{1}{2}V_{G,t,\text{after selection}} + \frac{2n\nu V_{G,max}}{1+4n\nu}. \tag{16}$$

Incorporating the environmental variance V_E yields:

$$\begin{aligned} V_{P,t+1} &= V_{G,t+1} + V_E && (17)\\ &= \frac{1}{2}h^2V_{P,t} + \frac{1}{2}h^2\frac{\omega''(\mu)}{\omega(\mu)}V_{P,t}^2 + \frac{2n\nu V_{G,max}}{1+4n\nu} + (1-h^2)V_{P,t} && (18)\\ &= (1-\frac{1}{2}h^2)V_{P,t} + \frac{1}{2}h^2\frac{\omega''(\mu)}{\omega(\mu)}V_{P,t}^2 + \frac{2n\nu V_{P,max}\cdot h^2}{1+4n\nu} && (19) \end{aligned}$$

where $V_E = V_{P,t} - V_{G,t}$ is the environmental variance which is assumed to be constant.

The general trait and population coevolution model

In summary, the trait and population dynamics are governed by the following set of equations:

$$\begin{aligned} \omega(\mu_{i,t}) &= \frac{N_{i,t+1}}{N_{i,t}} && (20)\\ \mu_{i,t+1} &= \mu_{i,t} + h^2V_{P,t}\frac{\omega'(\mu_{i,t})}{\omega(\mu_{i,t})} && (21)\\ V_{P,t+1} &= (1-\frac{1}{2}h^2)V_{P,t} + \frac{1}{2}h^2\frac{\omega''(\mu_{i,t})}{\omega(\mu_{i,t})}V_{P,t}^2 + \frac{2N_{i,t}\nu V_{P,max}\cdot h^2}{1+4N_{i,t}\nu} && (22) \end{aligned}$$

This novel model tells us that the growth of abundance for one specific species depends on the fitness of its average trait. The evolution of the average trait in response to selection is in the direction that increases the mean fitness in the population (Lande 1976). In the subsequent section, we build a model for the mean fitness function associated with stabilizing selection to an environmental optimum and competitive interactions.

Mean fitness function on species scale

We have derived a general trait and population dynamics model with dynamical variance at the species scale. To complete the model we have to specify the mean fitness function. We define the mean fitness function for species using the growth ratio of abundances in subsequent generations:

$$\begin{aligned} \omega(\mu_{i,t}) &= \frac{N_{i,t+1}}{N_{i,t}}\\ &= Re^{-\beta/\beta_0} && (23) \end{aligned}$$

where $\mu_{i,t}$ denotes the trait mean of species i at the t-th generation. The growth factor$R$ depends on the trait value and the parameter θ that can be interpreted as the optimum trait favored by abiotic stabilizing selection:

$$R = R_0 e^{-\gamma(\theta-\mu_{i,t})^2}. \tag{24}$$

where R_0 is the optimal growth factor. Here, γ determines the strength of stabilizing selection towards the optimum. In Eq. 23 , β quantifies the intensity of competition. Assuming a Gaussion competition kernel, we define β as

$$\beta = \sum_{j=1}^{n} (e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} N_{j,t}). \tag{25}$$

Eq.25 tells us that the strength of competition between two species depends on the similarity in their mean traits weighted by their abundances. Thus, we have fitness function in the t-th generation

$$\omega(\mu_{i,t}) = R_0 e^{-\gamma(\theta-\mu_{i,t})^2} \cdot e^{-\frac{\beta}{\beta_0}}. \tag{26}$$

The parameter β_0 in Eq.23 and in Eq. 26 is a coefficient that can be interpreted as the carrying capacity of the community.

Finally, we incorporate demographic stochasticity in the population dynamics by drawing species abundances from a zero-truncated Poisson distribution (denoted by$\mathbf{Pois}(\cdot)$) with the mean determined by the population dynamics. The zero-truncation ensures that the population cannot go extinct when present in the phylogeny. Only when extinction noccurs in the phylogeny, we remove the species.

Inserting this fitness function in model Eq.20-22 and adding demographic stochasticity $\eta_{i,t}$ yields

$$N_{i,t+1} \sim \mathbf{Pois}\left(N_{i,t} R_0 e^{-\gamma(\theta-\mu_{i,t})^2} \cdot e^{-\beta/\beta_0}\right) \tag{27}$$

$$\mu_{i,t+1} = \mu_{i,t} + h^2 V_{i,t}\left(2\gamma(\theta-\mu_{i,t}) + \frac{2\alpha}{\beta_0}\sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} N_{j,t}\right) + \eta_{i,t} \tag{28}$$

$$\begin{aligned} V_{i,t+1} = {} & (1-\frac{1}{2}h^2)V_{i,t} + \frac{2N_{i,t}\nu V_{P,max}}{1+4N_{i,t}\nu}\cdot h^2 \\ & + \frac{1}{2}h^2 V_{i,t}^2 \Bigg[2\gamma(-1+2\gamma\theta^2) - \beta_0^{-1}\sum_j \left(4\alpha^2(\mu_{i,t}-\mu_{j,t})^2 - 2\alpha\right) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} N_{j,t} \\ & - \left(2\gamma(2\theta-\mu_{i,t}) + 2\alpha\beta_0^{-1}\sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} N_{j,t}\right) \\ & \cdot \left(2\gamma\mu_{i,t} - 2\alpha\beta_0^{-1}\sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} N_{j,t}\right)\Bigg]. \end{aligned} \tag{29}$$

The demographic stochasticity follows a normal distribution with mean 0 and a population-related variance. It is defined as

$$\eta_{i,t} \sim N(0, \frac{1}{2}\frac{V_{i,t}}{N_{e,i,t}})$$

where $V_{i,t}$ is still the phenotypic variance, $N_{e,i,t}$ is the effective population size which for simplicity is assumed to be $N_{i,t}$ in our case. This random variable describes the relationship between the variance and the population size. For large population size, the trait distribution will have small variance.

Two variants of the trait-population coevolution model

To answer the question whether population dynamics really matters in the trait evolution, we developed two additional models for comparison. One model is to simply remove the population dynamics and abundance weights in the competition kernel, yielding:

$$\mu_{i,t+1} = \mu_{i,t} + h^2 V_{i,t} \left(2\gamma(\theta - \mu_{i,t}) + 2\alpha \sum_j (\mu_{i,t} - \mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} \right) + \eta_{i,t} \tag{30}$$

$$\begin{aligned} V_{i,t+1} = {} & (1 - \frac{1}{2}h^2)V_{i,t} + \frac{2N_{i,t}\nu V_{max}}{1+4N_{i,t}\nu} \cdot h^2 + \\ & \frac{1}{2}h^2 V_{i,t}^2 \Bigg[2\gamma(-1+2\gamma\theta^2) - \sum_j \left(4\alpha^2(\mu_{i,t}-\mu_{j,t})^2 - 2\alpha\right) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} \\ & - \left(2\gamma(2\theta - \mu_{i,t}) + 2\alpha \sum_j (\mu_{i,t} - \mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} \right) \\ & \cdot \left(2\gamma\mu_{i,t} - 2\alpha \sum_j (\mu_{i,t} - \mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} \right) \Bigg] . \end{aligned} \tag{31}$$

The competition kernel of this model is similar to Drury et al. (2017)'s model in the sense that when the traits of two competitors are very different, repulsion will be small due to competition being avoided. But it differs from Drury et al. (2017)'s model in that when the traits of the two competitors are similar, Drury et al.'s model just assumes strong repulsion, whereas our model shows a weak directional repulsion which emerges as a result of modelling the population dynamics. Such a weak directional repulsion for similar trait values is realistic because if the trait distributions of two populations largely overlap, there is no strong tendency to evolve either way. We refer to this model as the Unweighted competition (UWC) model.

The other model we consider weighs competition by the total metabolic rate and hence we refer to it as the Metabolism-weighted competition (MWC) model. Here we assume that it is not so much the abundance that affects competitive strength, but the species' energy consumption.The basal metabolic rate (Brody and Procter 1932; Brody 1945; Kleiber 1947) scales with body mass and body mass scales with body length in whales (Lockyer 1976), so the basal metabolic rate scales with body length as

$$B_{i,t,\text{per capita}} = B_0 \mu_{i,t}^{9/4}. \tag{32}$$

The total metabolic rate of species i is therefore

$$B_{i,t} = B_{i,t,\text{per capita}} N_{i,t}. \tag{33}$$

Thus, the trait-metabolism coevolution model yields:

$$
\begin{aligned}
N_{i,t+1} &\sim \mathbf{Pois}\left(N_{i,t} R_0 e^{-\gamma(\theta-\mu_{i,t})^2} \cdot e^{-1/\beta_0 \left(\sum_j e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} B_{j,t} \right)} \right) && (34) \\
\mu_{i,t+1} &= \mu_{i,t} + h^2 V_{i,t} \left(2\gamma(\theta-\mu_{i,t}) + \frac{2\alpha}{\beta_0} \sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} B_{j,t} \right) + \eta_{i,t} && (35) \\
V_{i,t+1} &= (1-\frac{1}{2}h^2)V_{i,t} + \frac{2N_{i,t}\nu V_{max}}{1+4N_{i,t}\nu} \cdot h^2 && (36) \\
&\quad + \frac{1}{2}h^2 V_{i,t}^2 \Bigg[2\gamma(-1+2\gamma\theta^2) - \beta_0^{-1} \sum_j \left(4\alpha^2(\mu_{i,t}-\mu_{j,t})^2 - 2\alpha \right) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} B_{j,t} \\
&\quad - \left(2\gamma(2\theta-\mu_{i,t}) + 2\alpha\beta_0^{-1} \sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} B_{j,t} \right) \\
&\quad \cdot \left(2\gamma\mu_{i,t} - 2\alpha\beta_0^{-1} \sum_j (\mu_{i,t}-\mu_{j,t}) e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2} B_{j,t} \right) \Bigg].
\end{aligned}
$$

We used an ABC-SMC algorithm to compare the three models on the baleen whales data. We estimated 5 free variables, i.e. $\gamma, \alpha, \nu, V_m, \theta$.

Parameter inference and model selection using the ABC-SMC algorithm

The complexity of our model precludes analytical approaches to fit the model to data. Hence, we developed an inference framework using Approximate Bayesian Computation based on Sequential Monte Carlo (ABC-SMC). In ABC-SMC, first introduced by Toni et al. 2009, one starts with a large number of parameter sets sampled from the prior (these are called particles in the terminology of the field), which are then used to simulate data. The similarities of the simulated data to the empirical data (measured by one or more summary statistics) are then used as weights to sample parameter sets in the next generation, with some random noise added to it. After a few iterations (generations in the terminology of the field), the parameter sets will form the posterior distribution.

Specifically, the algorithm for parameter inference consists of the following steps for our model:

1. Set a particle indicator $j = 1, 2, \cdots, n$, where n is the total number of particles in each iteration for each candidate model, and a maximum T to the number of iterations.

2. Initialize the algorithm with parameter sets $\theta^{(j)}_{t,sample}$. The parameters are sampled independently from a prior distribution $\pi_k(\theta)$ of model k. We used a uniform distribution, indicating no informative prior information on the parameters.

3. Simulate data under all sampled parameter sets $\theta^{(j)}_{t,sample}$ and obtain the trait data vector $\boldsymbol{Z}^{(j)}_{t,sim}$ and trait variance $\boldsymbol{V}^{(j)}_{t,sim}$ from each simulation. Order the trait vector $\boldsymbol{Z}^{(j)}_{t,sim}$ and accordingly adjust trait variance $\boldsymbol{V}^{(j)}_{t,sim}$. Order the empirical trait vector $\boldsymbol{Z}_{obs}$ as well. (This is for the algorithm using the sorted mean trait distance (SMTD) as the summary statistic. Change it accordingly when using the phylogenetic independent contrasts only (PICs) or the unsorted mean trait distance combined with the phylogenetic independent contrasts (UMTD+PICs).)

4. Compute the goodness-of-fit (GOF) of each particle which is defined as

$$\mathrm{GOF}_j = \left(1 - \frac{d_{z^{(j)}}}{max\{d_{z^{(i)}}, i = 1, \cdots, n\}}\right) + \left(1 - \frac{d_{V^{(j)}}}{max\{d_{V^{(i)}}, i = 1, \cdots, n\}}\right)$$

   where $d_{z^{(j)}} = \left\| \boldsymbol{Z}^{(j)}_{t,sim} - \boldsymbol{Z}_{obs} \right\|, d_{V^{(j)}} = \left\| \sqrt{\boldsymbol{V}^{(j)}_{t,sim}} - \sqrt{\boldsymbol{V}_{obs}} \right\|$.

5. Calculate weights within the particles using

$$w^{(j)}_{t,k} = \begin{cases} 1, & \text{if } t = 0 \\ \frac{\pi(\theta^{(j)}_{t,sample})}{\sum_{j=1}^{n} w^{(m)}_{t-1} K_t(\theta^{(j)}_{t,sample} | \theta^{(m)}_{t-1,sample})}, & \text{if } t > 0. \end{cases}$$

6. Normalize the weights and sample parameters set $\theta^{(j)}_{t+1,sample}$ from the previous particles $\left\{\theta^{(m)}_{t,sample}\right\}$ that have the 5% largest GOF values, using weights $w^{(m)}_{t-1}$, and perturb the particles under a perturbation kernel $K_t(\theta^{(j)}_{t,sample} | \theta^{(m)}_{t-1,sample})$ equal to a Gaussian distribution with the previous sampled parameters as the mean and a weighted variance given by

$$\mathrm{var}(\theta^{(j)}_{t,sample}) = 2 \times \sum_{l \in \{m\}} \left\{ \left(\theta^{(l)}_{t-1,sample} - \bar{\theta}_{t-1,sample}\right)^2 \cdot w^{(l)}_{t-1} \right\}$$

   where $\bar{\theta}_{t-1,sample} = \sum_{l \in \{m\}} \theta^{(l)}_{t-1,sample} \cdot w^{(l)}_{t-1}$ is the weighted mean.

7. Increment t by 1 and return to step 2 until reaching the prescribed maximum number of iterations T.

Finally, we obtain a vector of particles with GOF values and the corresponding parameter estimates.

This ABC-SMC algorithm thus represents an evolutionary algorithm in which the parameter values generating the best GOF-values are selected. A similar idea can be used to perform model selection: the algorithm will then tend to select the model with the best GOF values. However, as in a real evolutionary process, such an algorithm can suffer from drift effects, by which a model can quickly "become extinct" due to stochasticity, particularly if the "population size" (i.e. the number of particles of each model) is not large enough. This may be remedied by increasing the number of particles which is computationally unfeasible. Therefore we chose to run the algorithm for each model (or each model scenario) separately and then at the end select the simulations with the 5% best GOF-values. The proportion in which a model contributes simulations to this 5% elite is a measure of the support for that model.

Finally, we sampled 1000 parameters from the posterior estimates to generate the evolutionary history for the baleen whale study.

Parameter inference and model comparison on the untransformed body length

We believe that using the log-transformed data biologically makes more sense than using the untransformed data, because it would be approximately equal to increase a certain proportion of weight for individuals at different sizes which is correctly captured by the log-transformation. However, we are still curious to see whether the results are very different when using the untransformed body length.

Our parameter inference using the untransformed data for our main model (AWC) again shows weak environmental stabilizing selection and fairly strong competition in baleen whale body size evolution. Generation time and heritability tend to have little influence on the inference of environmental stabilizing selection coefficient γ, competition coefficient α and mutation rate ν. Alternative sets of summary statistics lead to similar results in the estimation of the parameters except V_m (see Fig. S54 for SMTD, Fig. S55 for PICs and Fig. S56 for UMTD+PICs). The estimated γ is smallest, around 2.5×10^{-7}, when using PICs. The value increases when the summary statistics include absolute trait information (SMTD and UMTD+PICs), reaching 1×10^{-6} for the algorithm using SMTD only. The estimations for α and ν are relatively consistent across sets of summary statistics: α is inferred to be $2 \times 10^{-4} \sim 4 \times 10^{-4}$, more than two orders of magnitude larger than γ, with ν being estimated around 1×10^{-3}. The maximum variance by mutation V_m is consistently inferred to be 110 when using PICs and UMTD+PICs. We did not find substantial differences in the estimates under different time scaling parameters, although with SMTD as the set of summary statistics, the predicted V_m decreases when increasing the time scaling parameter and/or heritability. The estimation of θ is estimated to be around 1100 when using PICs, while with absolute trait data (SMTD and UMTD+PICs) the estimation is a bit larger, around 1300.

The parameter estimations all suggest a small environmental stabilizing selection coefficient and a large competition coefficient (see Fig. S60-S62 for the three sets of summary statistics). When using the summary statistics SMTD, the best fitted value of the optimum trait θ is close to the mean of the extant species traits (1300.2) for the AWC model and the UWC model, and around 750 for the MWC model. A similar pattern is also found in the estimation using UMTD+PICs but a large variance emerges for the MWC model. Using PICs (but not SMTD and UMTD+PICs) leads to large variance to estimations of θ in all three models.

References

Brody, S. (1945). *Bioenergetics and growth, with special reference to the efficiency complex in domestic animals.* Reinhold, New York.

Brody, S. and R. Procter (1932). *Relation between basal metabolism and mature body weight in different species of mamals.* Univ.Mo.Agr.Exp.Sta.Res.Bull.

Drury, J. P., G. F. Grether, T. Garland, H. Morlon, T. Garland, Jr., H. Morlon, T. Garland, H. Morlon, T. Garland, Jr., H. Morlon, T. Garland, H. Morlon, T. Garland Jr., and H. Morlon (2017). An Assessment of Phylogenetic Tools for Analyzing the Interplay Between Interspecific Interactions and Phenotypic Evolution. *Systematic Biology 67*(May), 413–427.

Kimura, M. and J. F. Crow (1964). The number of alleles that can be maintained in a finite population. *Genetics 49*(4), 725–738.

Kleiber, M. (1947). Body size and metabolic rate. *Physiological Reviews 27*(4).

Lande, R. (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Source: Evolution 30*(2), 314–334.

Lockyer, C. (1976). Body weights of some species of large whales. *ICES Journal of Marine Science 36*(3), 259–273.

Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface 6*(31), 187–202.

Supplementary results

Figures S2-S16. Trait trees under different parameter combinations of Scenarios 1, 2, 3 under the AWC model and the UWC model for the tree values of scaling parameters, $s = 100, 1000, 10000$. For the figures of the rest of the secenarios, they are stored in the folder "**abundance_test**".

Figures S17-S18. Trait trees under different parameter combinations of different scenarios.

Figures S19-S21. Prediction of the abundance distribution versus body length across 1000 simulations using the estimated parameters for Scenarios 1-3. For the figures of the rest of the scenarios, they are stored in the folder "**PredictionDis**".

Figures S22-S24. Predictions of the trait variance distribution across 100 simulations using the estimated parameters for Scenarios 1-3. For the figures of the rest of the scenarios, they are stored in the folder "**PredictionDis**".

Figures S25-S27. Comparison of the parameter estimates for those generating parameters for which the ratio $\frac{\sqrt{\alpha/\gamma}}{\text{Richness}}$ is in the range of $(0.5, 1.5)$. The figures of other scenarios are stored in the folder "**Estimation**".

Figures S28-S58. Comparison of the estimates among different scenarios.

Figures S59-S60. Parameter estimates under PICs and UMTD+PICs for the AWC model on the logtransformed body length.

Figures S61-S63. Parameter estimates under SMTD, PICs and UMTD+PICs for the AWC model on the untransformed body length.

Figures S64-S66. Results of the model comparison analysis on the untransformed body length.

Figures S67-S69. Results of the model comparison analysis on the logtransformed body length**.**

Figure S70. Abundance distributions of the AWC and MWC models.

Figure S71. Trait variance distributions of the AWC and MWC models.

Figures S72-74. Distribution of the goodness-of-fit of the three models using three summary statistics. The red dashed line represents the 5% best GOF-values.

Figure S75. The predicted contrasts using the estimated parameters under the three models and the true contrasts for PICs.

Figure S76. The predicted contrasts using the estimated parameters under the three models and the true contrasts for UMTD+PICs.

Phylogenetic tree data for Scenarios 1-22. They are stored in the folder "**SimTreeInfo**" and indexed accordingly.

All the data of generated trees used in the simulation study (Scenarios 1-22) are stored in the folder “**SimTreeInfo**”.

The inference results of the simulation study are stored in the folder “**SimEstResults**”

The baleen whale data and results are stored in the folder “**BaleenWhaleData&Results**”.

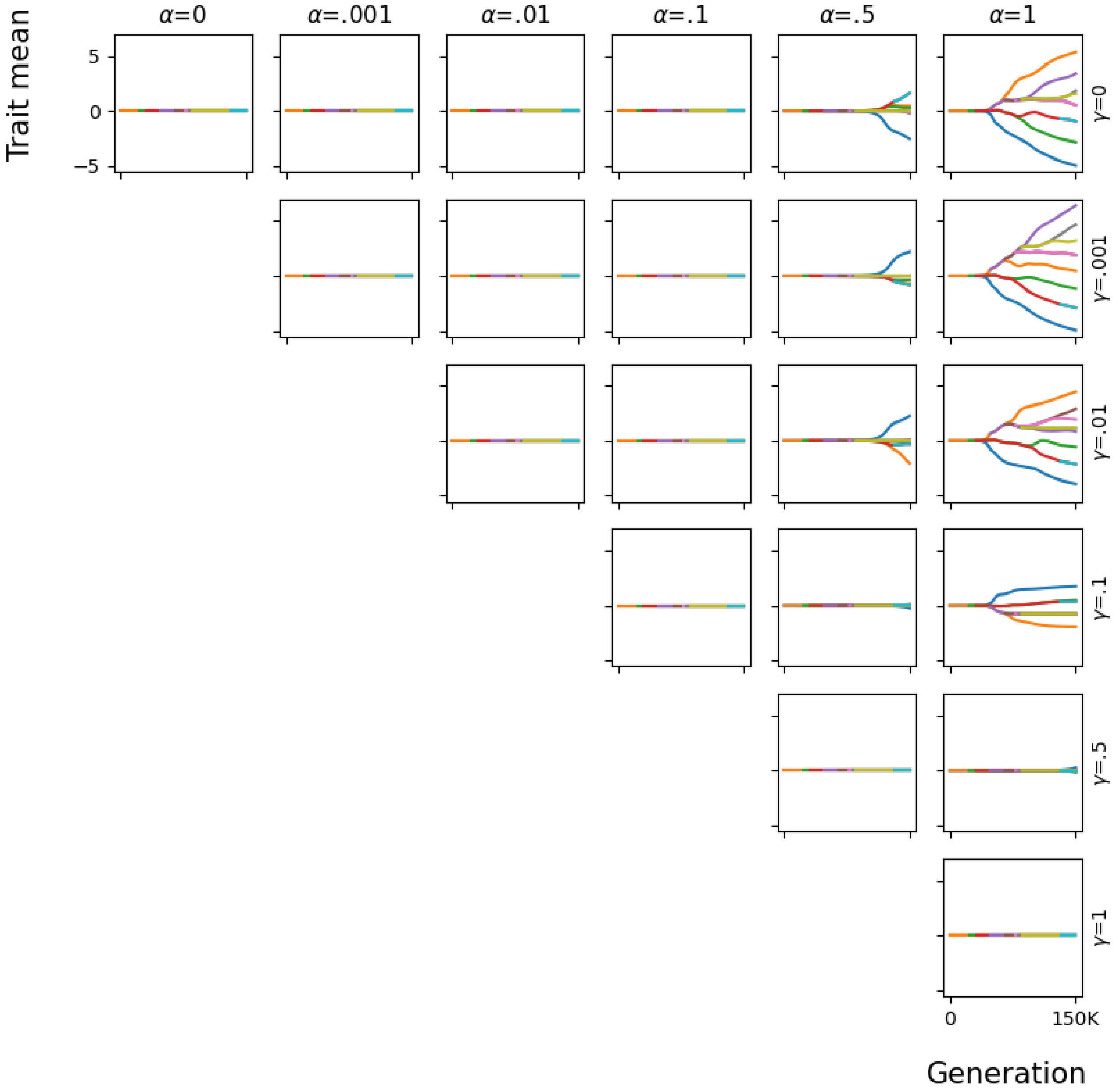

Fig. S2 – Trait trees along Tree 1 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

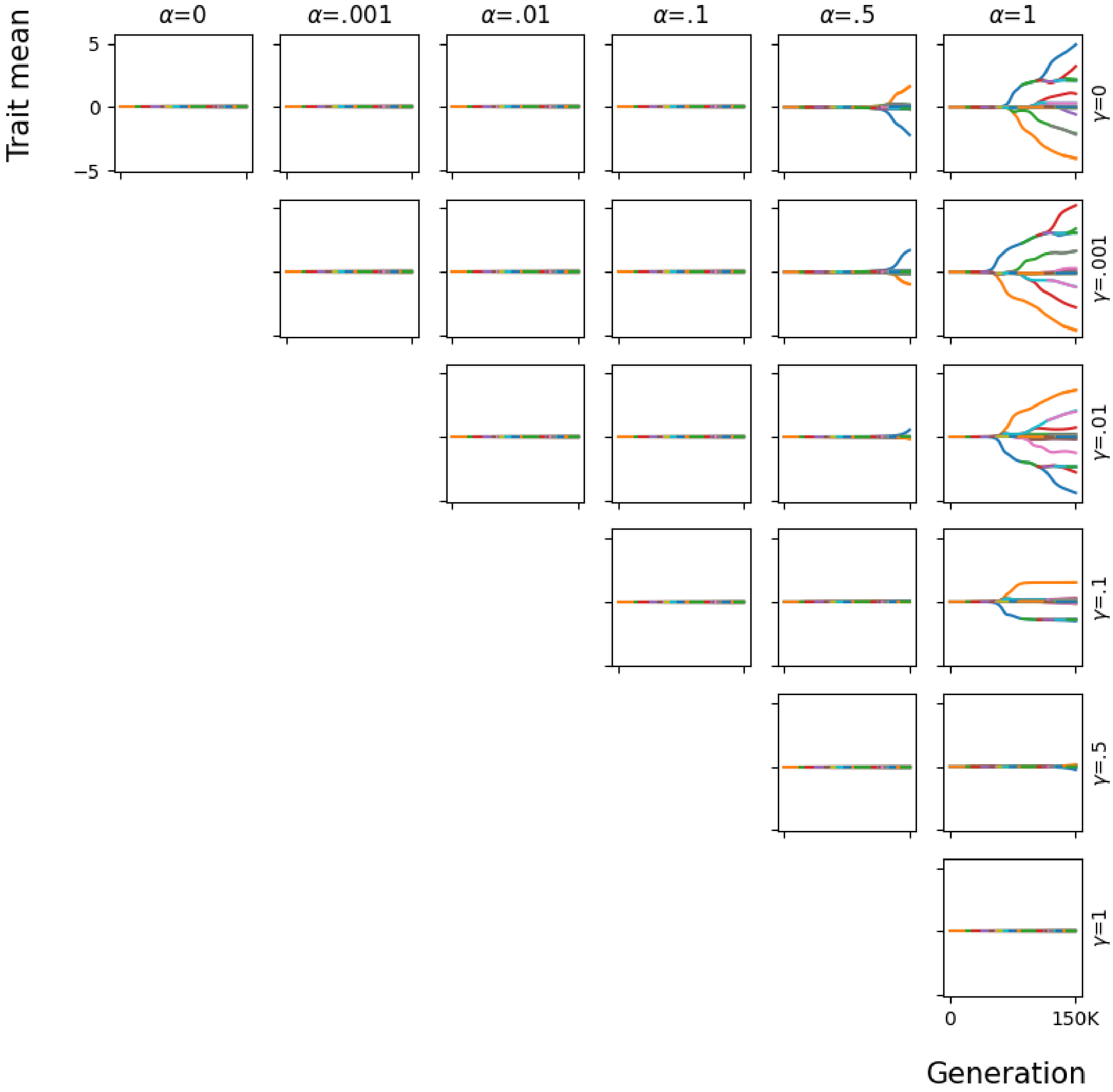

Fig. S3 – Trait trees along Tree 2 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

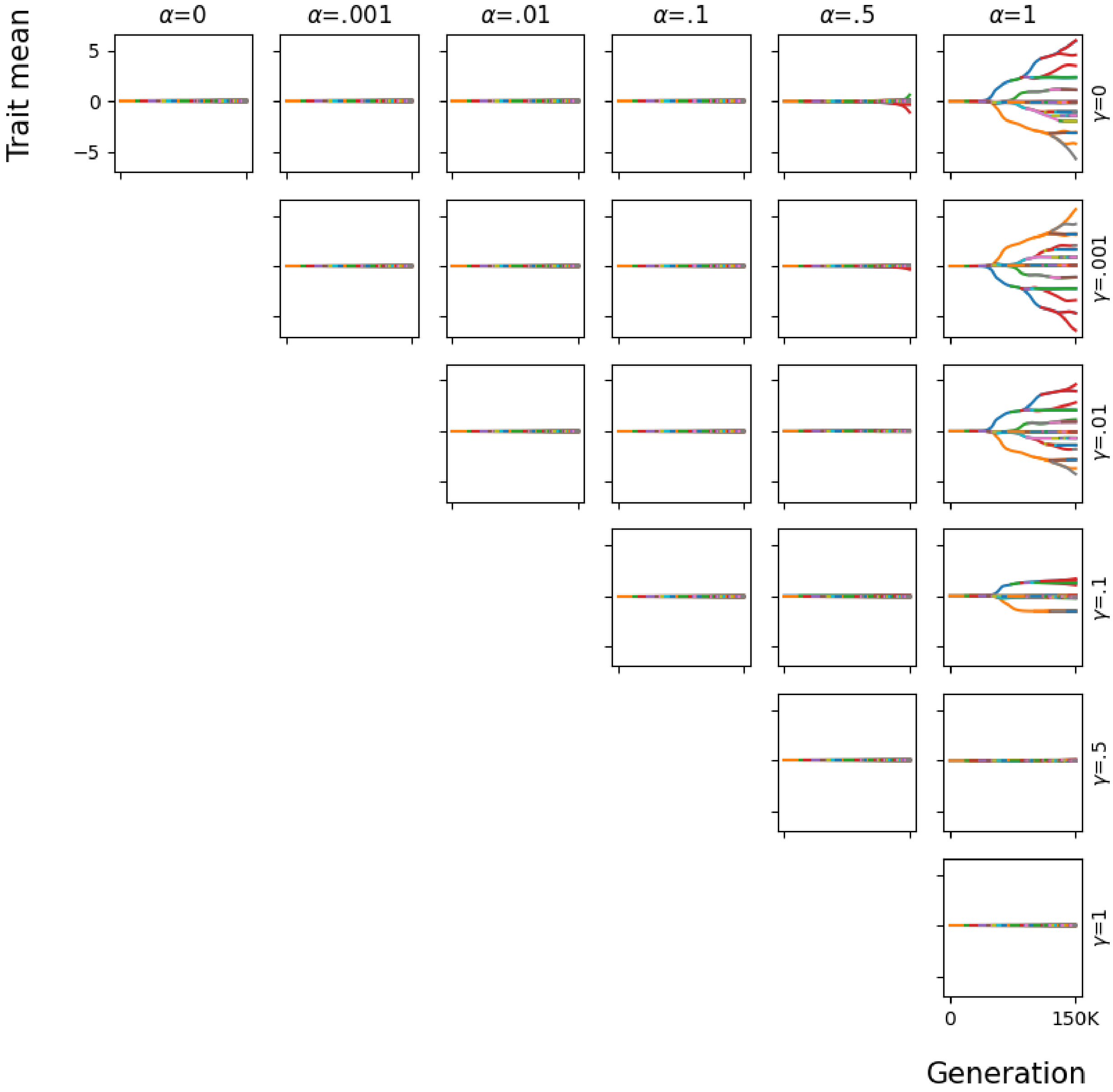

Fig. S4 – Trait trees along Tree 3 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

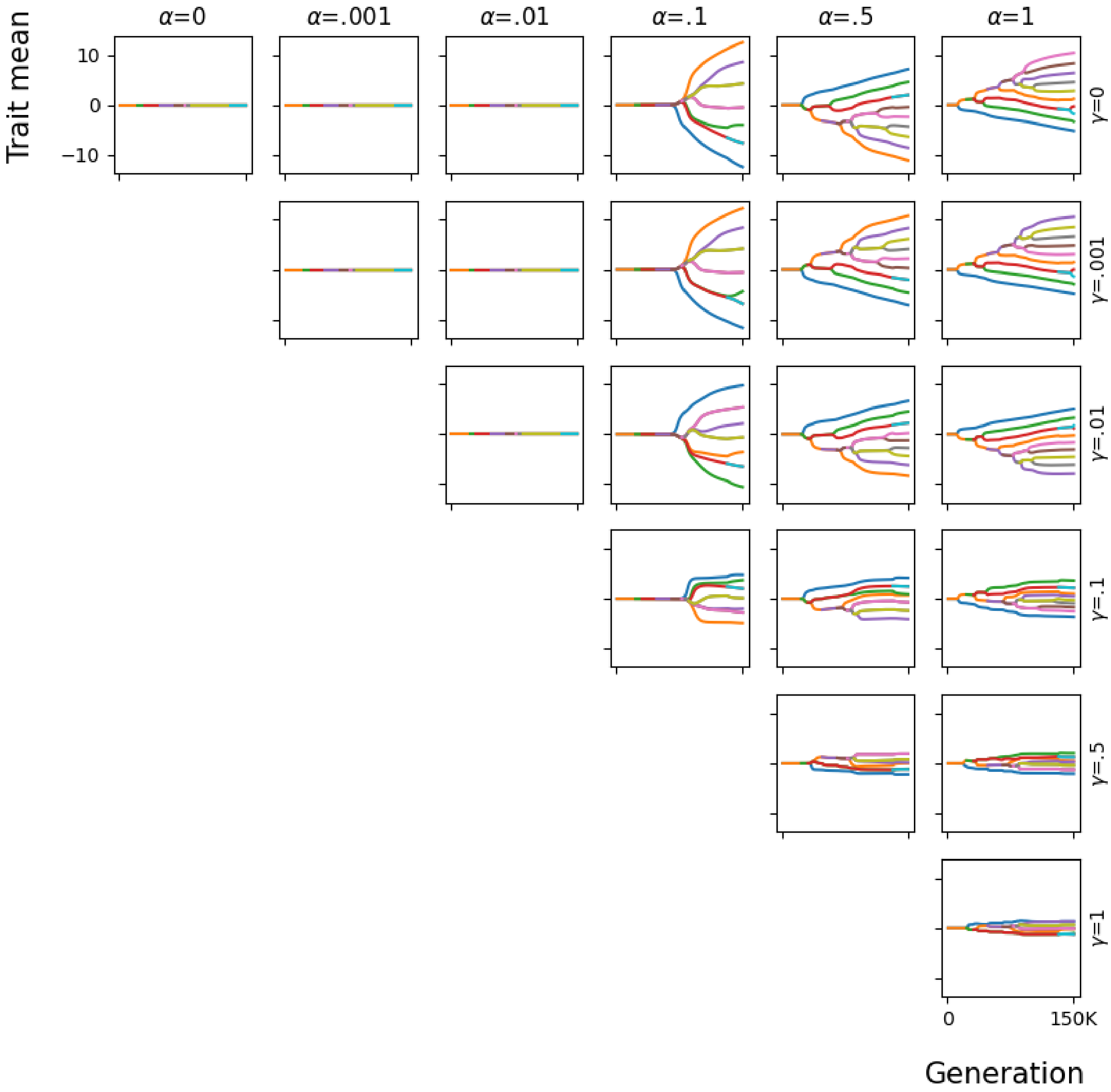

Fig. S5 – Trait trees along Tree 1 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

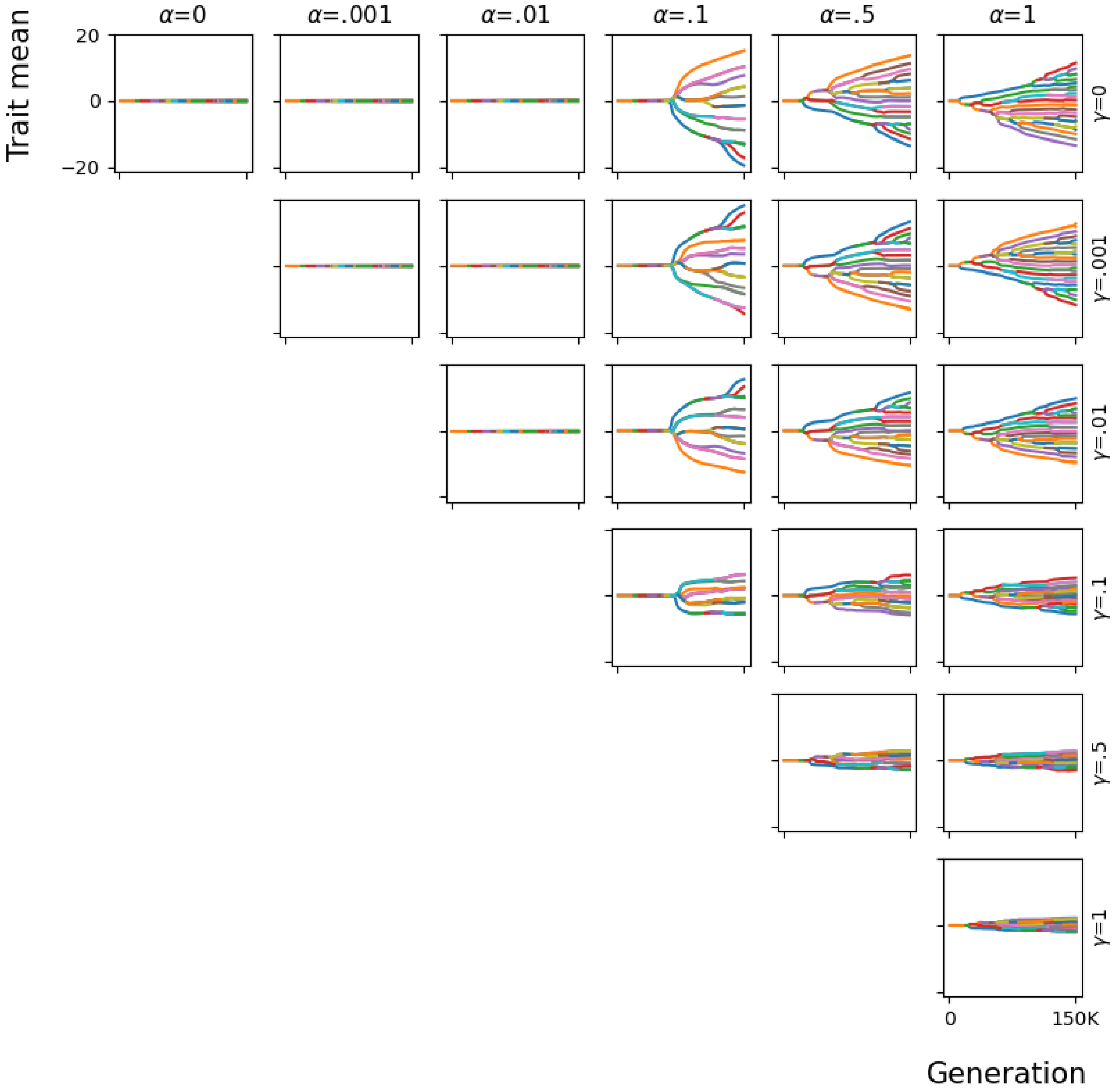

Fig. S6 – Trait trees along Tree 2 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

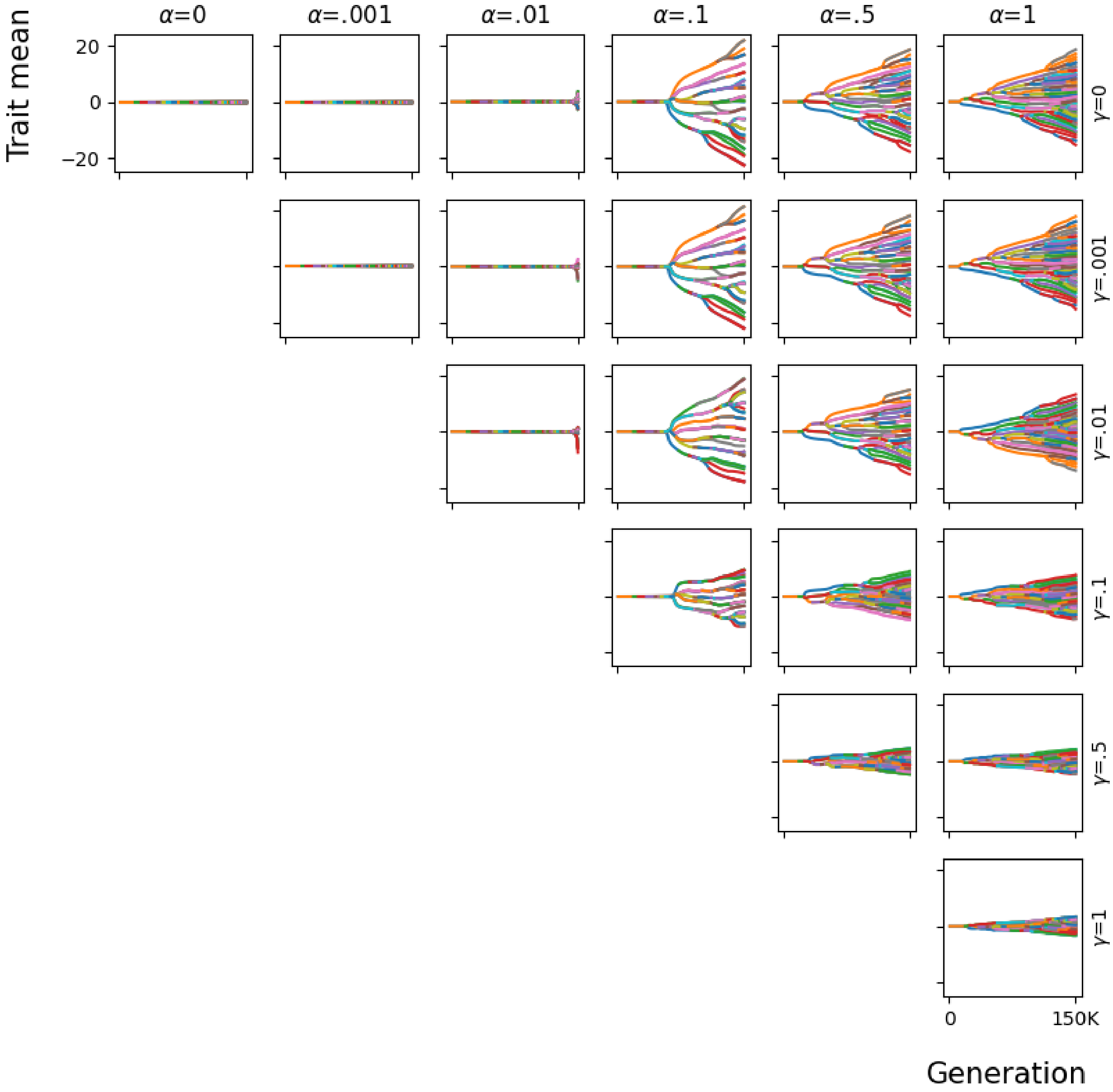

Fig. S7 – Trait trees along Tree 3 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 100$ indicating in total 1,500 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

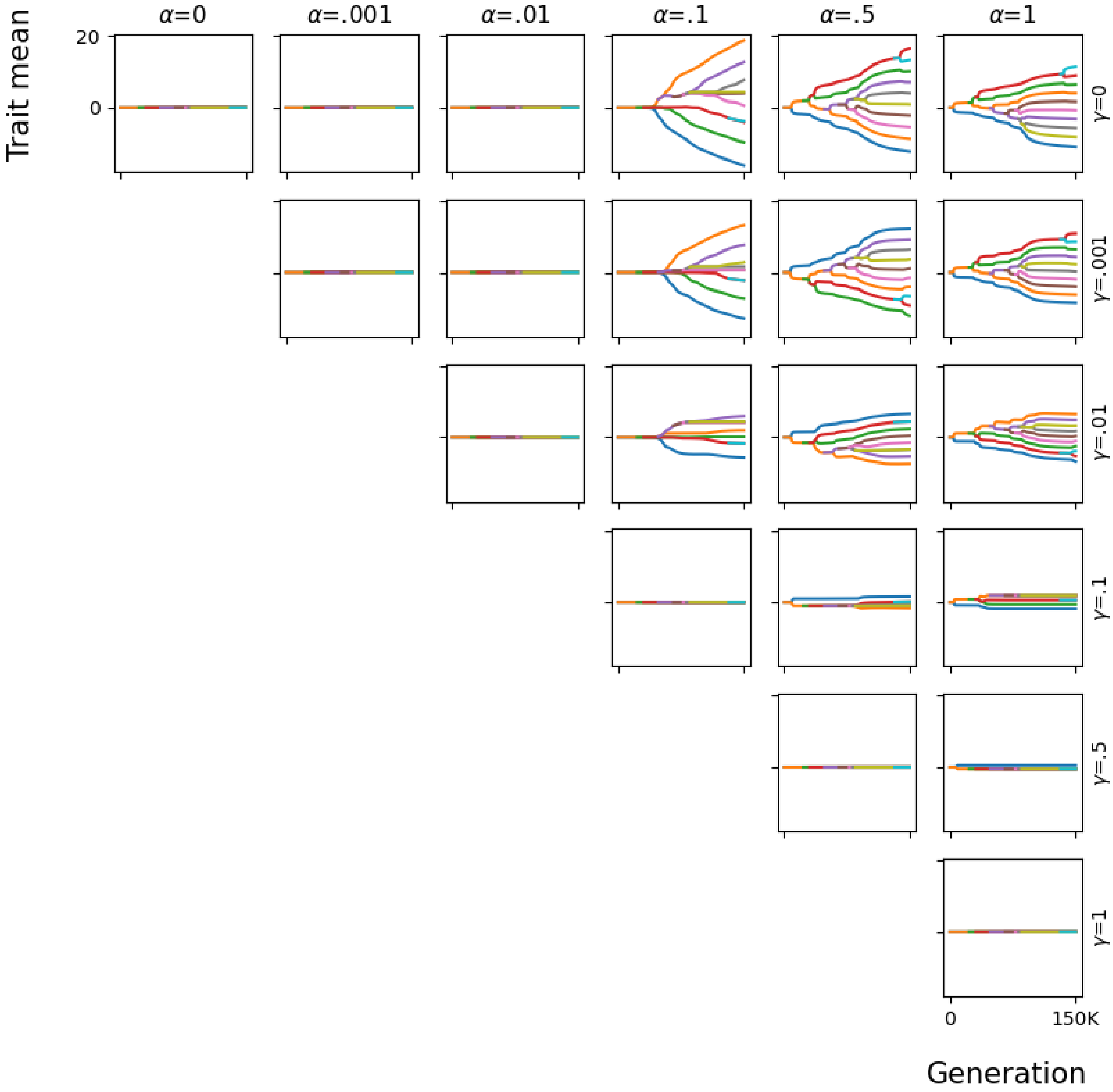

Fig. S8 – Trait trees along Tree 1 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

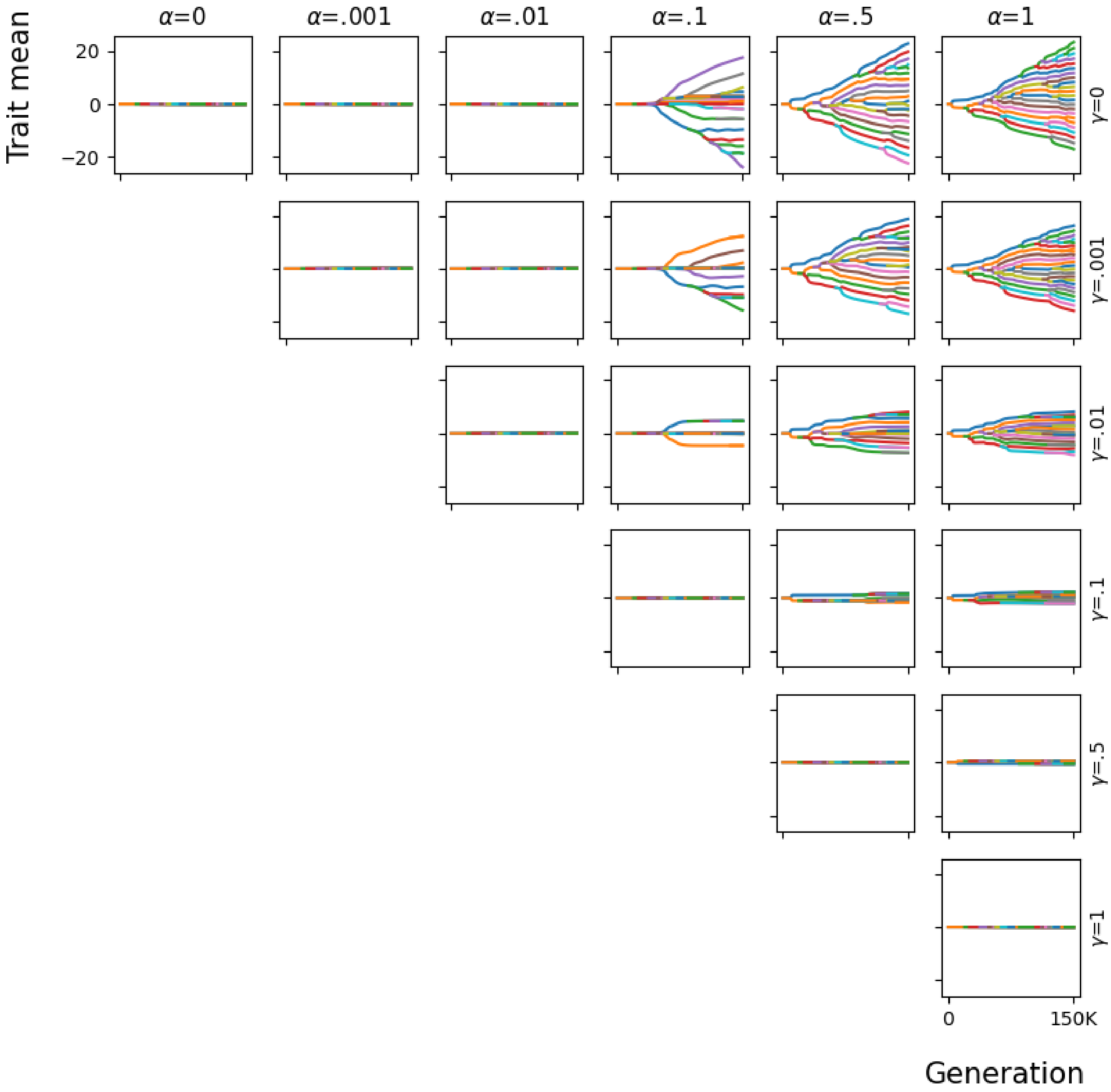

Fig. S9 – Trait trees along Tree 2 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

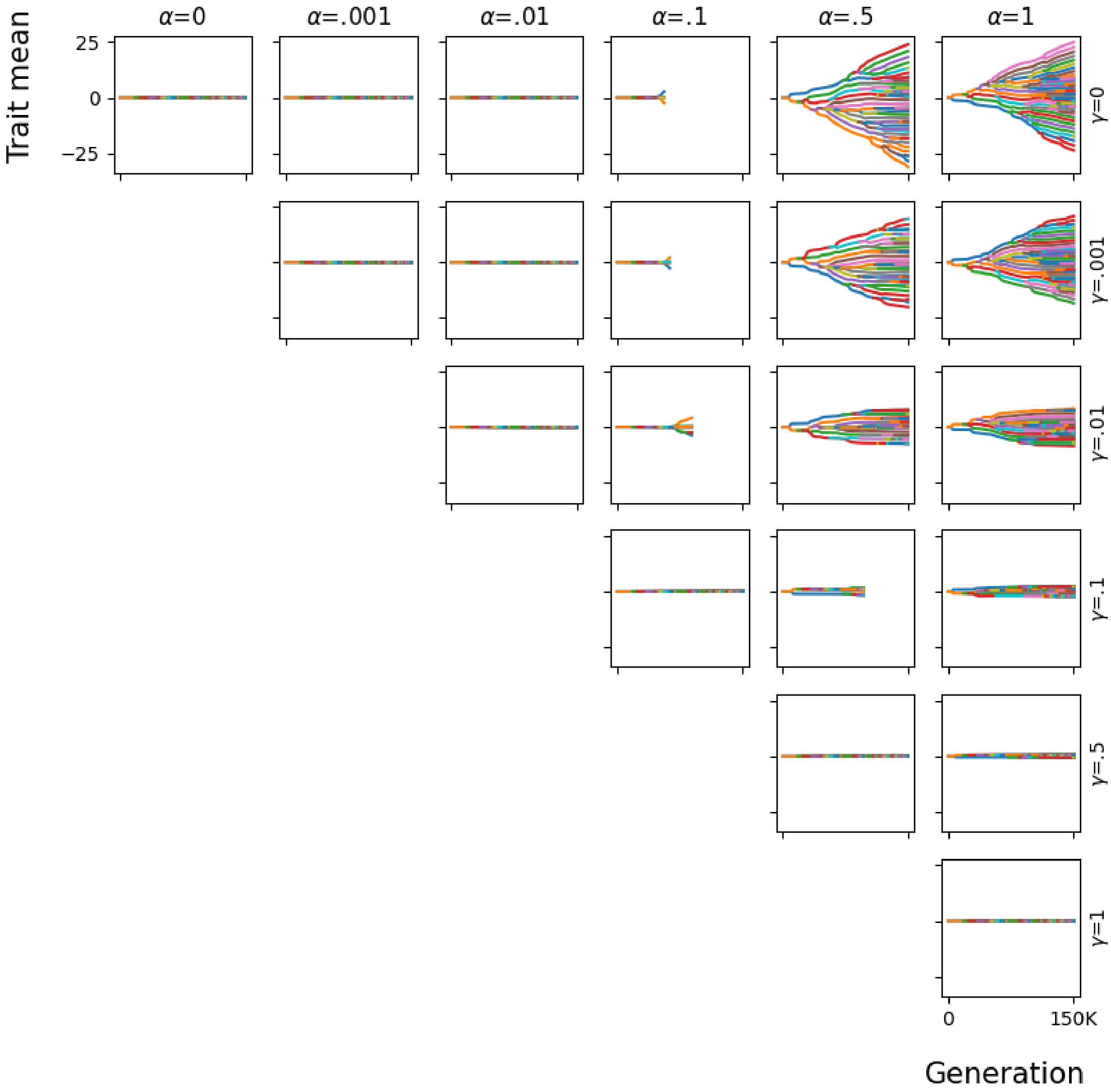

Fig. S10 – Trait trees along Tree 3 across all parameter combinations under the AWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

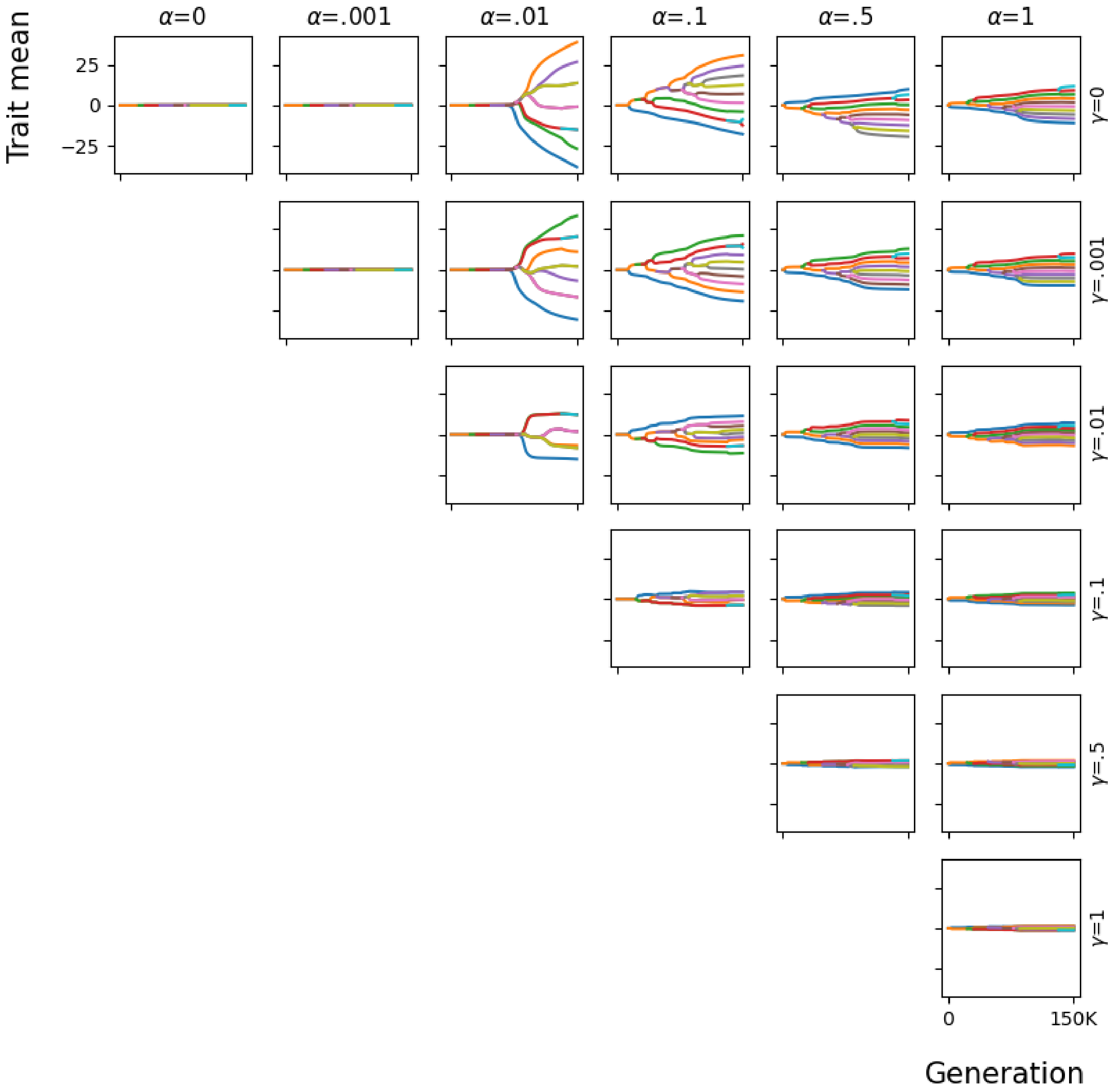

Fig. S11 – Trait trees along Tree 1 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

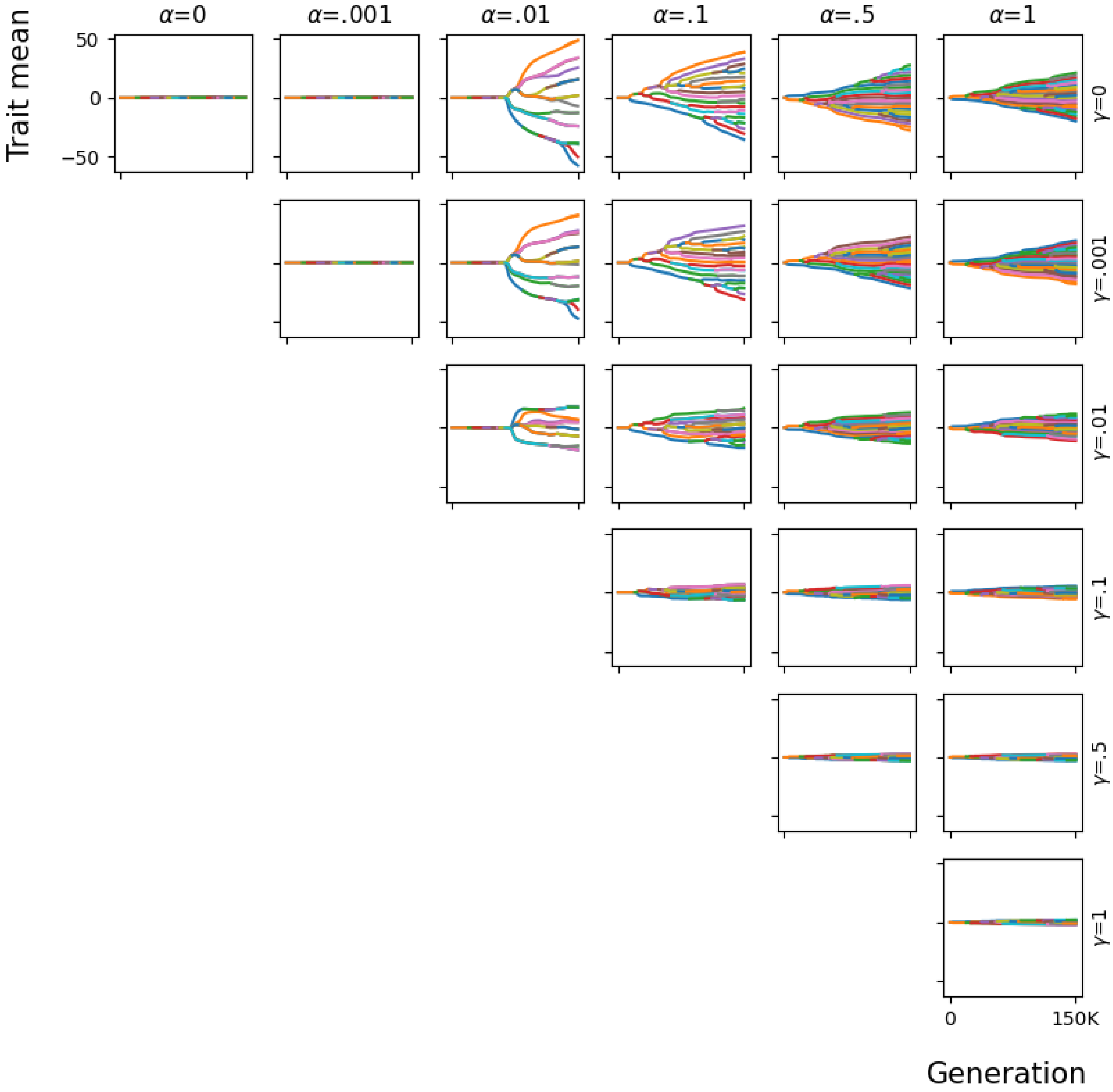

Fig. S12 – Trait trees along Tree 2 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

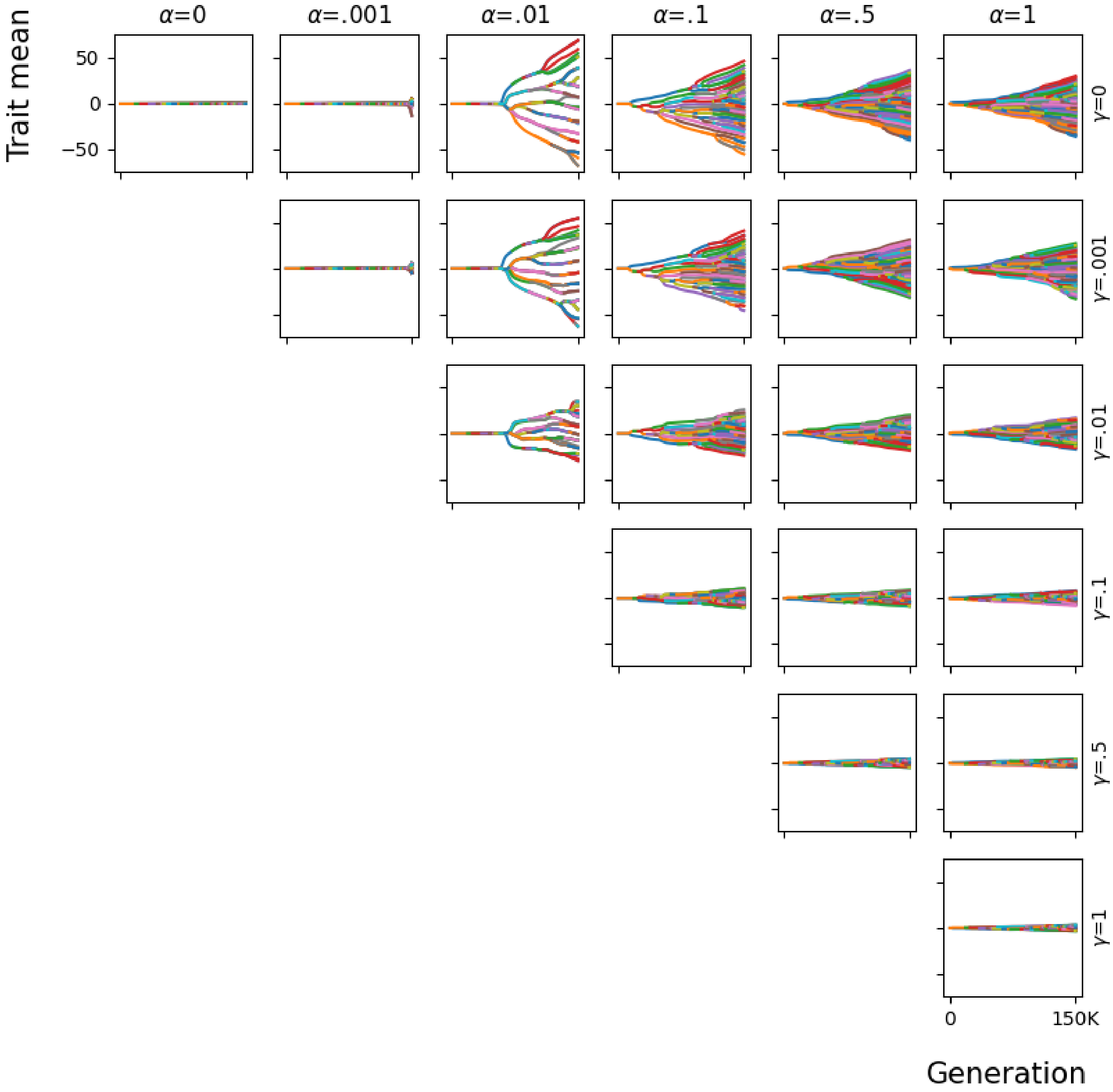

Fig. S13 – Trait trees along Tree 3 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 1000$ indicating in total 15,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

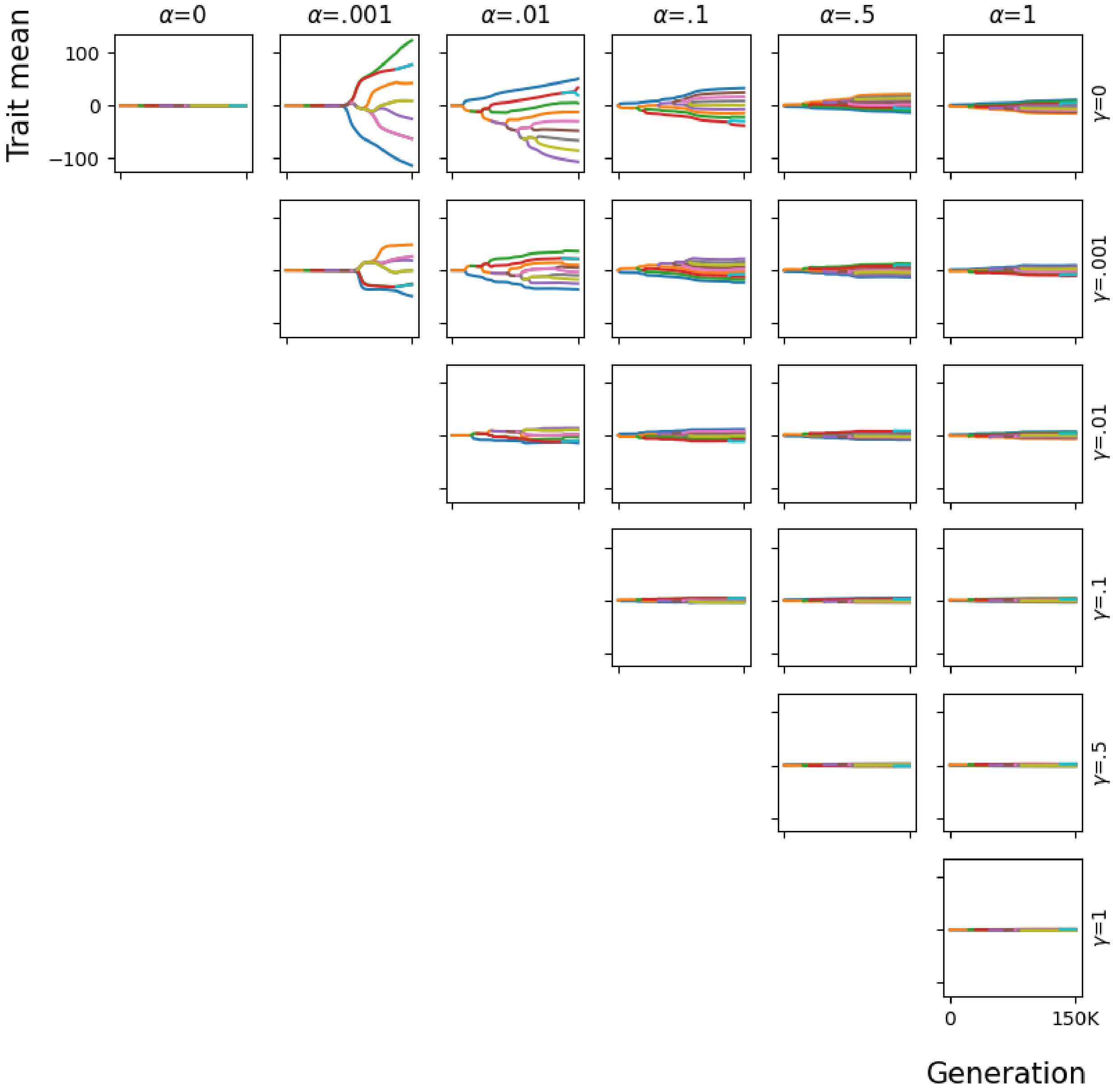

Fig. S14 – Trait trees along Tree 1 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 10000$ indicating in total 150,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

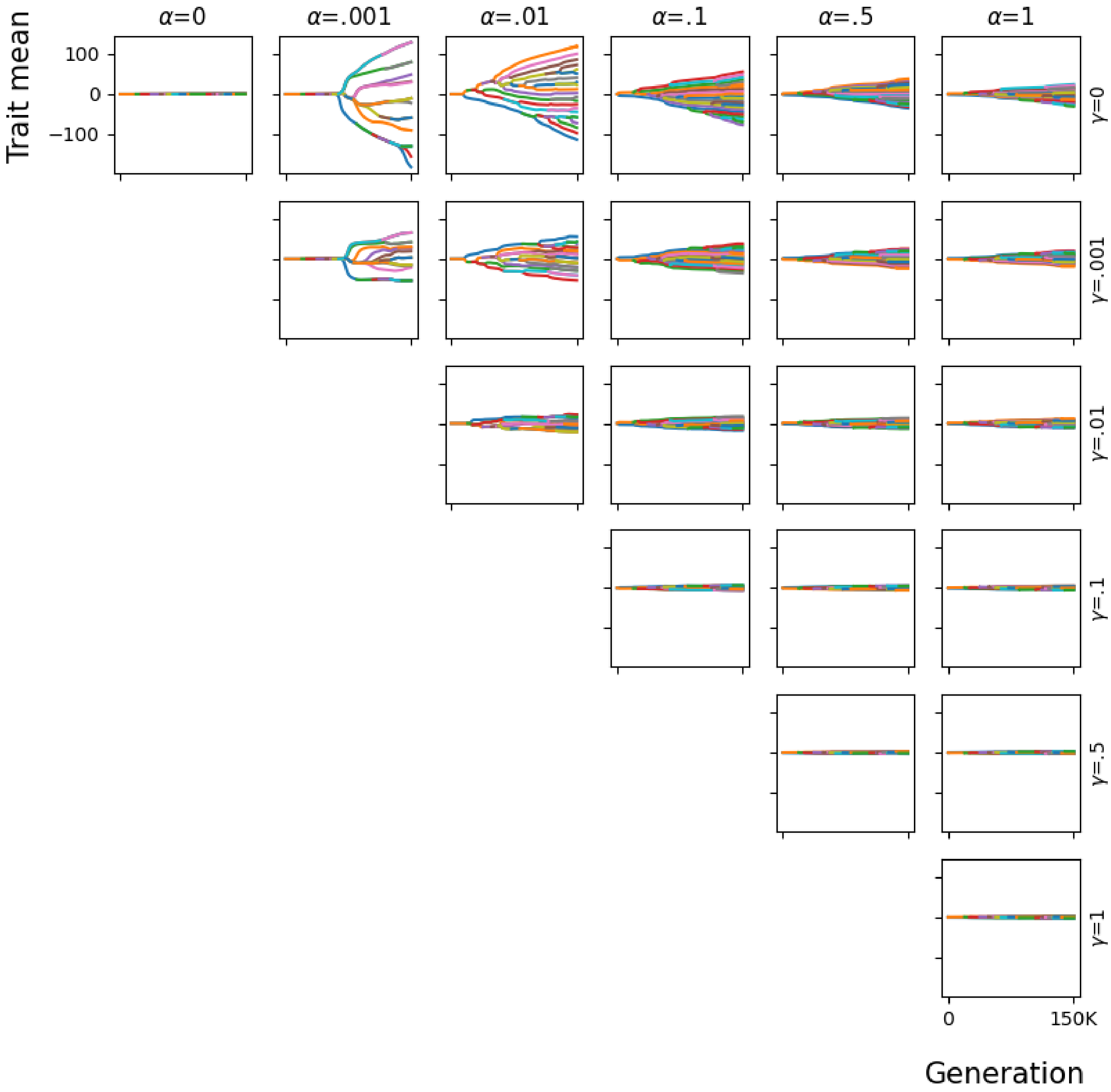

Fig. S15 – Trait trees along Tree 2 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 10000$ indicating in total 150,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

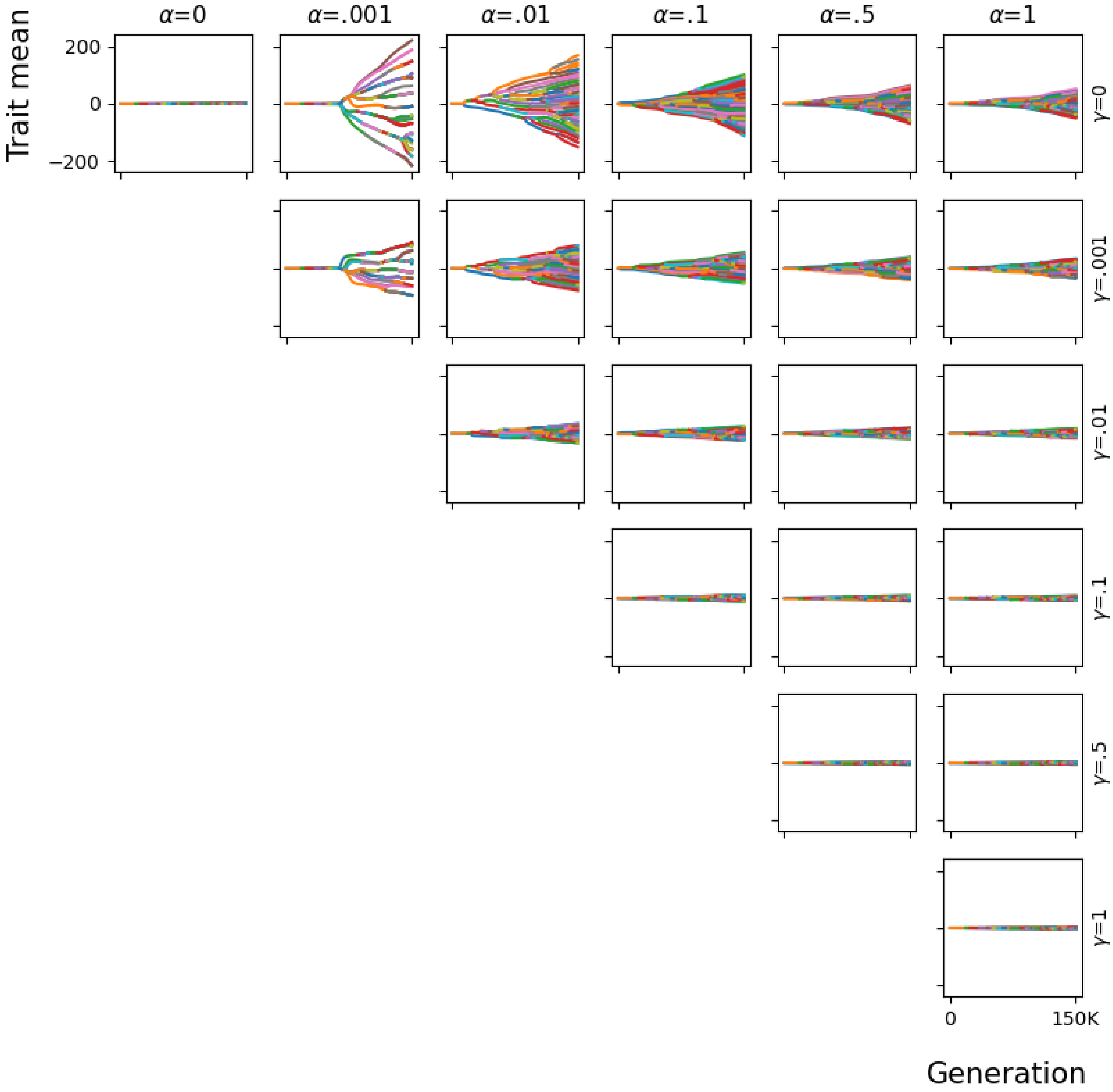

Fig. S16 – Trait trees along Tree 3 across all parameter combinations under the UWC model. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 10000$ indicating in total 150,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

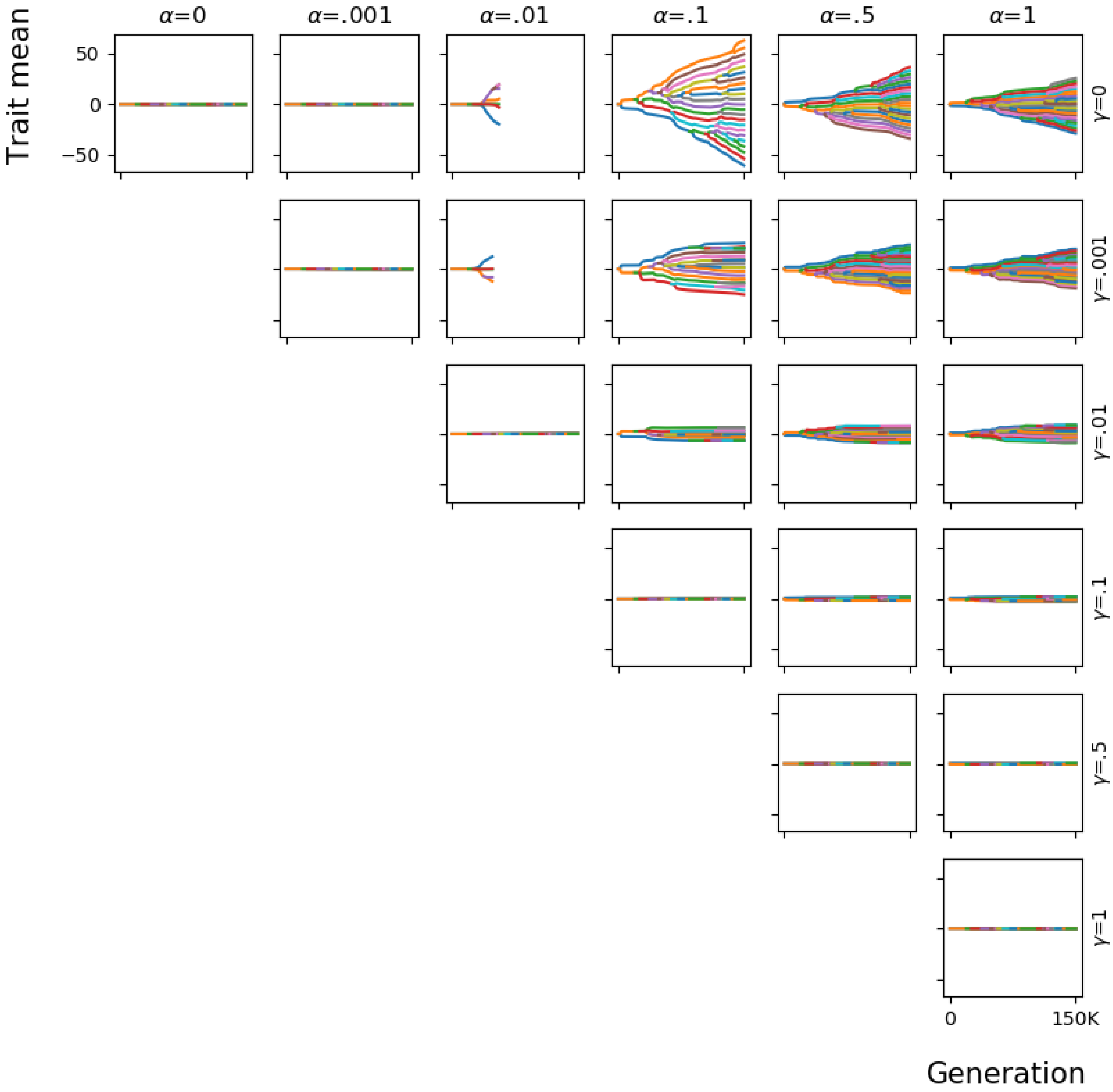

Fig. S17 – Trait trees along Tree 2 across all parameter combinations. Different colors denote the traits of different species. The simulation is initialized with two ancestral species with identical trait mean 0 and variance 1/2 and the population size is randomly chosen from a normal distribution centering 500 with variance of 100. The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 30$ with a crown age of 15 Myr. The time scaling parameter is $s = 10000$ indicating in total 150,000 time steps in trait simulation. Trees are shown only for the informative parameters domain ($\gamma < \alpha$); when $\gamma > \alpha$, the trait trees are all almost identical: a highly compressed line such as the pattern on the diagonal.

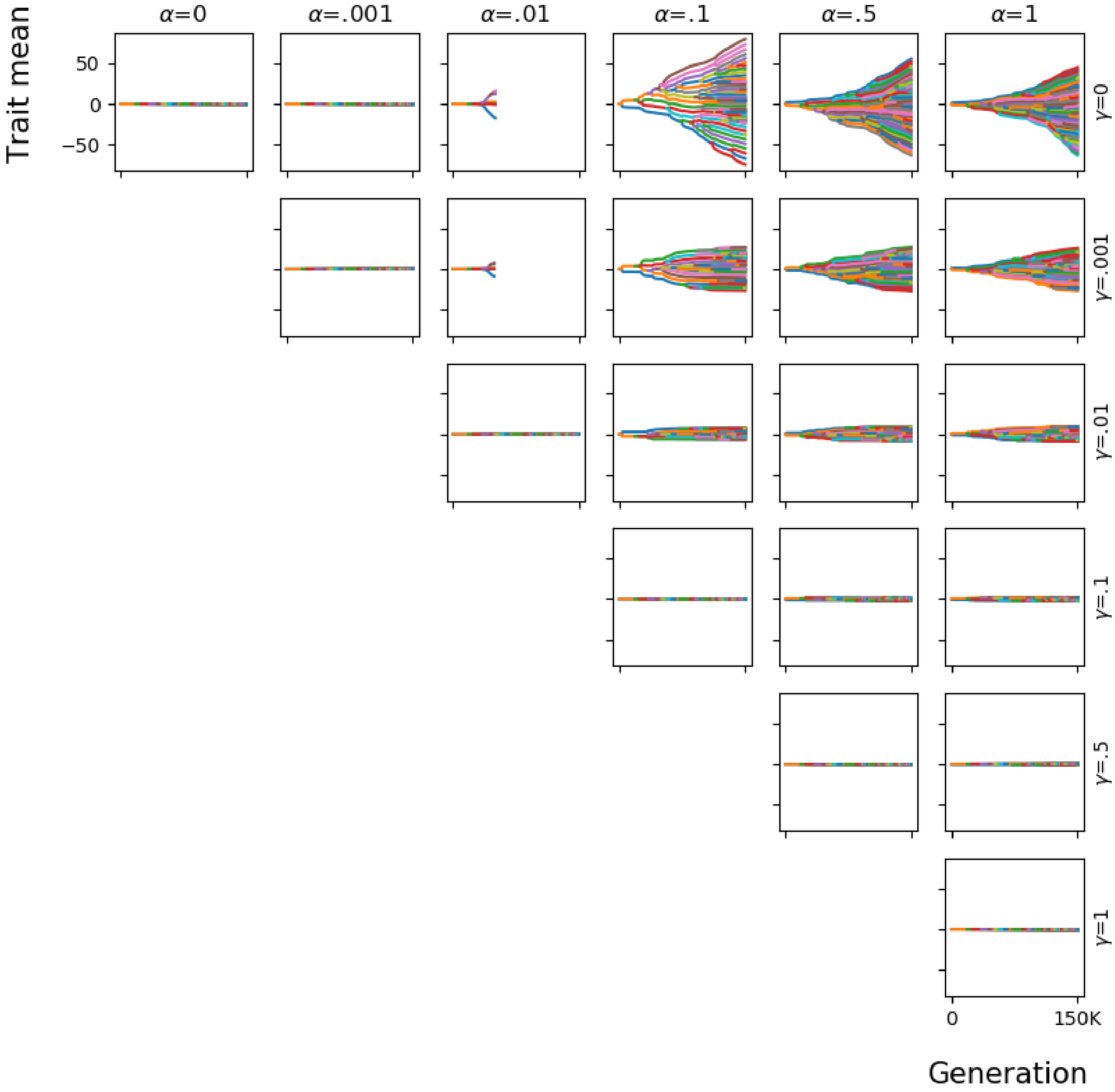

Fig. S18 – Trait trees along Tree 3 across all parameter combinations (similar to Fig. S17). The parameters used to generate the phylogenetic tree are $\lambda = 0.4, \mu = 0, K = 100$ with a crown age of 15 Myr. The time scaling parameter is $s = 10000$ indicating in total 150,000 time steps in trait simulation. Some plots show incomplete trait trees, because there is no complete simulation under these parameter combinations after 10,000 attempts.

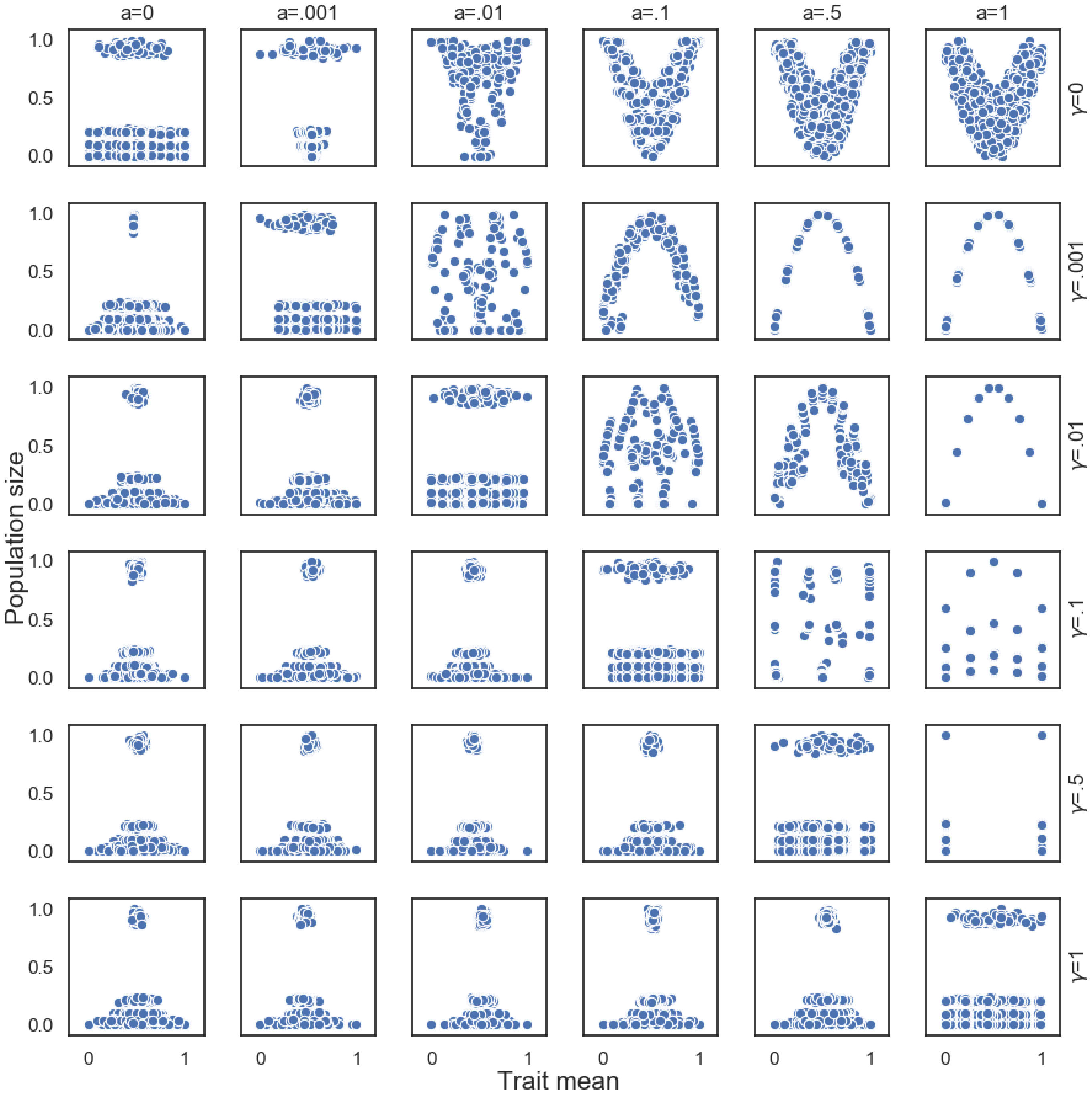

Fig. S19 – Abundance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 1. Both the abundance and the trait mean are normalized.

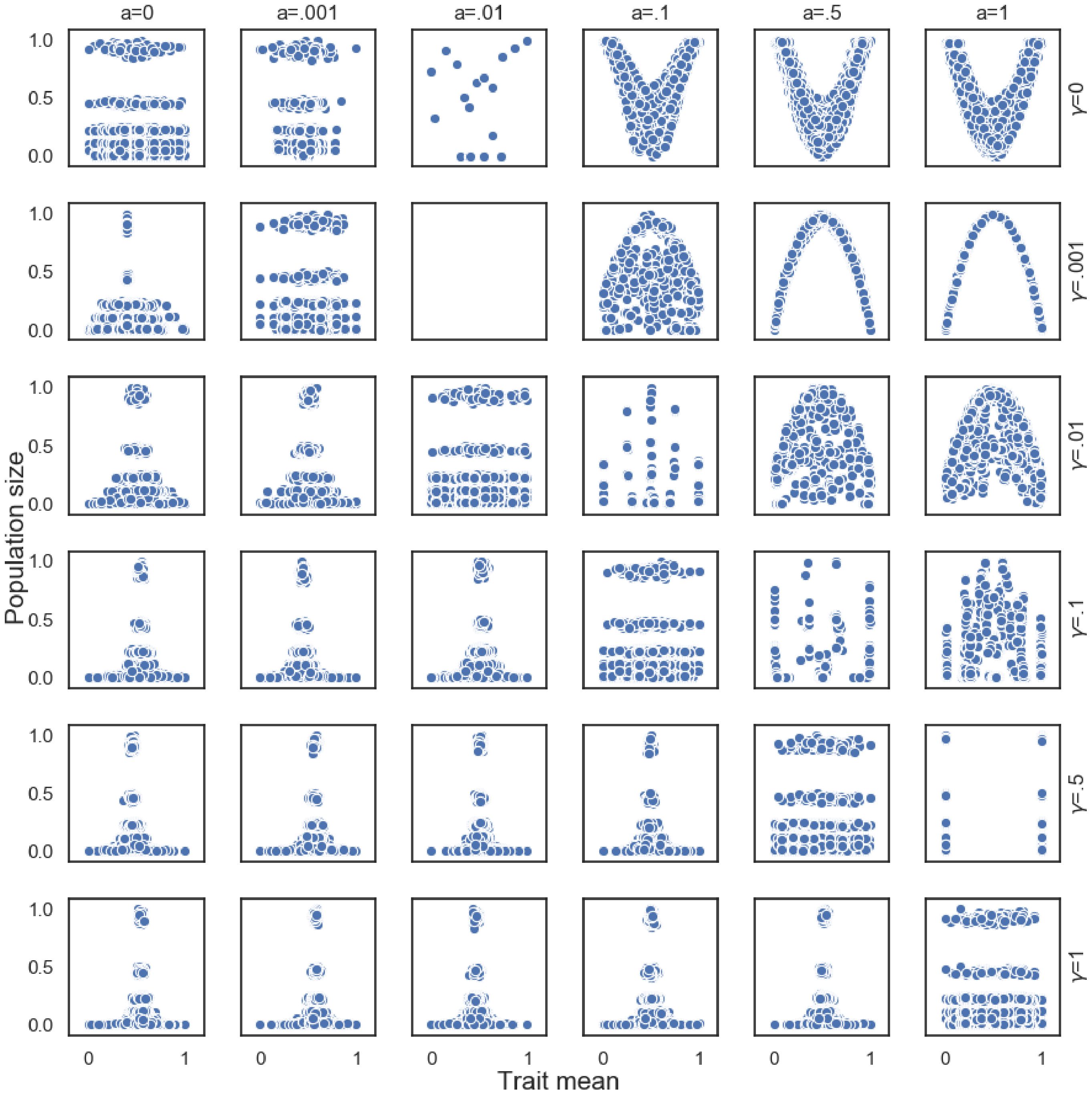

Fig. S20 – Abundance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 2. Both the abundance and the trait mean are normalized.

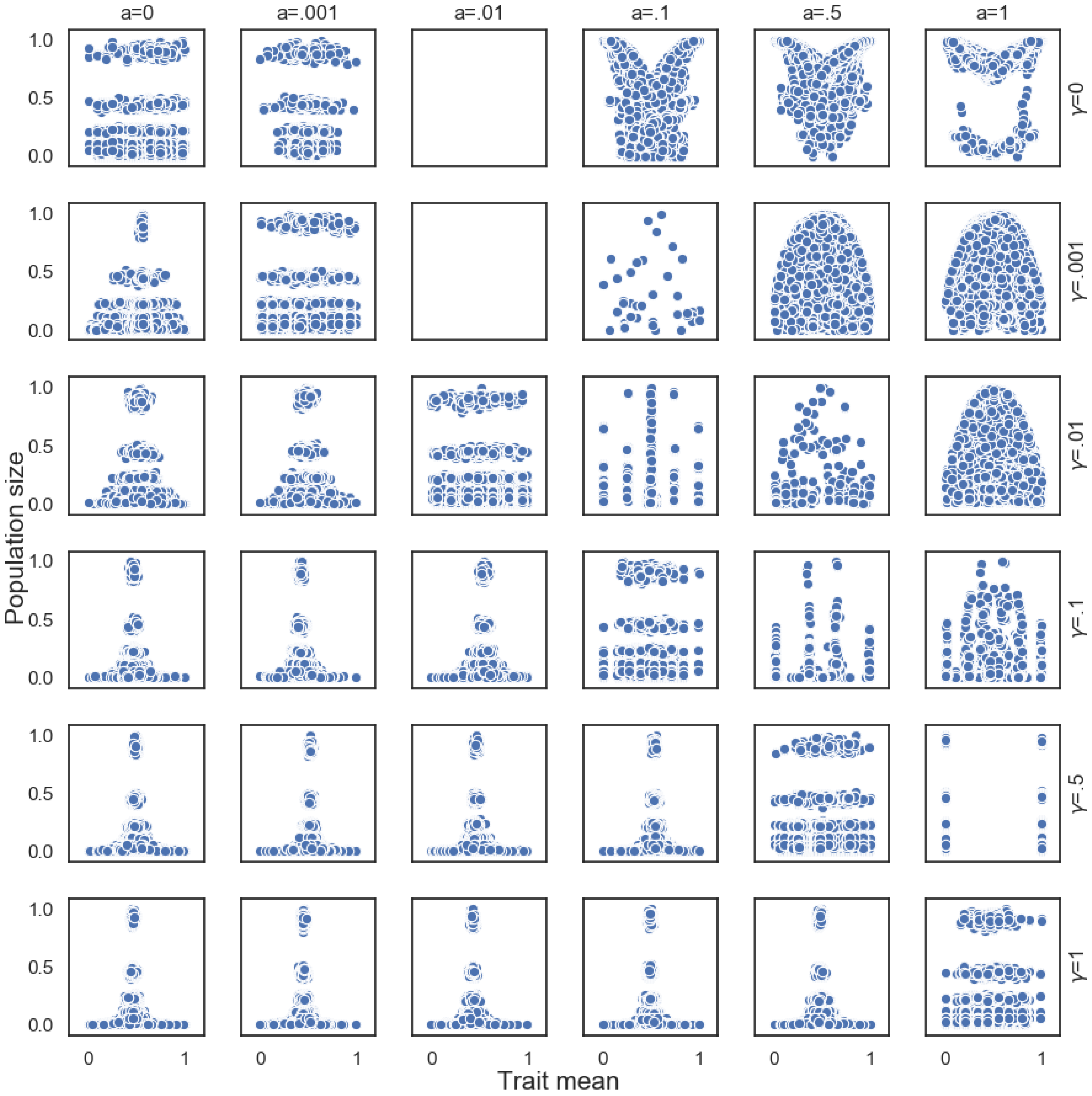

Fig. S21 – Abundance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 3. Both the abundance and the trait mean are normalized.

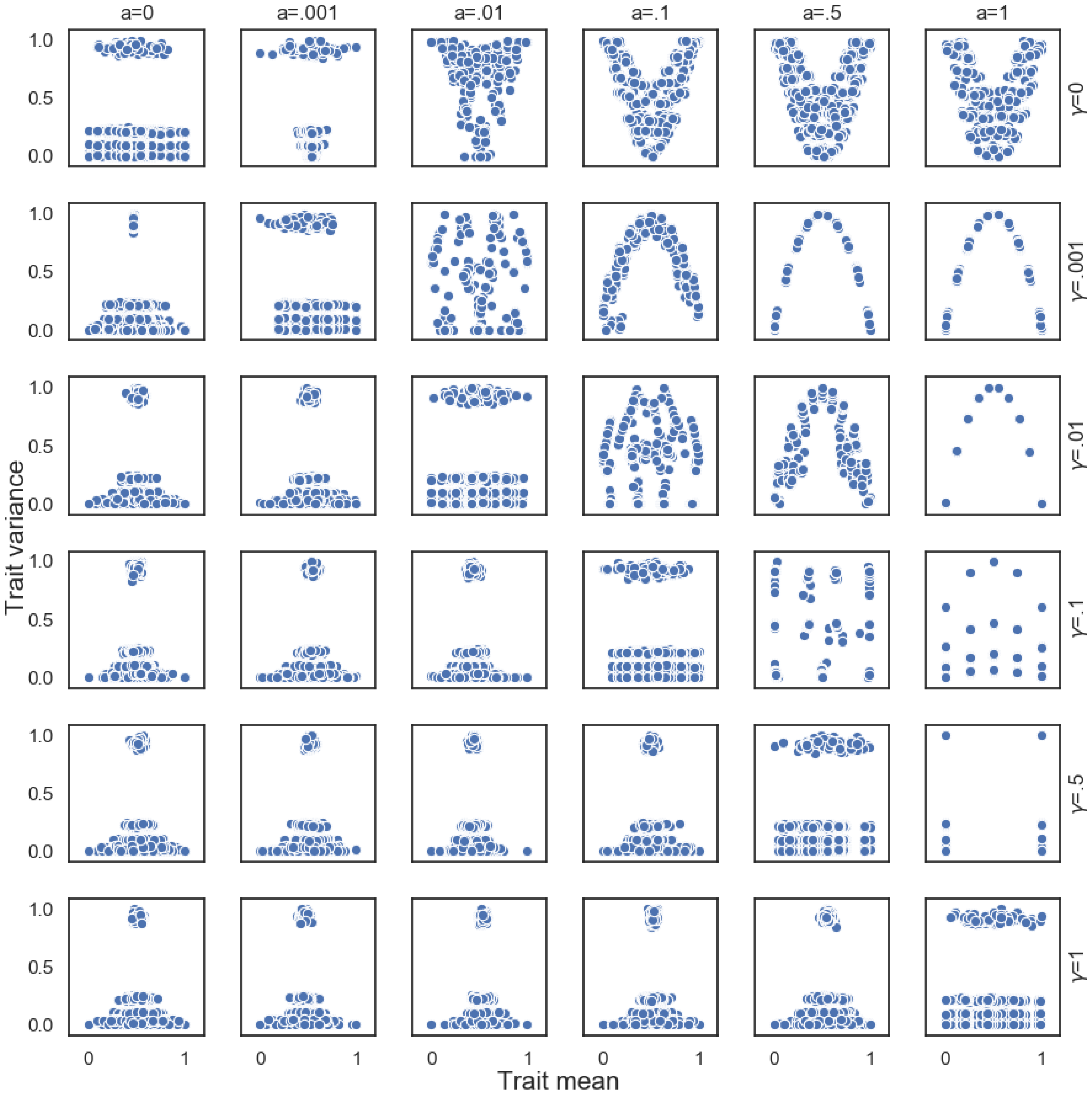

Fig. S22 – Trait variance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 1. Both the trait mean and variance are normalized.

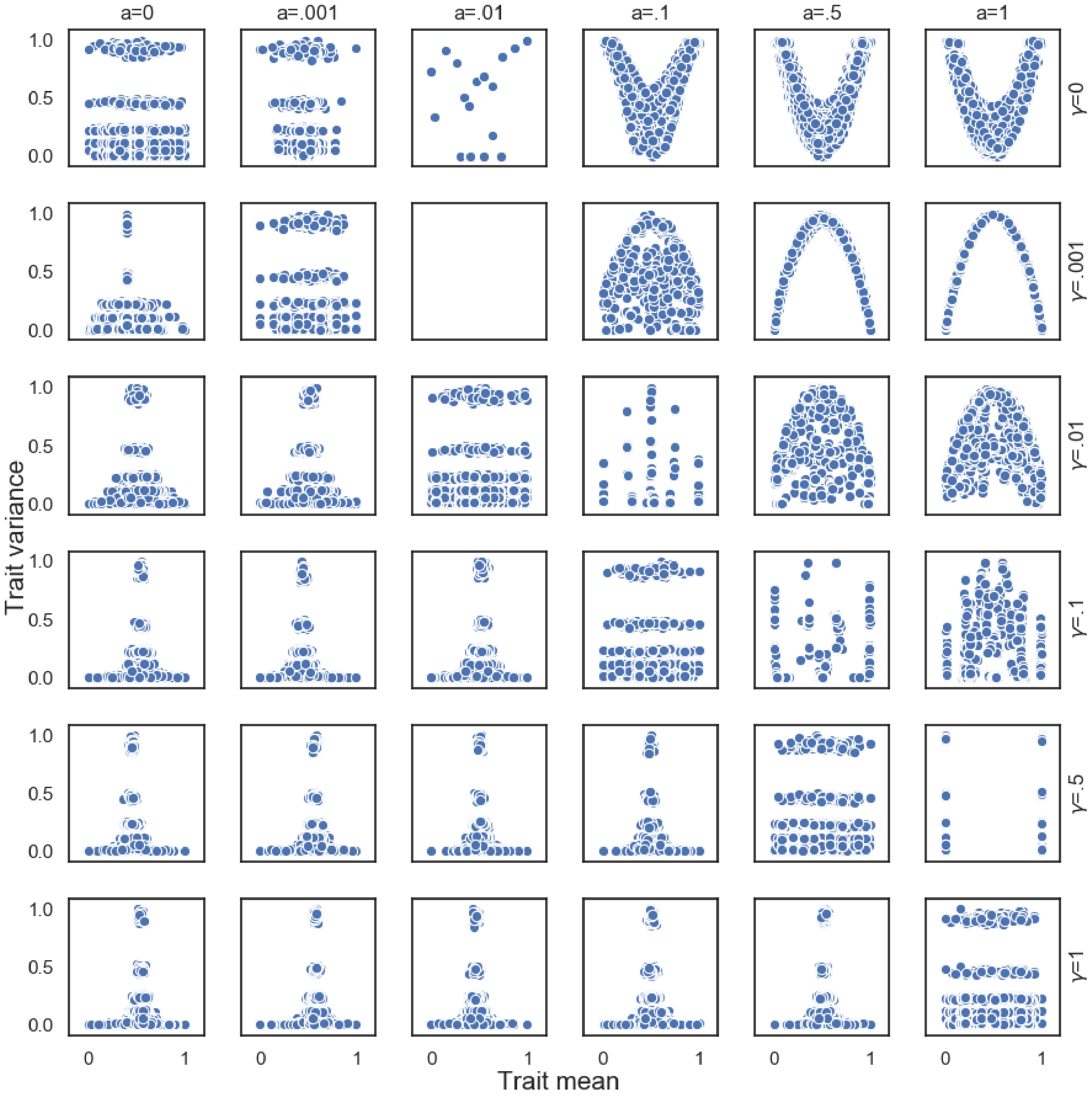

Fig. S23 – Trait variance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 2. Both the trait mean and variance are normalized.

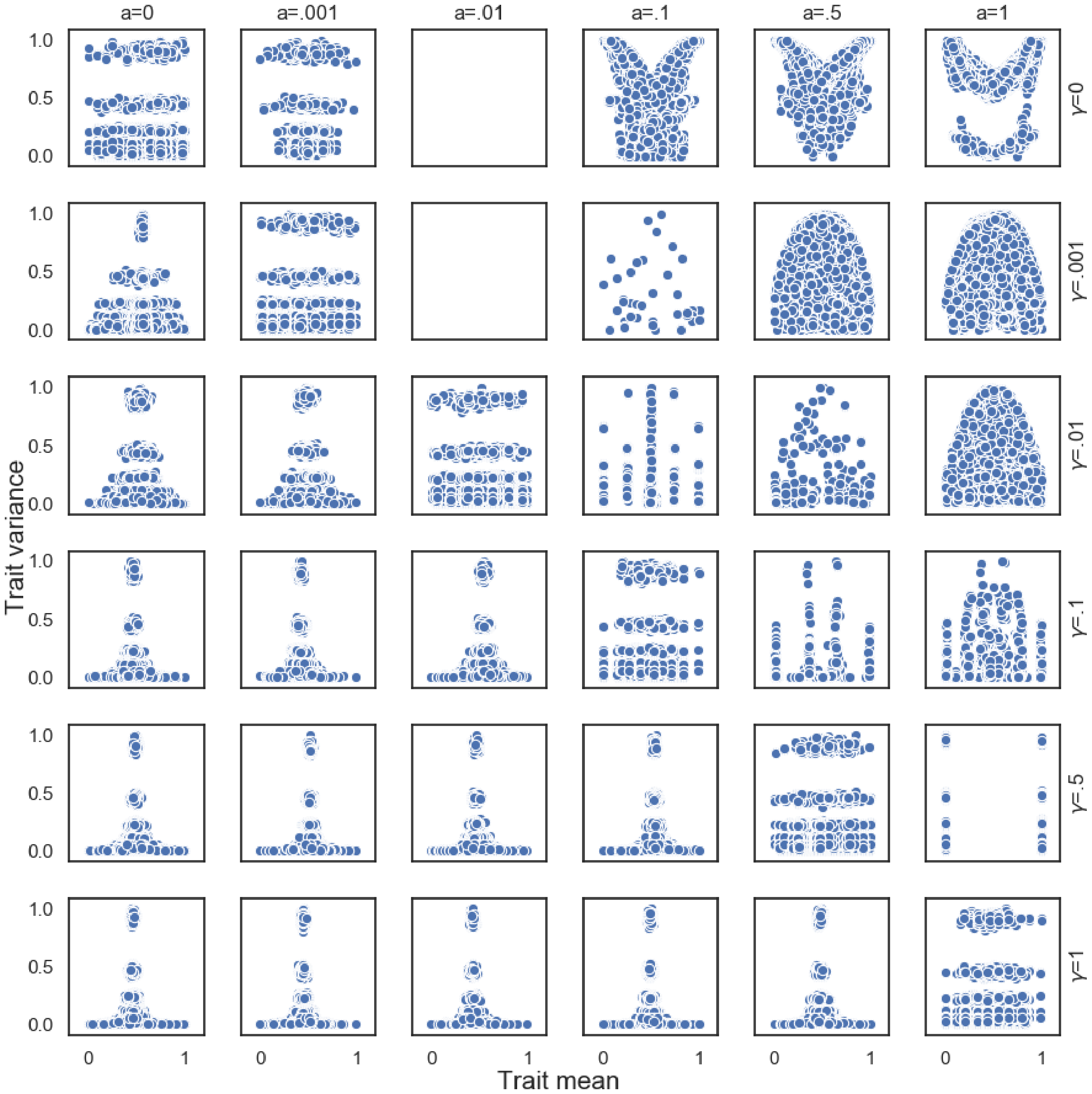

Fig. S24 – Trait variance distribution against trait means for 100 replicate simulations under the 36 parameter combinations of Scenarios 3. Both the trait mean and variance are normalized.

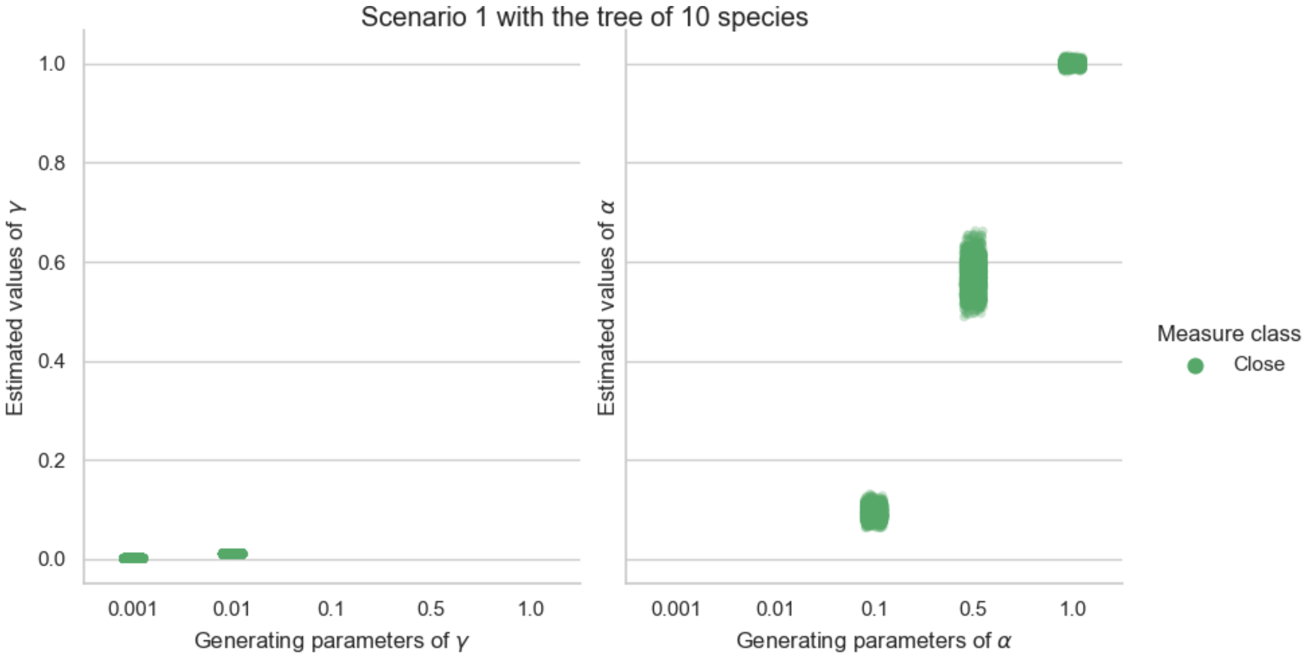

Fig. S25 – Comparison of generating and estimated parameter values for the generating parameter combinations for whichthe ratio $\frac{\sqrt{\alpha/\gamma}}{\text{Richness}}$ is in the range of $(0.5, 1.5)$ under Scenario 1. The green dots denote parameter estimates for which $m < 1$; the orange dots denote parameter estimations for which $1 < m < 2$; the red dots denote parameter estimates for which $m > 2$.

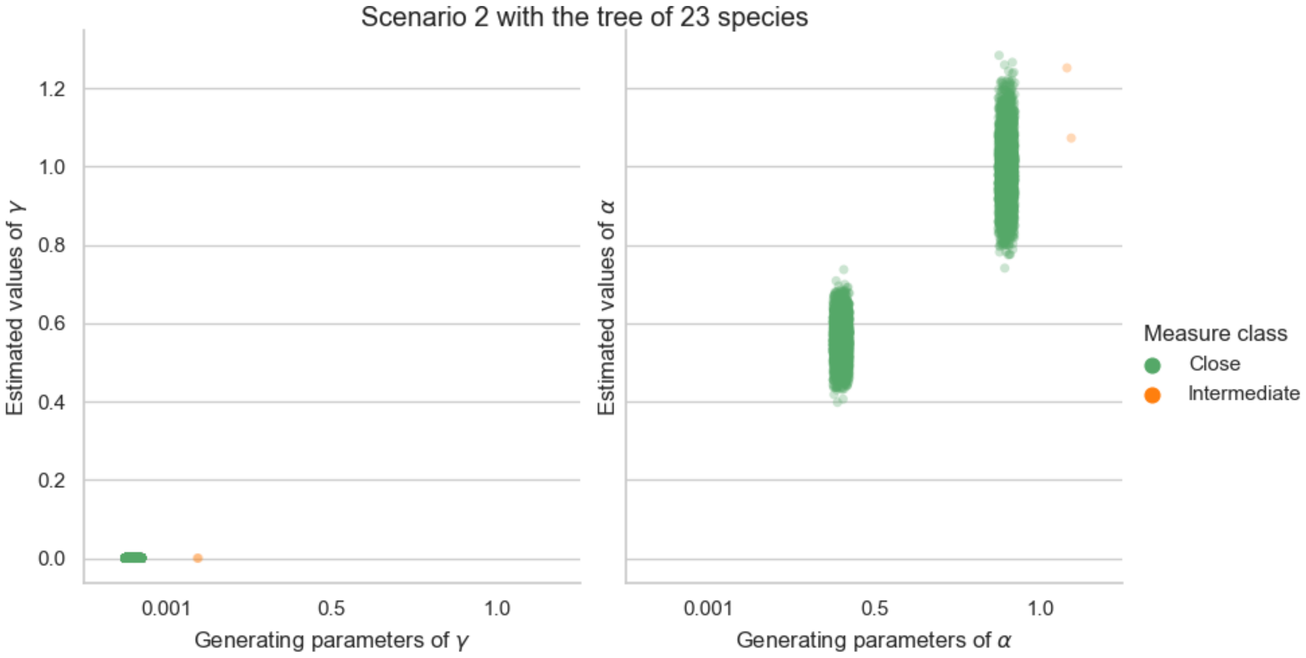

Fig. S26 – Comparison of generating and estimated parameter values for the generating parameter combinations for whichthe ratio $\frac{\sqrt{\alpha/\gamma}}{\text{Richness}}$ is in the range of $(0.5, 1.5)$ under Scenario 2. The green dots denote parameter estimates for which $m < 1$; the orange dots denote parameter estimations for which $1 < m < 2$; the red dots denote parameter estimates for which $m > 2$.

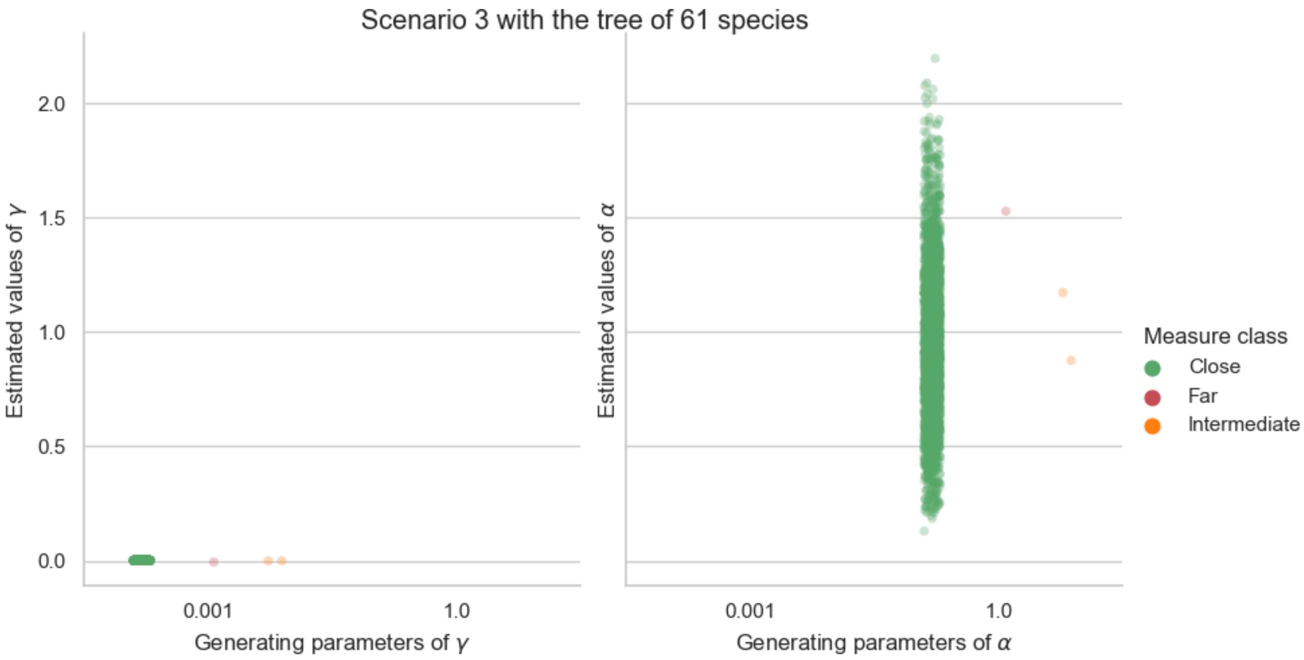

Fig. S27 – Comparison of generating and estimated parameter values for the generating parameter combinations for whichthe ratio $\frac{\sqrt{\alpha/\gamma}}{\text{Richness}}$ is in the range of $(0.5, 1.5)$ under Scenario 3. The green dots denote parameter estimates for which $m < 1$; the orange dots denote parameter estimations for which $1 < m < 2$; the red dots denote parameter estimates for which $m > 2$.

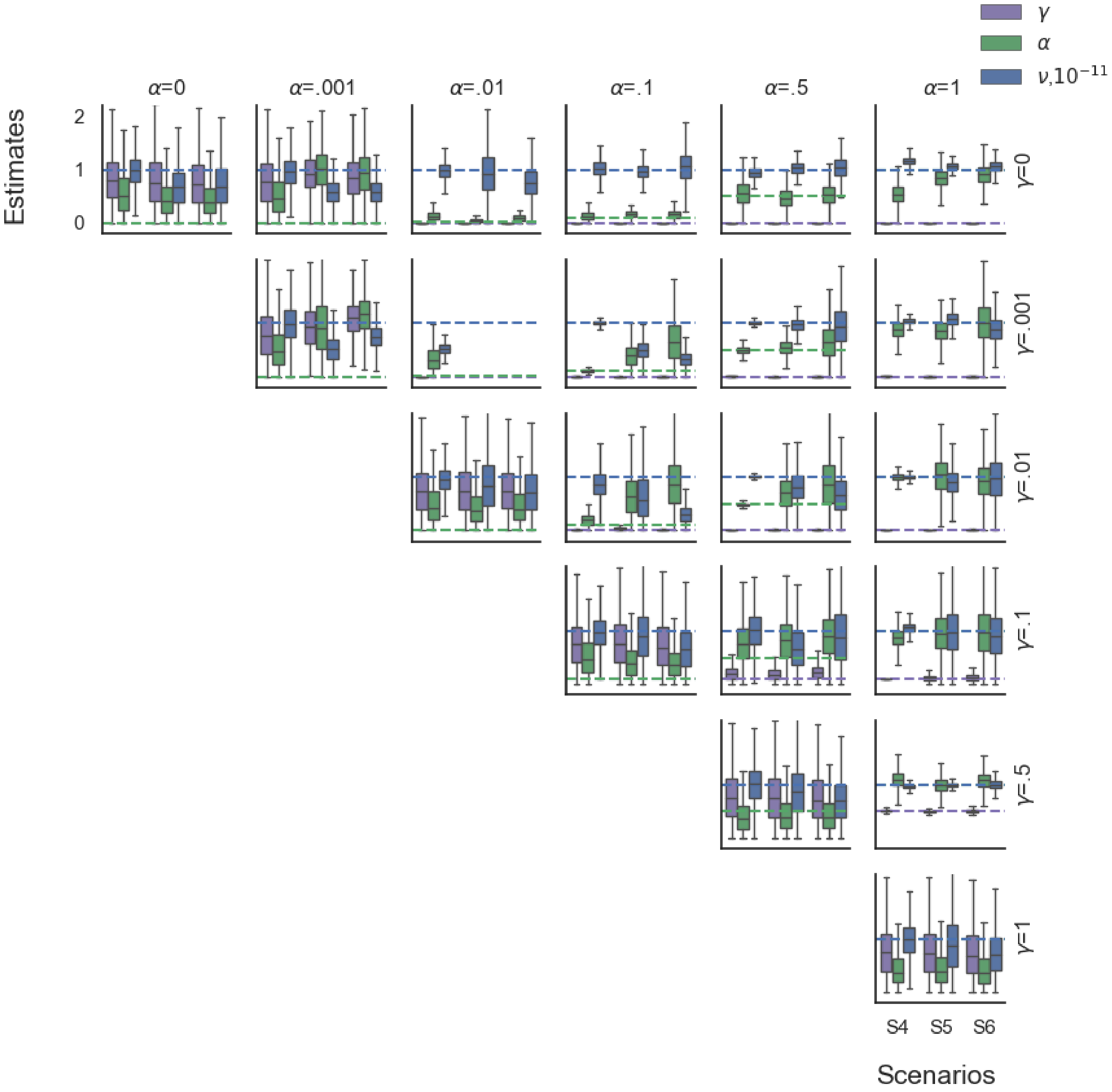

Fig. S28 – Parameter inference for the 36 parameter combinations under Scenarios 4, 5 and 6. The dashed lines in three colors denote three parameter values used to generate the data. Note that the scale of mutation rate is 10^{-11}. Box plots indicate the distribution of inferred parameter values, where the whiskers extend from the minimum to the maximum value. Some plots show no estimates for part of the parameter combinations, because there is no complete simulation under these parameter combinations after 10,000 attempts. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2$ with a crown age of 15 Myr. The clade-level carrying capacity is $K = 10, 30, 100$ for trees 4, 5, 6, respectively. Data are shown only for the informative parameter domain ($\gamma \leq \alpha$); when $\gamma > \alpha$, the parameter estimation show major bias and substantial variance.

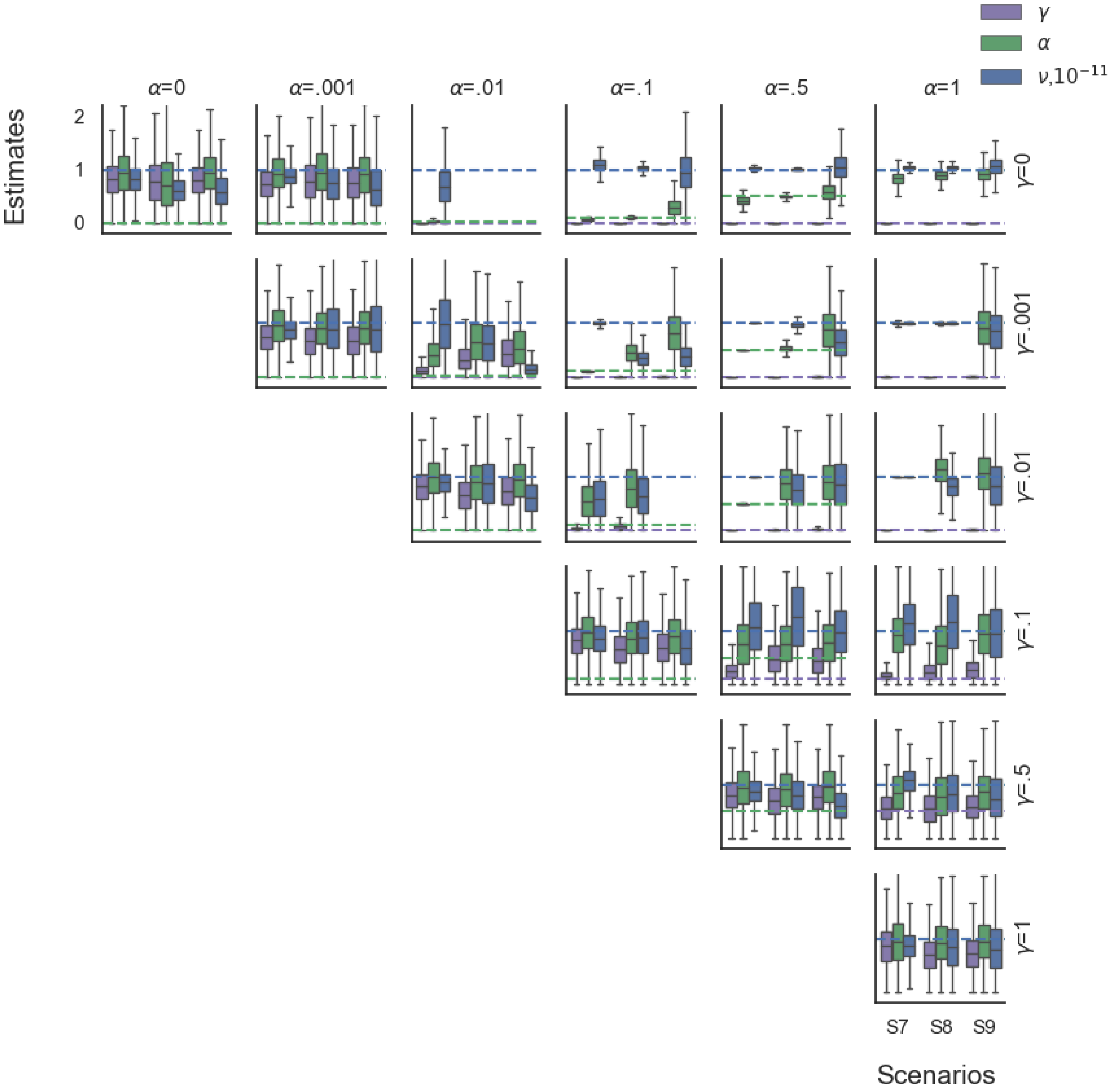

Fig. S29 – Parameter inference for the 36 parameter combinations under Scenarios 7, 8 and 9. The dashed lines in three colors denote three parameter values used to generate the data. Note that the scale of mutation rate is 10^{-11}. Box plots indicate the distribution of inferred parameter values, where the whiskers extend from the minimum to the maximum value. Some plots show no estimates for part of the parameter combinations, because there is no complete simulation under these parameter combinations after 10,000 attempts. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0$ with a crown age of 15 Myr. The clade-level carrying capacity is $K = 10, 30, 100$ for trees 7, 8, 9, respectively. Data are shown only for the informative parameter domain ($\gamma \leq \alpha$); when $\gamma > \alpha$, the parameter estimation show major bias and substantial variance.

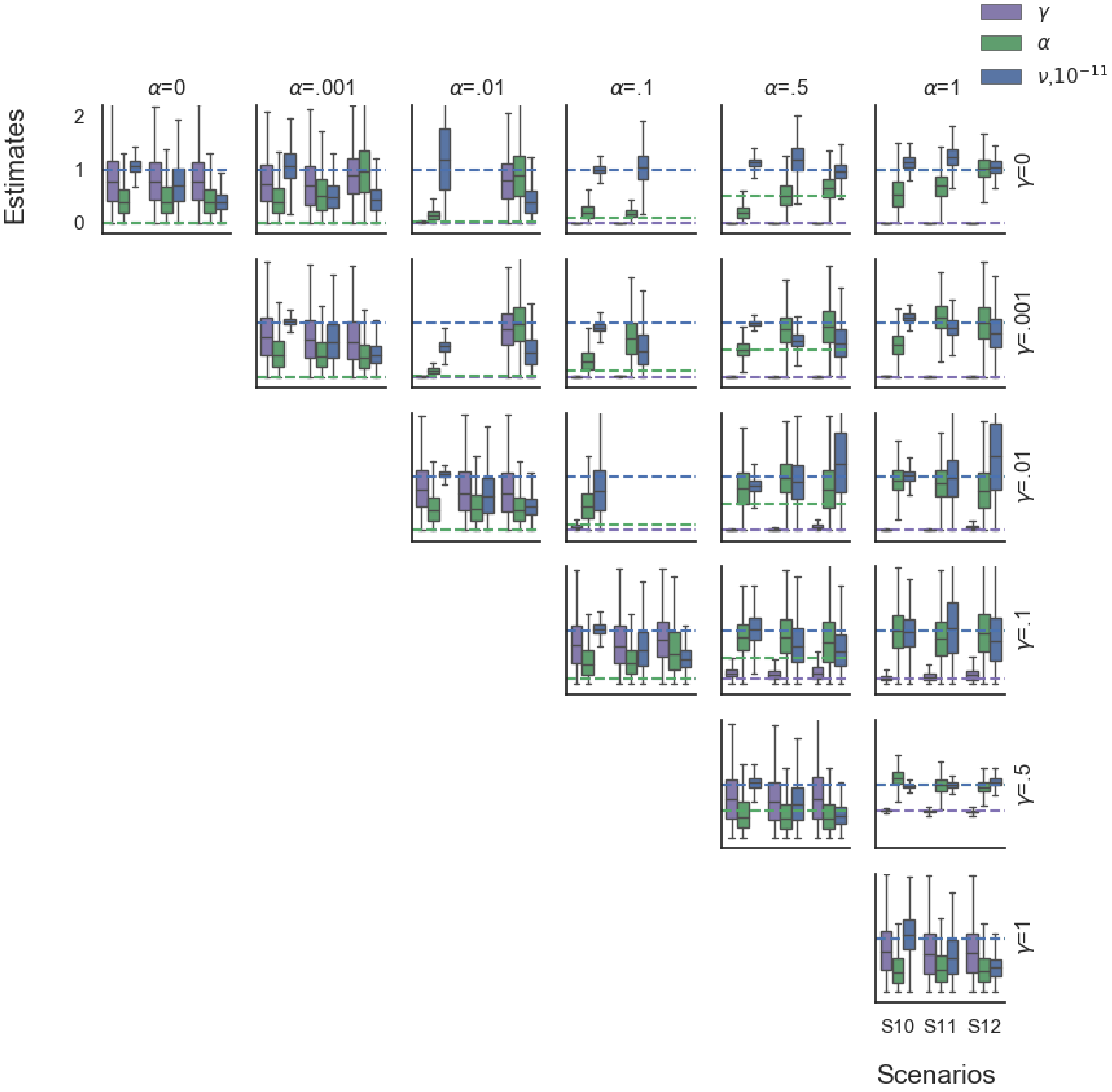

Fig. S30 – Parameter inference for the 36 parameter combinations under Scenarios 10, 11 and 12. The dashed lines in three colors denote three parameter values used to generate the data. Note that the scale of mutation rate is 10^{-11}. Box plots indicate the distribution of inferred parameter values, where the whiskers extend from the minimum to the maximum value. Some plots show no estimates for part of the parameter combinations, because there is no complete simulation under these parameter combinations after 10,000 attempts. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2$ with a crown age of 15 Myr. The clade-level carrying capacity is $K = 10, 30, 100$ for trees 10, 11, 12, respectively. Data are shown only for the informative parameter domain ($\gamma \leq \alpha$); when $\gamma > \alpha$, the parameter estimation show major bias and substantial variance.

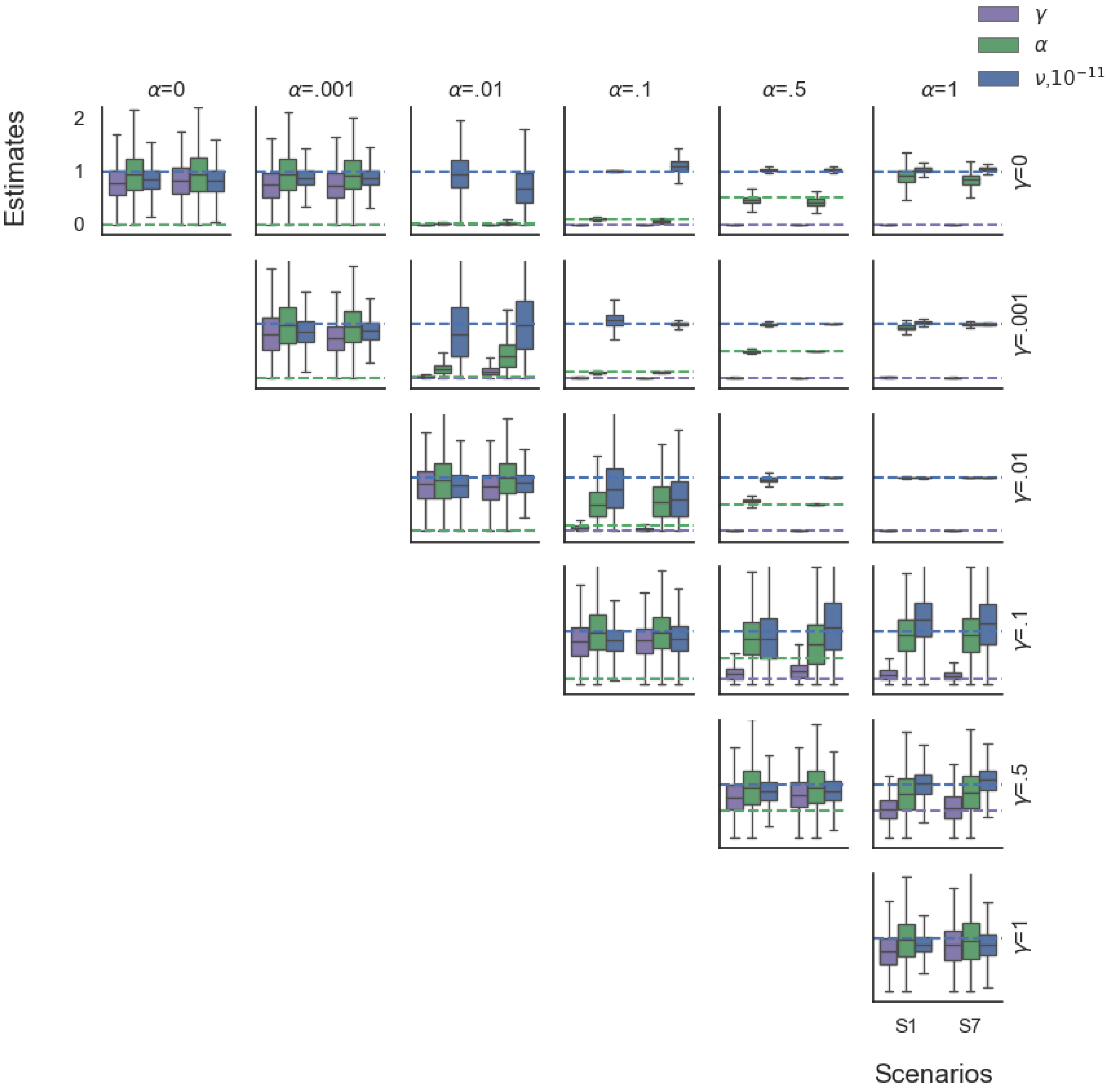

Fig. S31 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 1 and 7. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 10$ and $\lambda = 0.8, \mu = 0, K = 10$ with a crown age of 15 Myr, respectively.

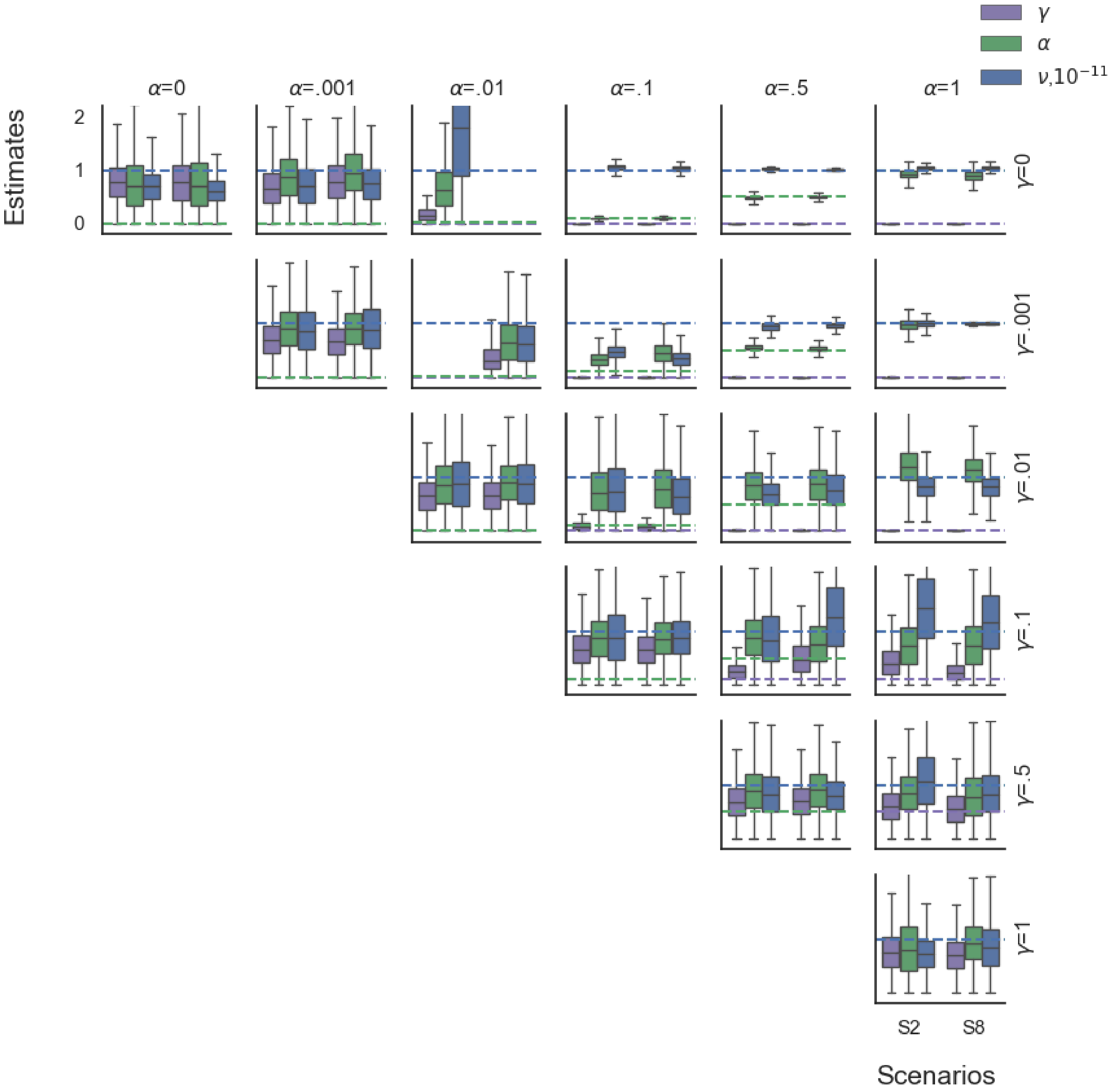

Fig. S32 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 2 and 8. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 30$ and $\lambda = 0.8, \mu = 0, K = 30$ with a crown age of 15 Myr, respectively.

Fig. S33 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 3 and 9. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 100$ and $\lambda = 0.8, \mu = 0, K = 100$ with a crown age of 15 Myr, respectively.

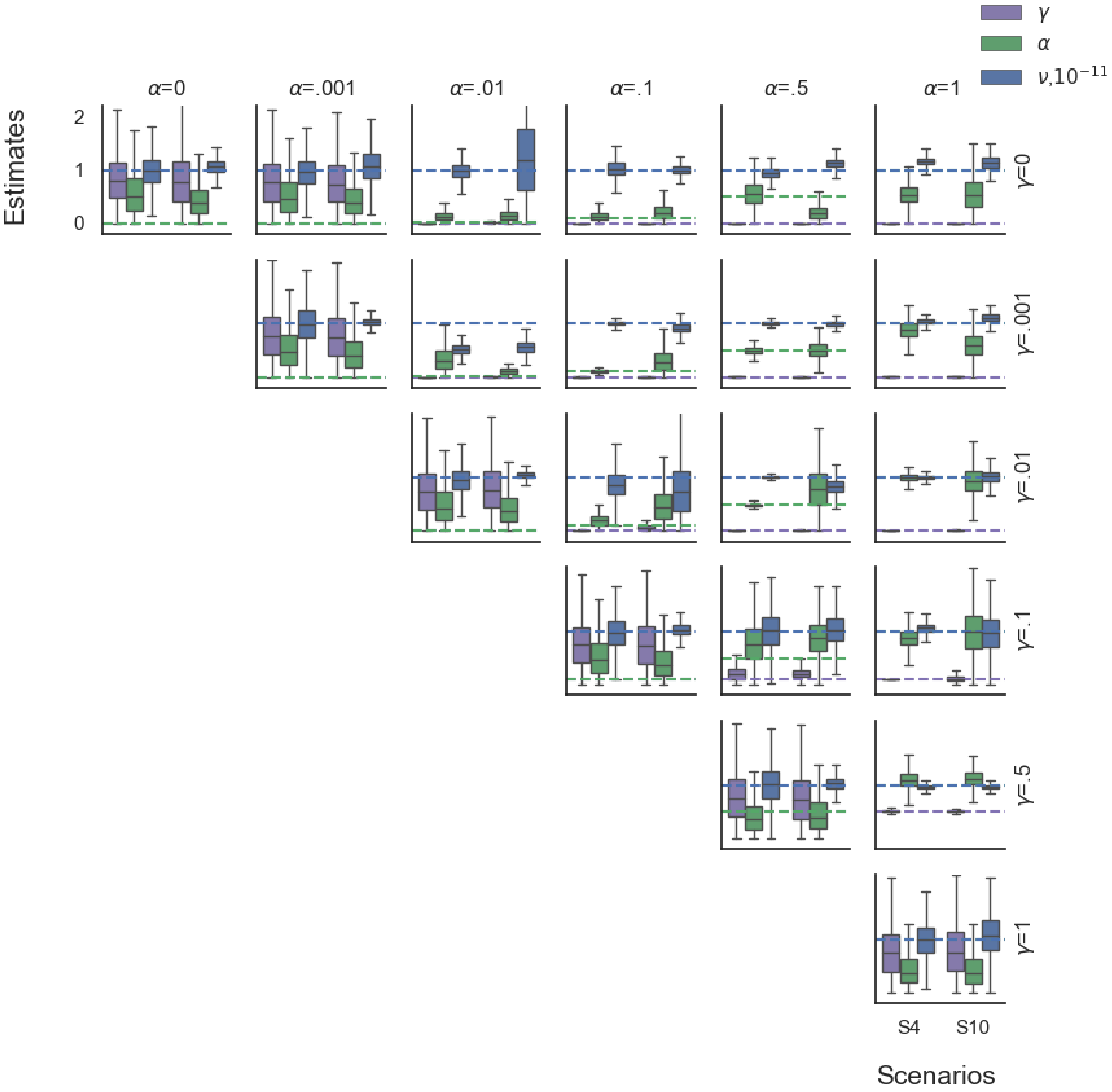

Fig. S34 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 4 and 10. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 10$ and $\lambda = 0.8, \mu = 0.2, K = 10$ with a crown age of 15 Myr, respectively.

Fig. S35 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 5 and 11. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 30$ and $\lambda = 0.8, \mu = 0.2, K = 30$ with a crown age of 15 Myr, respectively.

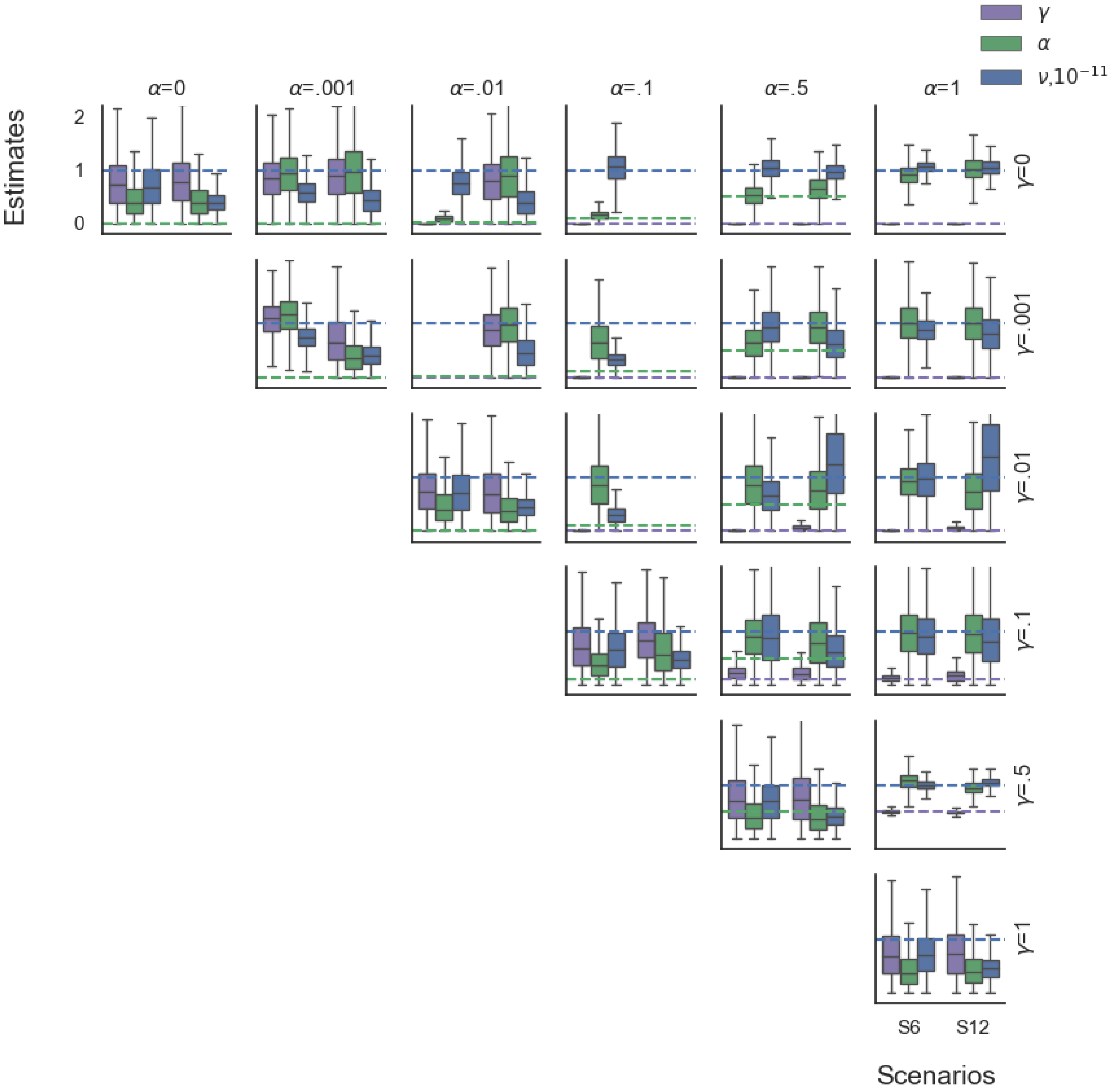

Fig. S36 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 6 and 12. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 100$ and $\lambda = 0.8, \mu = 0.2, K = 100$ with a crown age of 15 Myr, respectively.

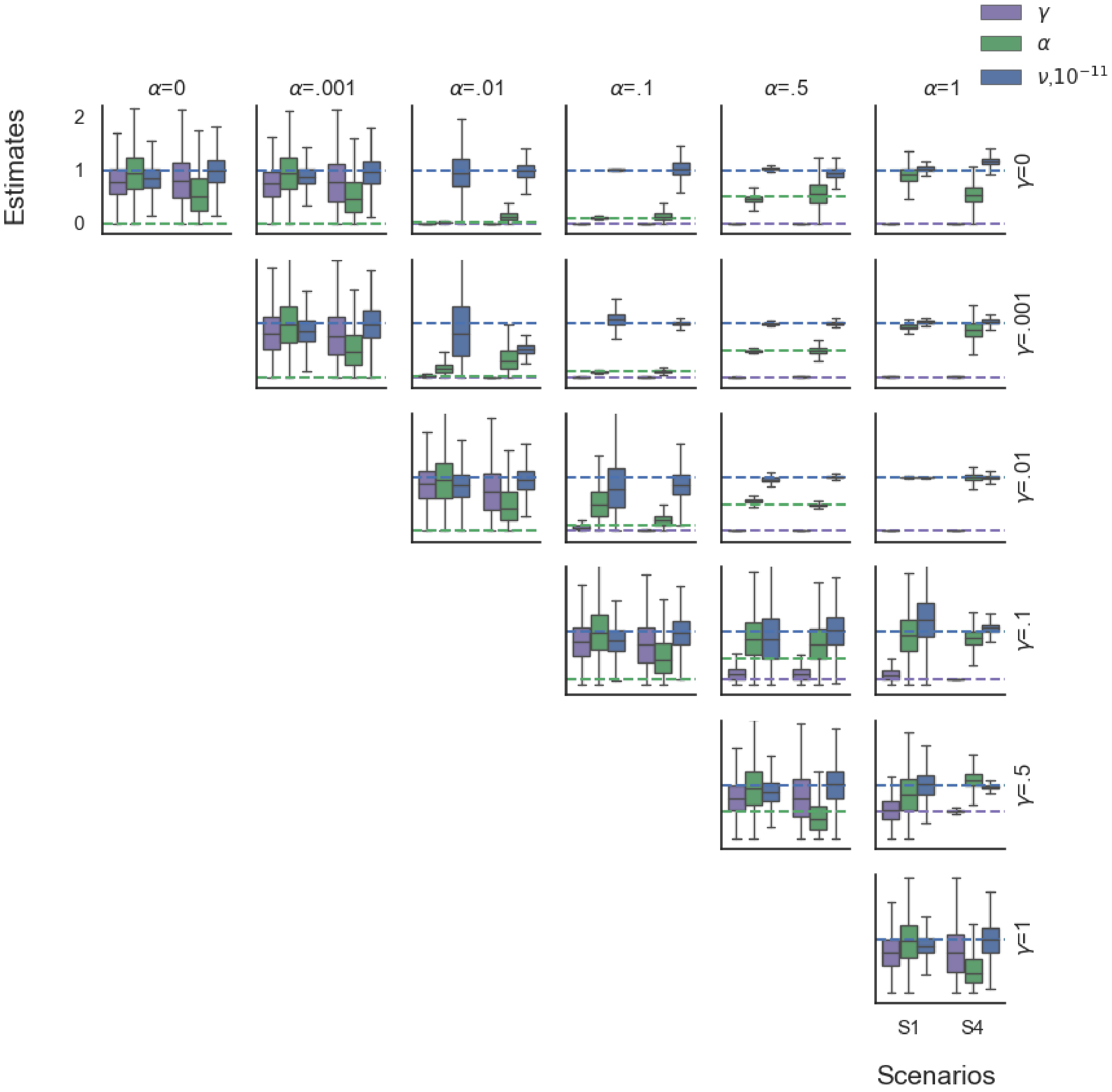

Fig. S37 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 1 and 4. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 10$ and $\lambda = 0.4, \mu = 0.2, K = 10$ with a crown age of 15 Myr, respectively.

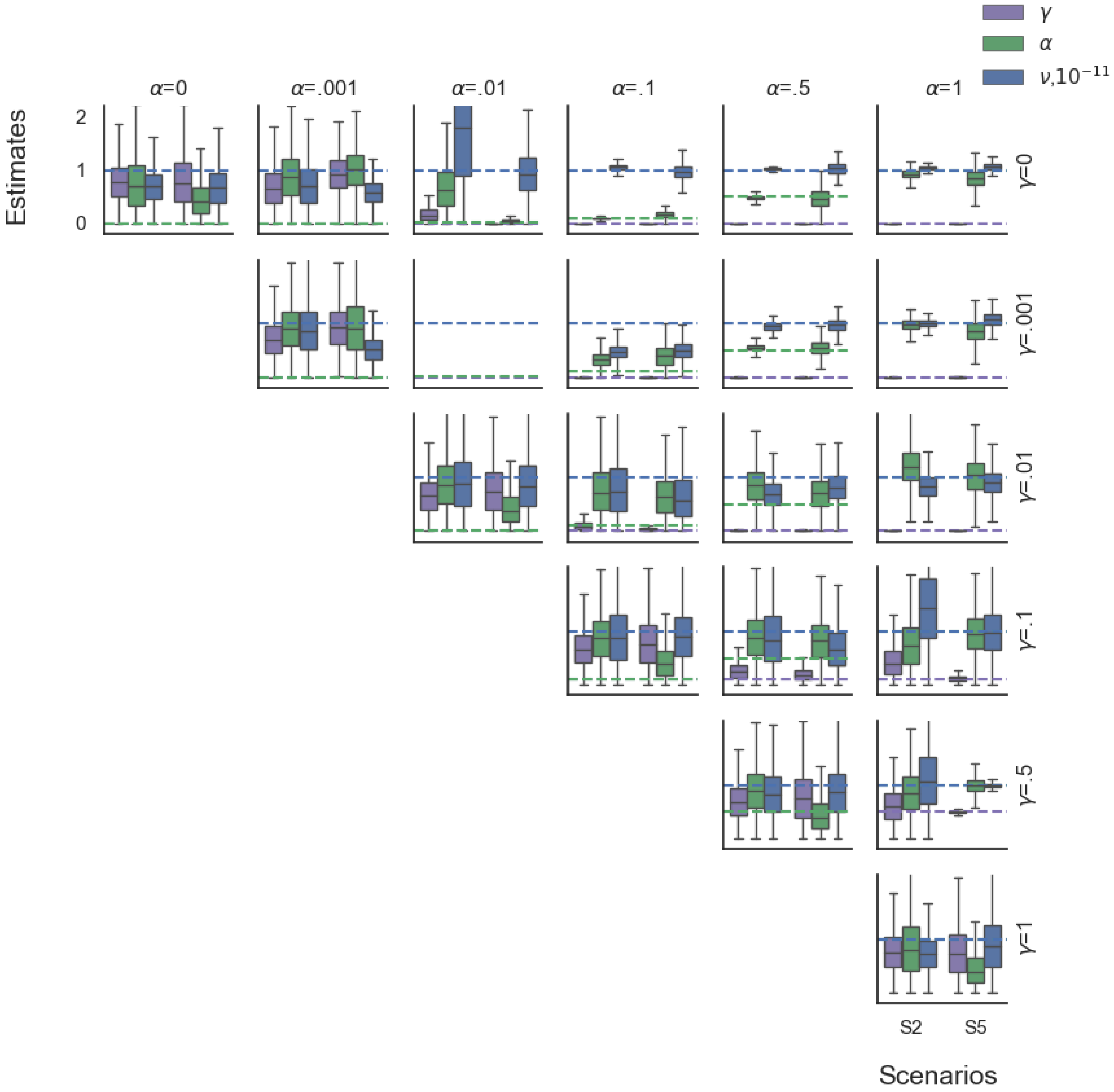

Fig. S38 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 2 and 5. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 30$ and $\lambda = 0.4, \mu = 0.2, K = 30$ with a crown age of 15 Myr, respectively.

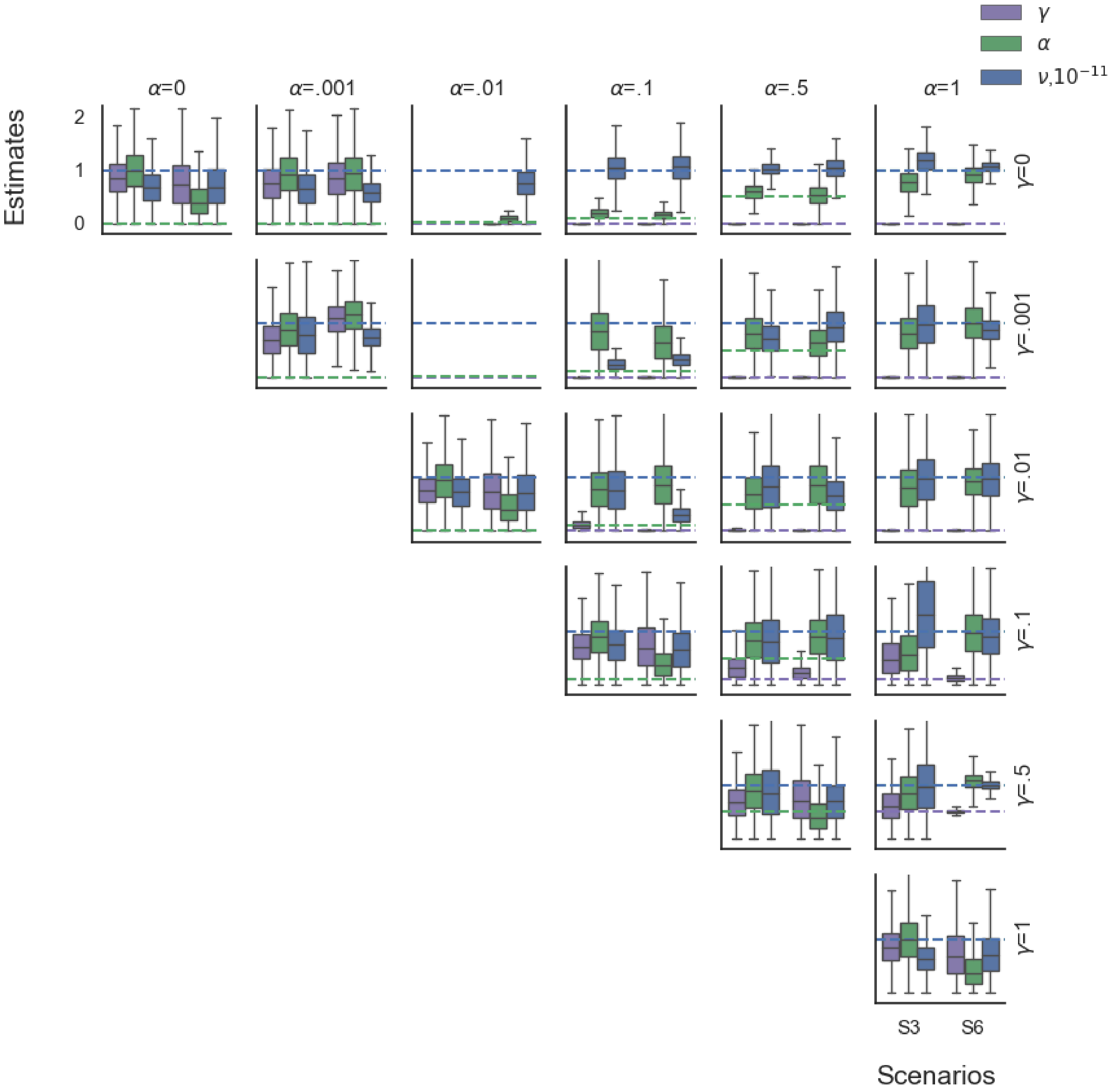

Fig. S39 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 3 and 6. The parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0, K = 100$ and $\lambda = 0.4, \mu = 0.2, K = 100$ with a crown age of 15 Myr, respectively.

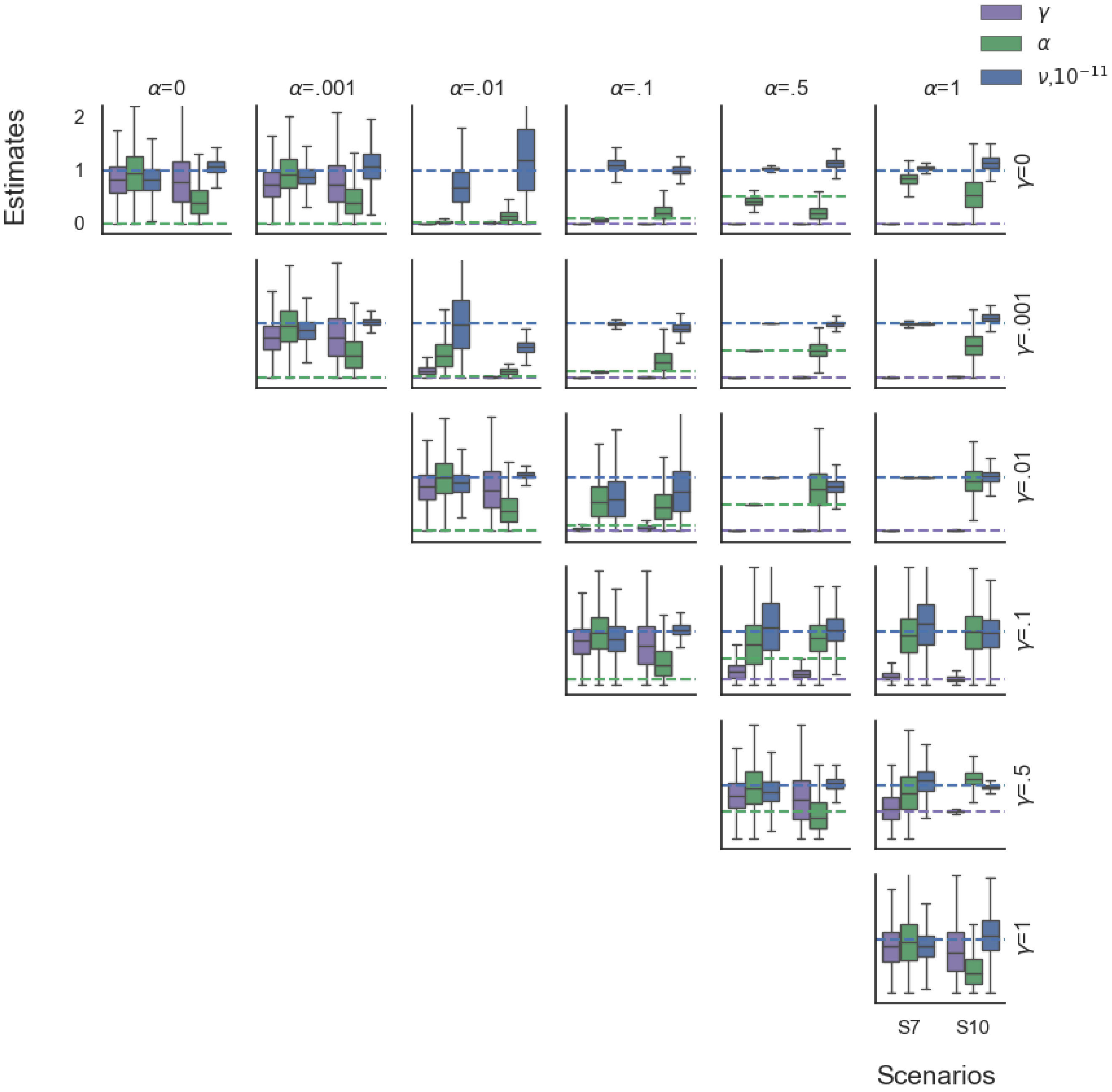

Fig. S40 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 7 and 10. The parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0, K = 10$ and $\lambda = 0.8, \mu = 0.2, K = 10$ with a crown age of 15 Myr, respectively.

Fig. S41 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 8 and 11. The parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0, K = 30$ and $\lambda = 0.8, \mu = 0.2, K = 30$ with a crown age of 15 Myr, respectively.

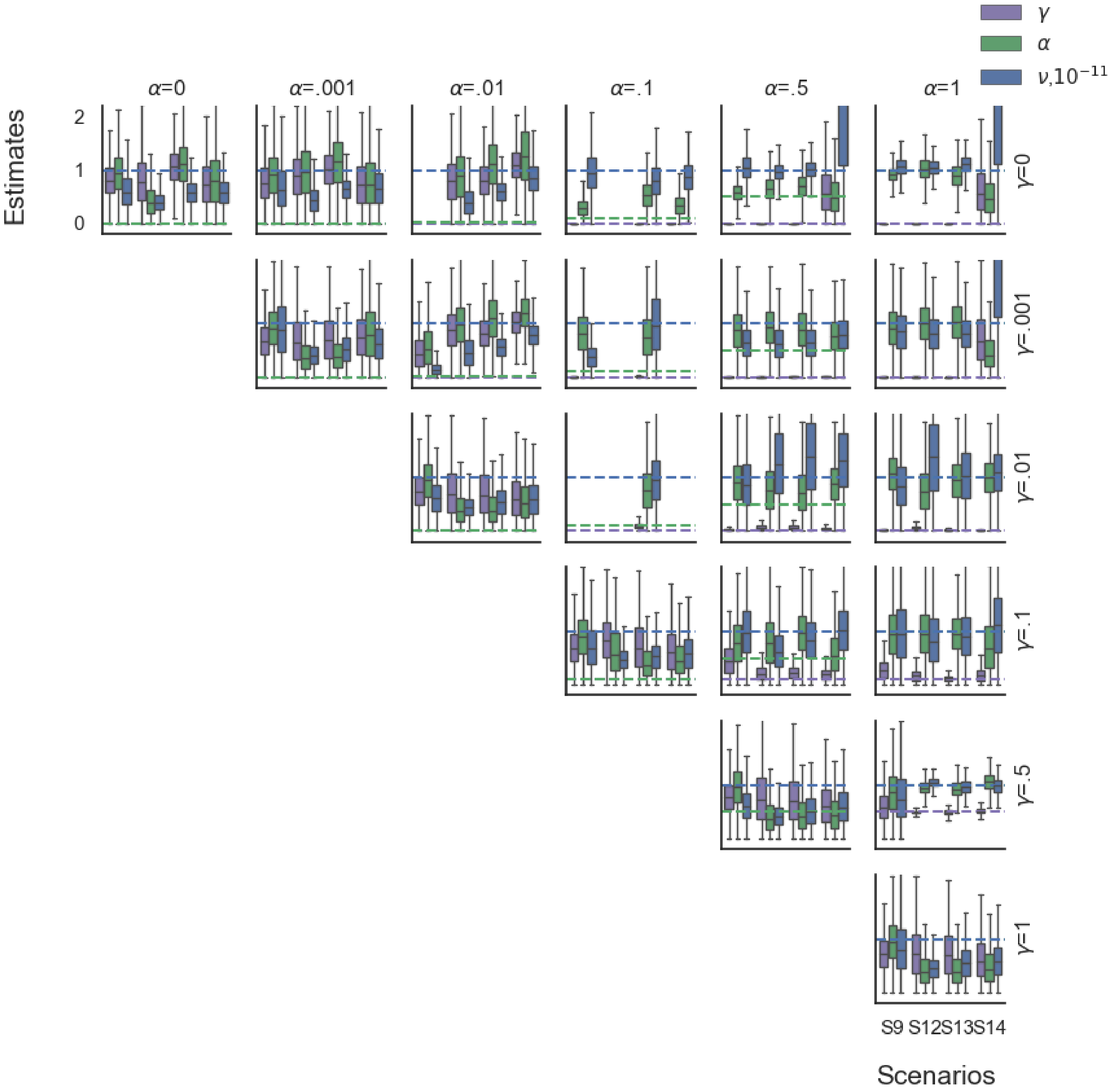

Fig. S42 – Parameter inference (similar to Fig. S28) for the 36 parameter combinations under Scenarios 9, 12, 13 and 14. The parameters used to generate the phylogenetic trees are $\lambda = 0.8, K = 100$ and $\mu = 0, 0.2, 0.4, 0.6$ with a crown age of 15 Myr, respectively.

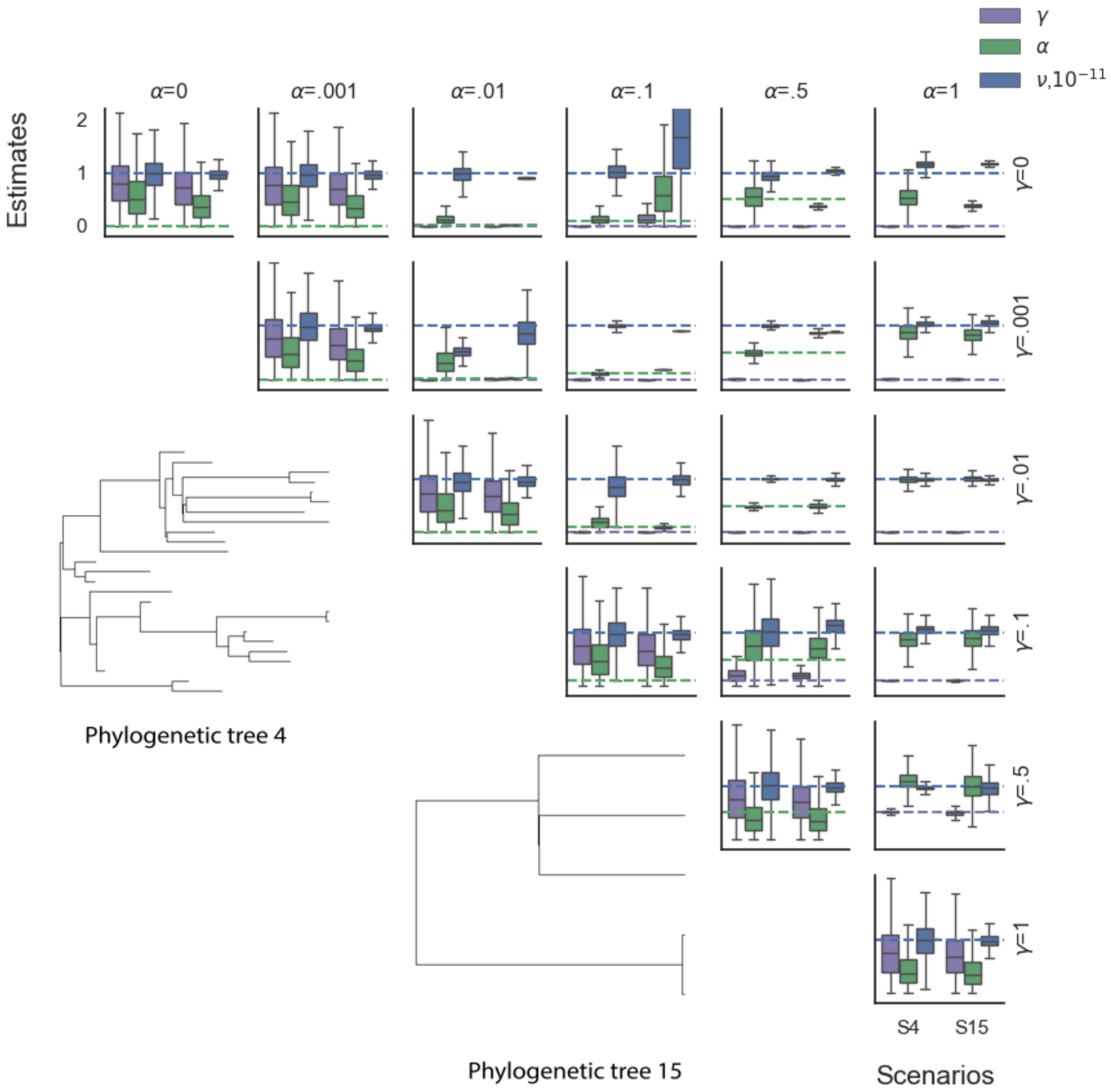

Fig. S43 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 4 and 15 (the corresponding phylogenetic trees are shown at the bottom left). The shared parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 10$ with a crown age of 15 Myr. Phylogenetic tree 15 is pruned from tree 4.

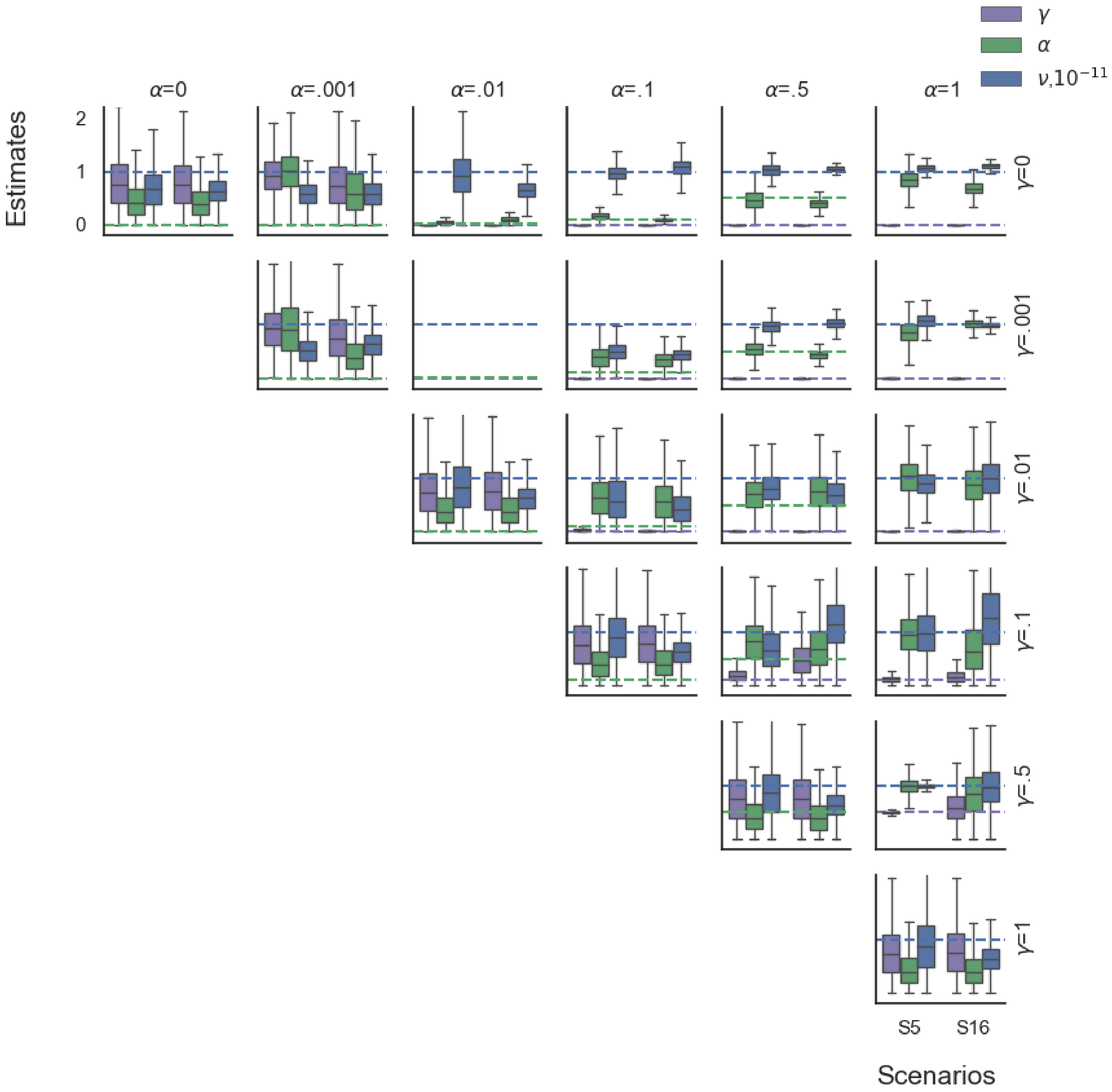

Fig. S44 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 5 and 16. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 30$ with a crown age of 15 Myr. Phylogenetic tree 16 is pruned from tree 5.

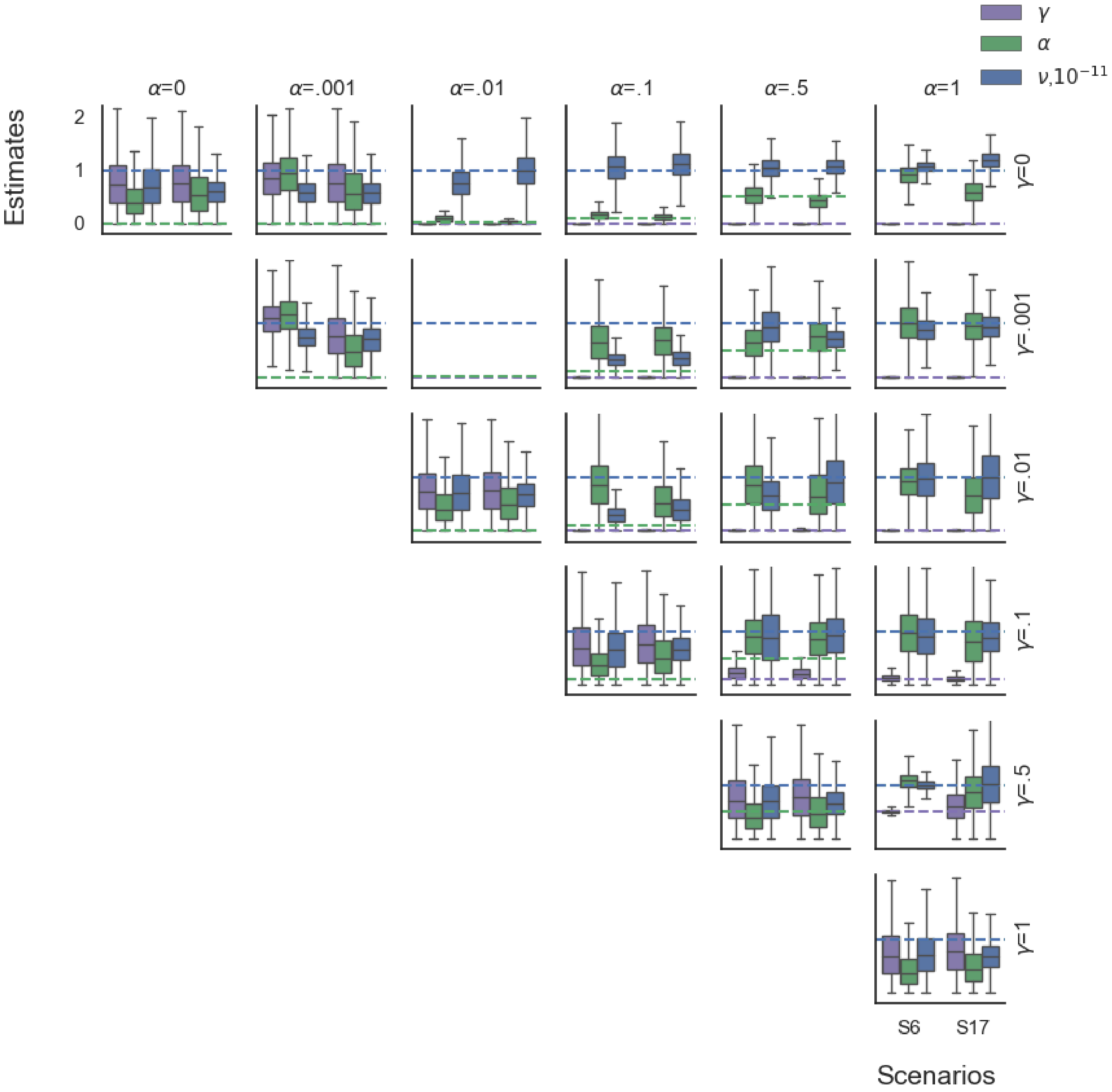

Fig. S45 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 6 and 17. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.4, \mu = 0.2, K = 100$ with a crown age of 15 Myr. Phylogenetic tree 17 is pruned from tree 6.

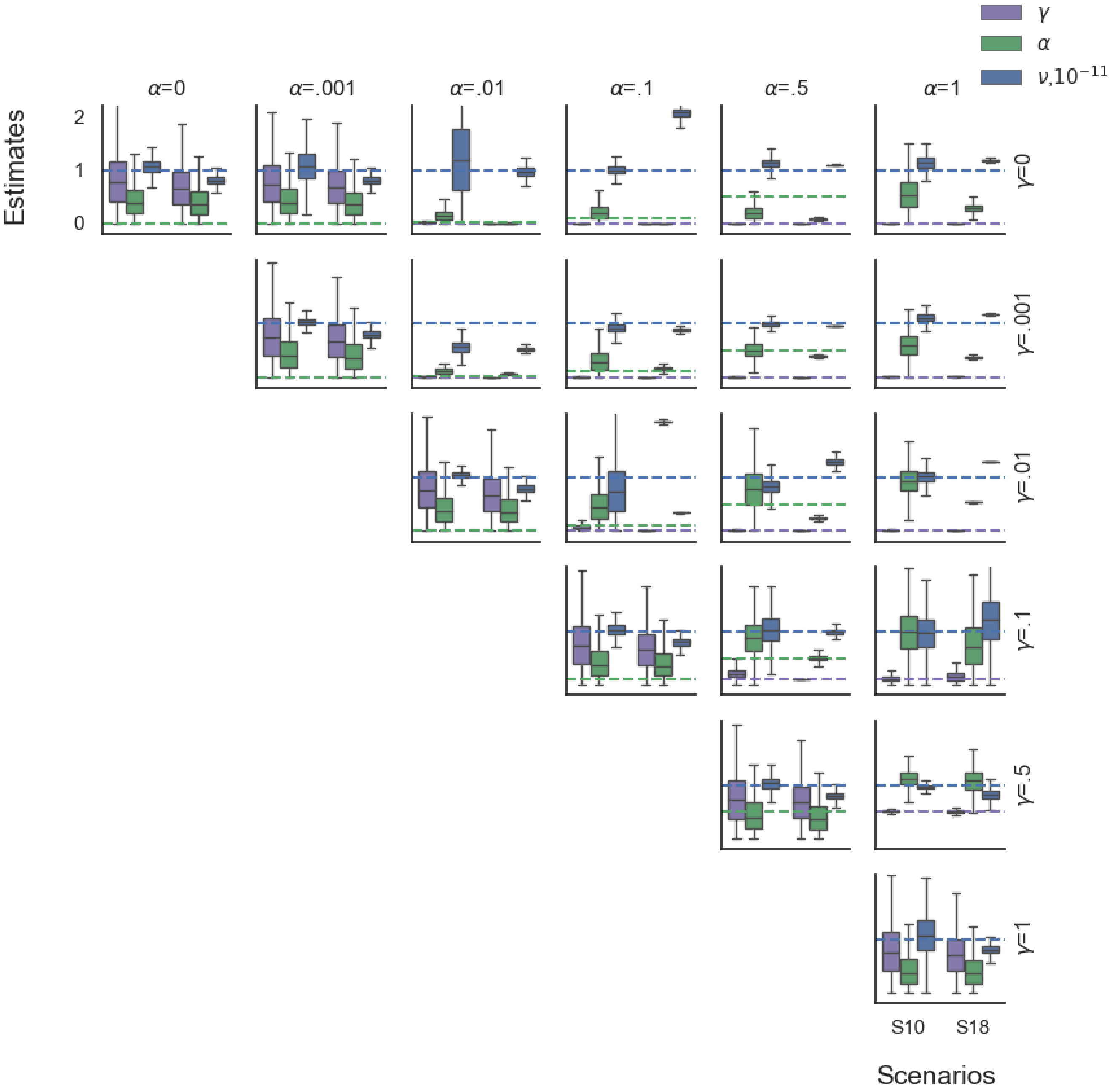

Fig. S46 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 10 and 18. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2, K = 10$ with a crown age of 15 Myr. Phylogenetic tree 18 is pruned from tree 10.

Fig. S47 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 11 and 19. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2, K = 30$ with a crown age of 15 Myr. Phylogenetic tree 19 is pruned from tree 11.

Fig. S48 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 12 and 20. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2, K = 100$ with a crown age of 15 Myr. Phylogenetic tree 20 is pruned from tree 12.

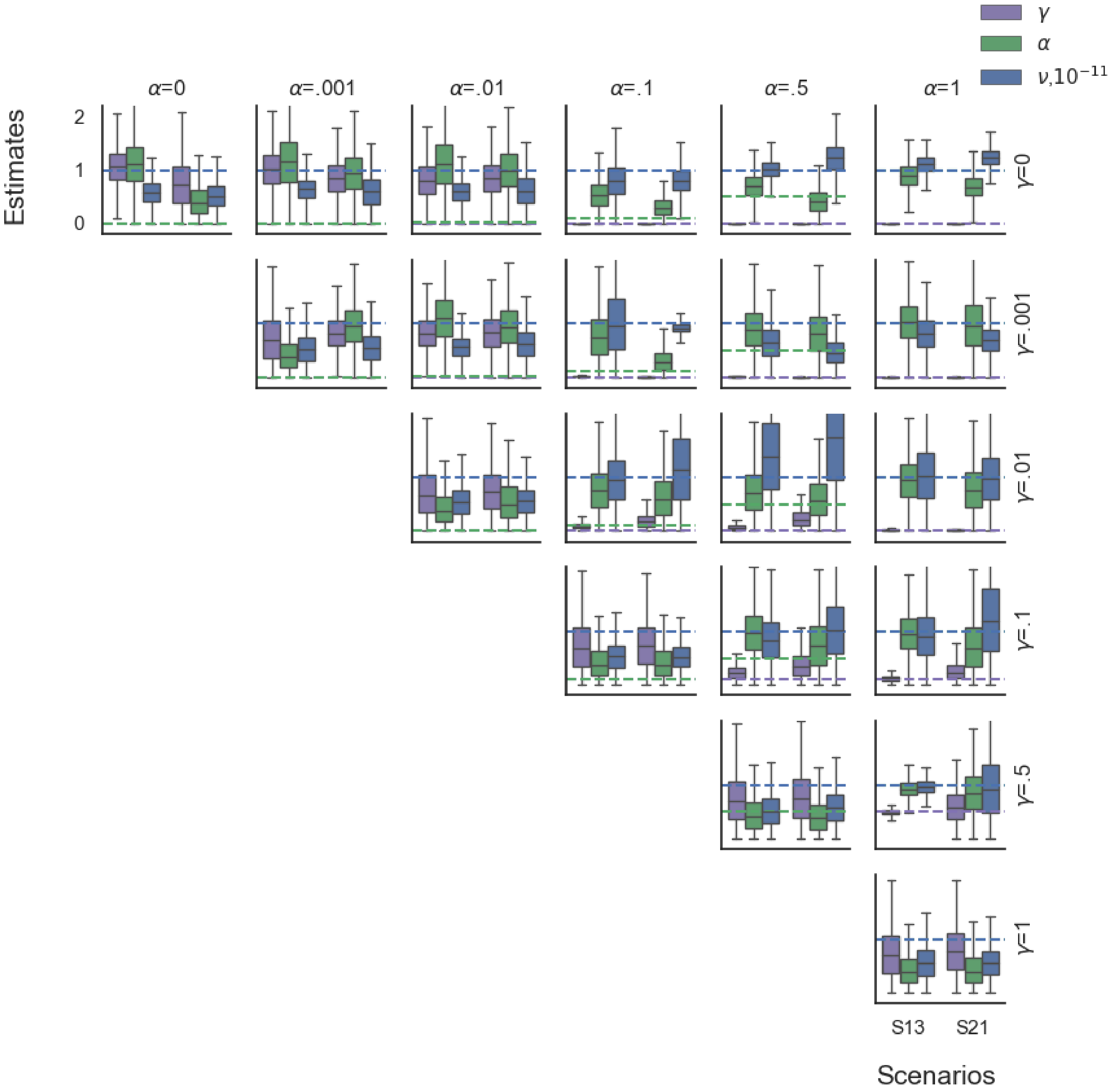

Fig. S49 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 13 and 21. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.4, K = 100$ with a crown age of 15 Myr. Phylogenetic tree 21 is pruned from tree 13.

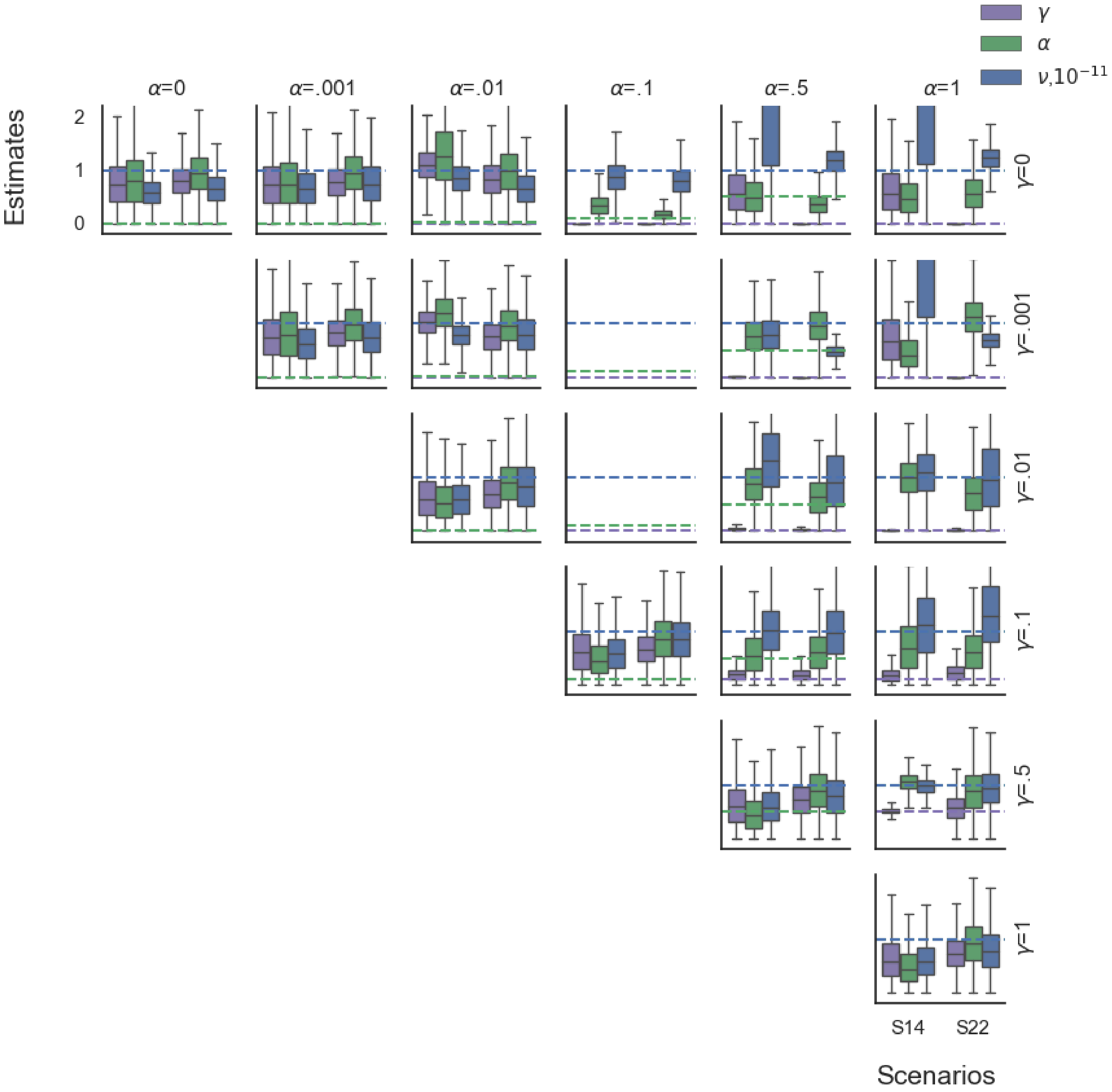

Fig. S50 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 14 and 22. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.6, K = 100$ with a crown age of 15 Myr. Phylogenetic tree 22 is pruned from tree 14.

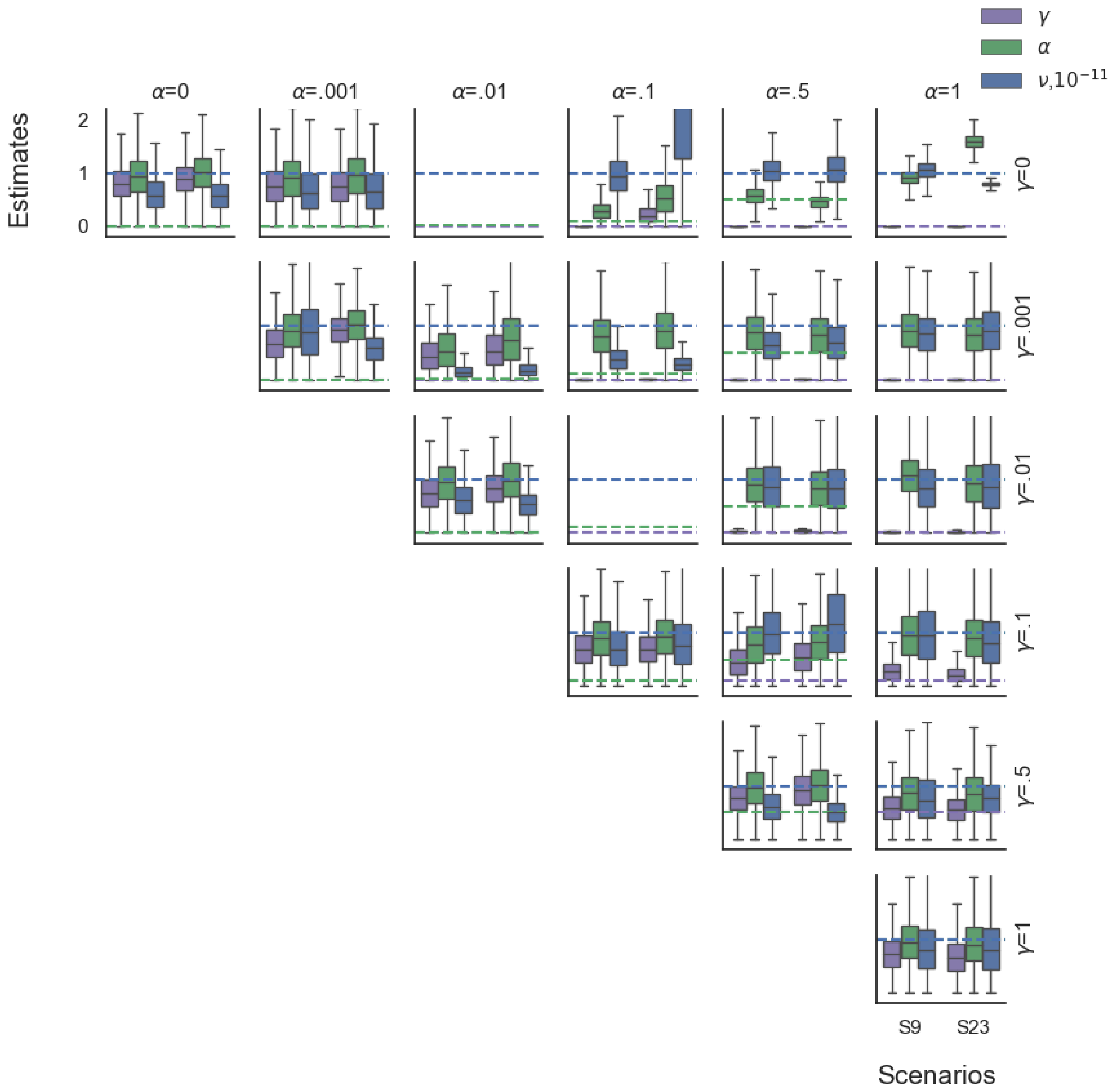

Fig. S51 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 9 and 23. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0, K = 100$ with a crown age of 15 Myr. In Scenario 23, the time scaling parameter $s = 10000$ is used to generate observed traits while $s = 20000$ is used in the inference algorithm.

Fig. S52 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 12 and 24. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2, K = 100$ with a crown age of 15 Myr. In Scenario 24, the time scaling parameter $s = 10000$ is used to generate observed traits while $s = 20000$ is used in the inference algorithm.

Fig. S53 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 13 and 25. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.4, K = 100$ with a crown age of 15 Myr. In Scenario 25, the time scaling parameter $s = 10000$ is used to generate observed traits while $s = 20000$ is used in the inference algorithm.

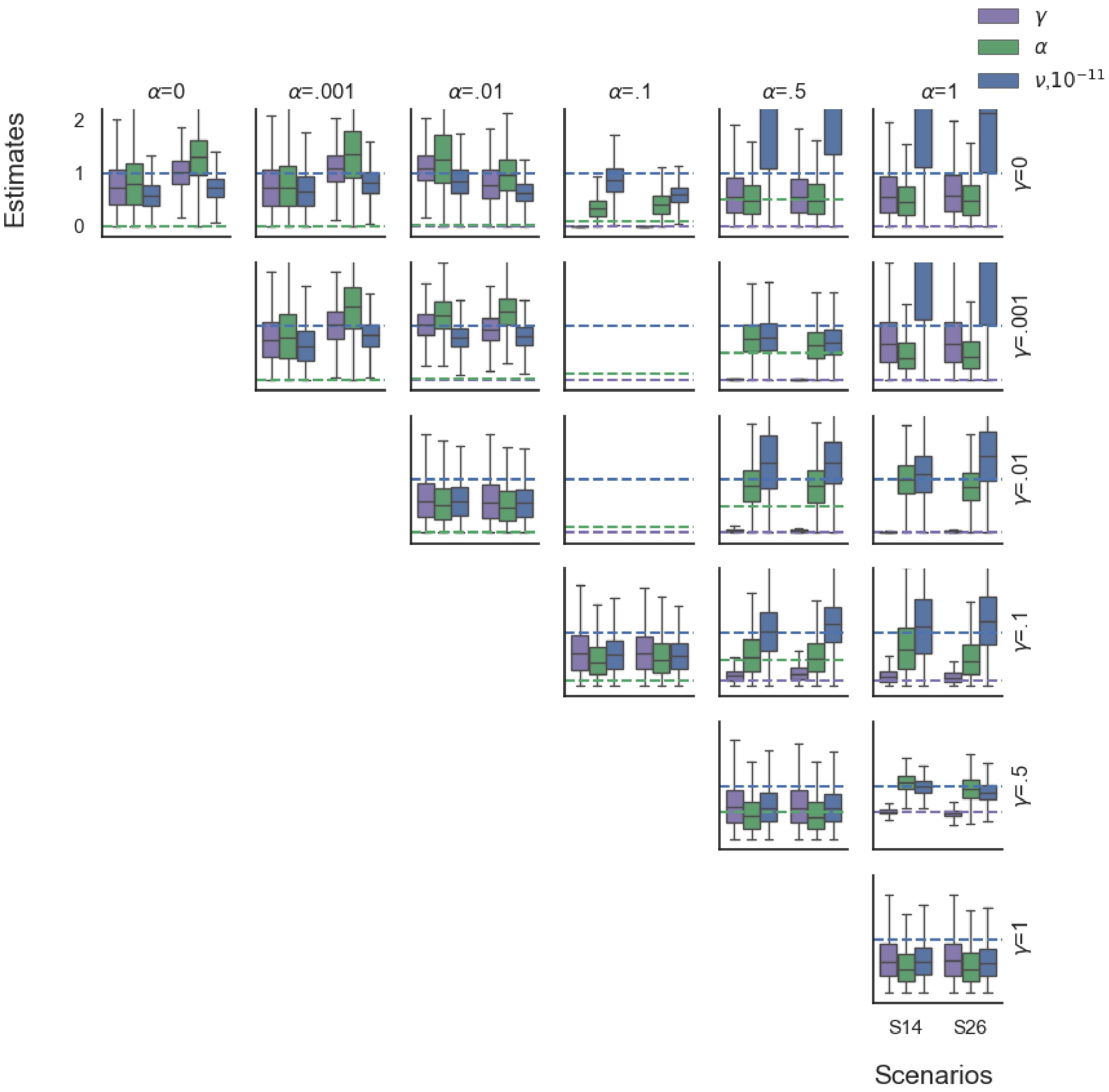

Fig. S54 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 14 and 26. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.6, K = 100$ with a crown age of 15 Myr. In Scenario 26, the time scaling parameter $s = 10000$ is used to generate observed traits while $s = 20000$ is used in the inference algorithm.

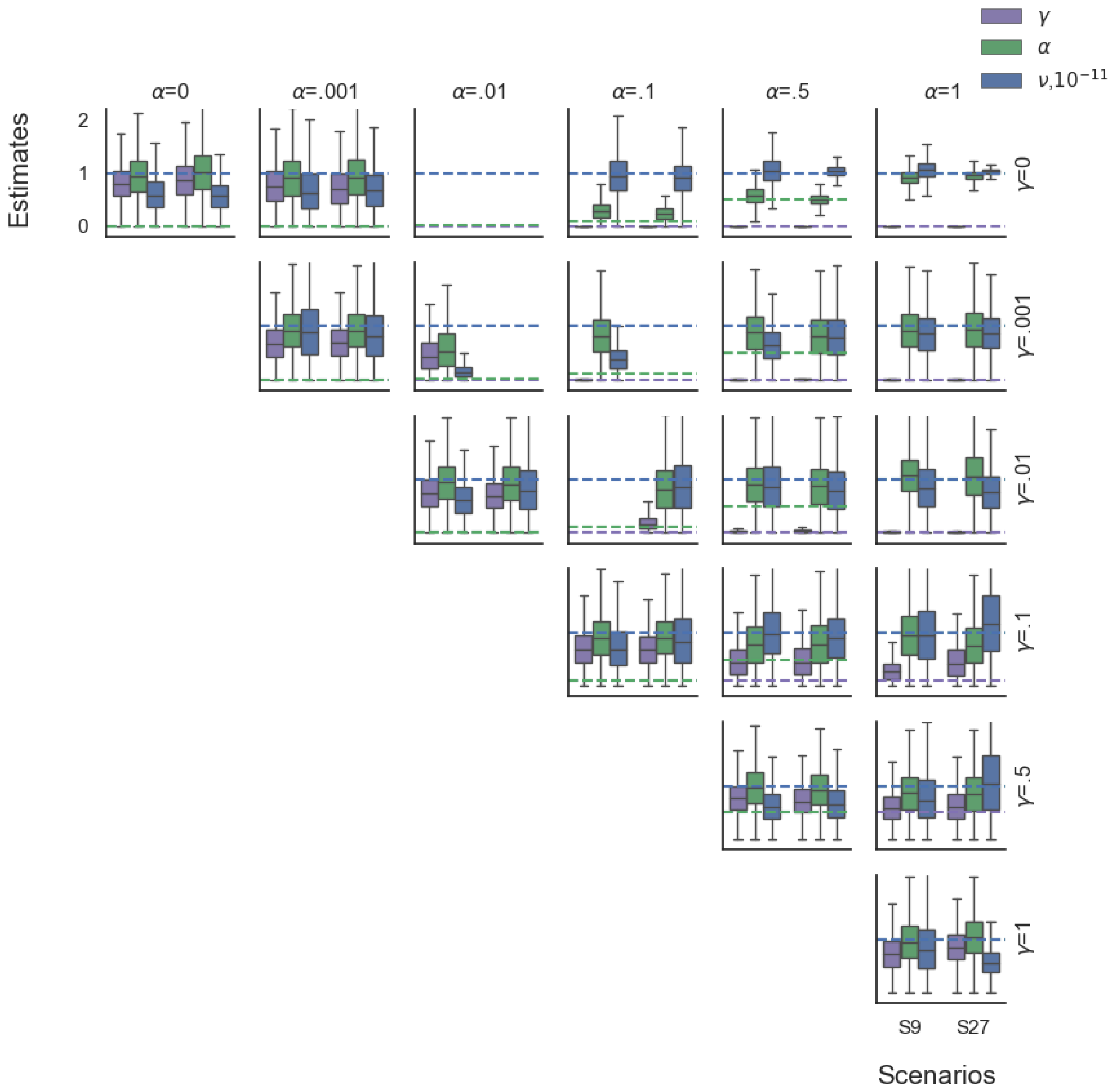

Fig. S55 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 9 and 27. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0, K = 100$ with a crown age of 15 Myr. In Scenario 9, the time scaling parameter $s = 10000$ is used both in generating observed traits and in inference algorithm while in Scenario 27 $s = 20000$.

Fig. S56 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 12 and 28. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.2, K = 100$ with a crown age of 15 Myr. In Scenario 12, the time scaling parameter $s = 10000$ is used both in generating observed traits and in inference algorithm while in Scenario 28 $s = 20000$.

Fig. S57 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 13 and 29. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.4, K = 100$ with a crown age of 15 Myr. In Scenario 13, the time scaling parameter $s = 10000$ is used both in generating observed traits and in inference algorithm while in Scenario 29 $s = 20000$.

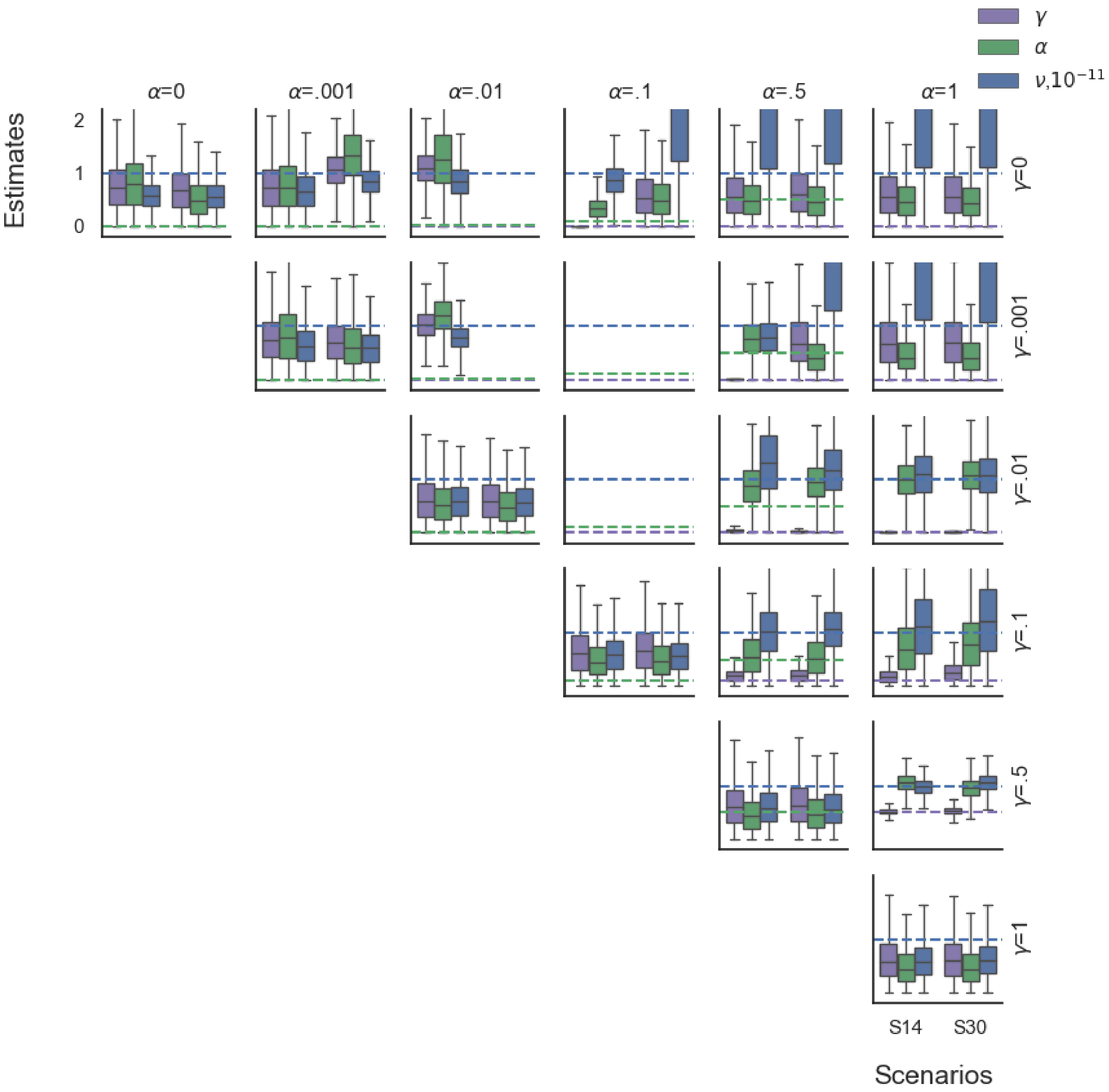

Fig. S58 – Parameter inference (similar to Figure S28) for the 36 parameter combinations under Scenarios 14 and 30. The shared parameters used to generate the phylogenetic trees are $\lambda = 0.8, \mu = 0.6, K = 100$ with a crown age of 15 Myr. In Scenario 14, the time scaling parameter $s = 10000$ is used both in generating observed traits and in inference algorithm while in Scenario 30 $s = 20000$.

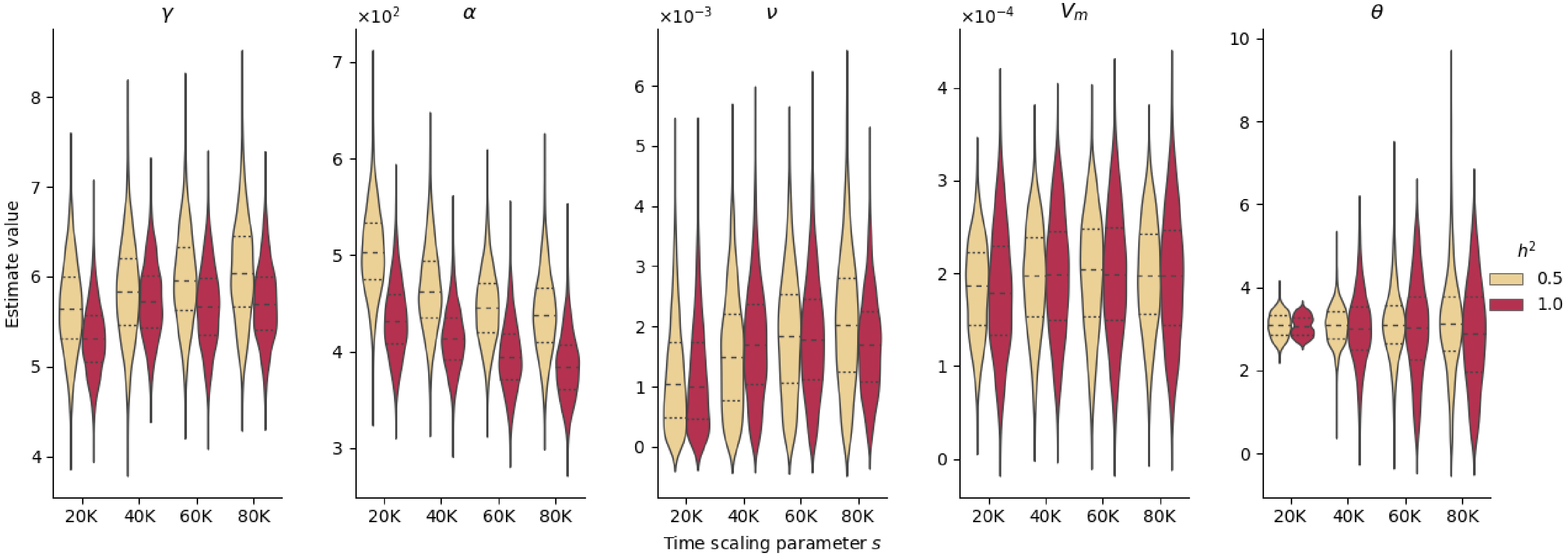

Fig. S59 – Parameter inference under 8 scenarios on the *Log*10 transformed body length to test the influence of the number of time steps and heritability on parameter estimation using PICs only as the summary statistic. The three dashed lines in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

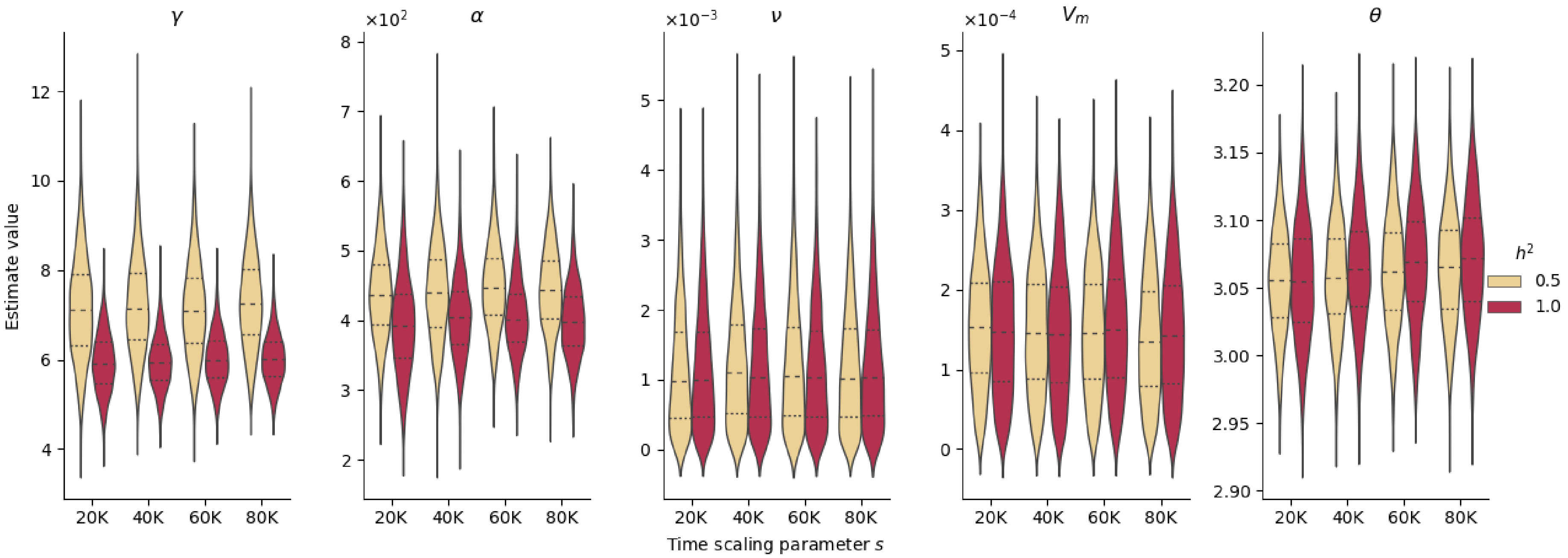

Fig. S60 – Parameter inference under 8 scenarios on the *Log*10 transformed body length to test the influence of the number of time steps and heritability on parameter estimation using UMTD+PICs as the summary statistic. The three dashed lines in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

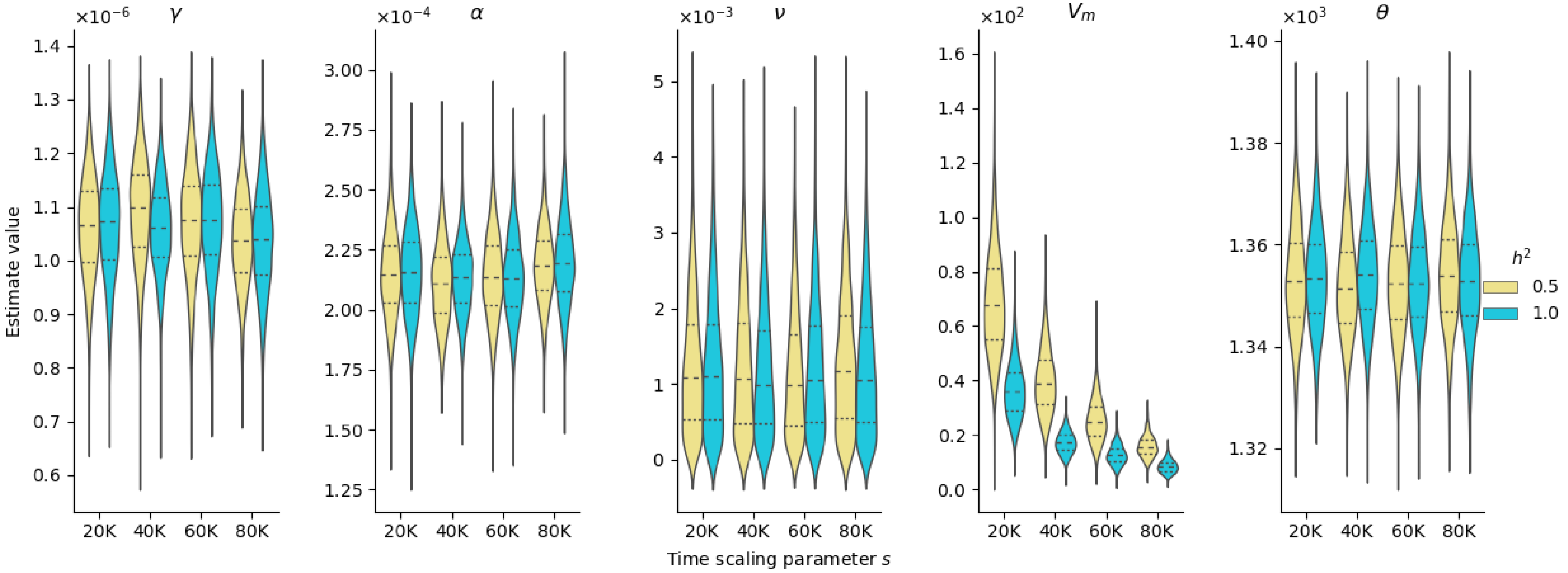

Fig. S61 – Parameter inference under 8 scenarios on the untransformed body length to test the influence of the number of time steps and heritability on parameter estimation using SMTD only as the summary statistic. The three dashed lines in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

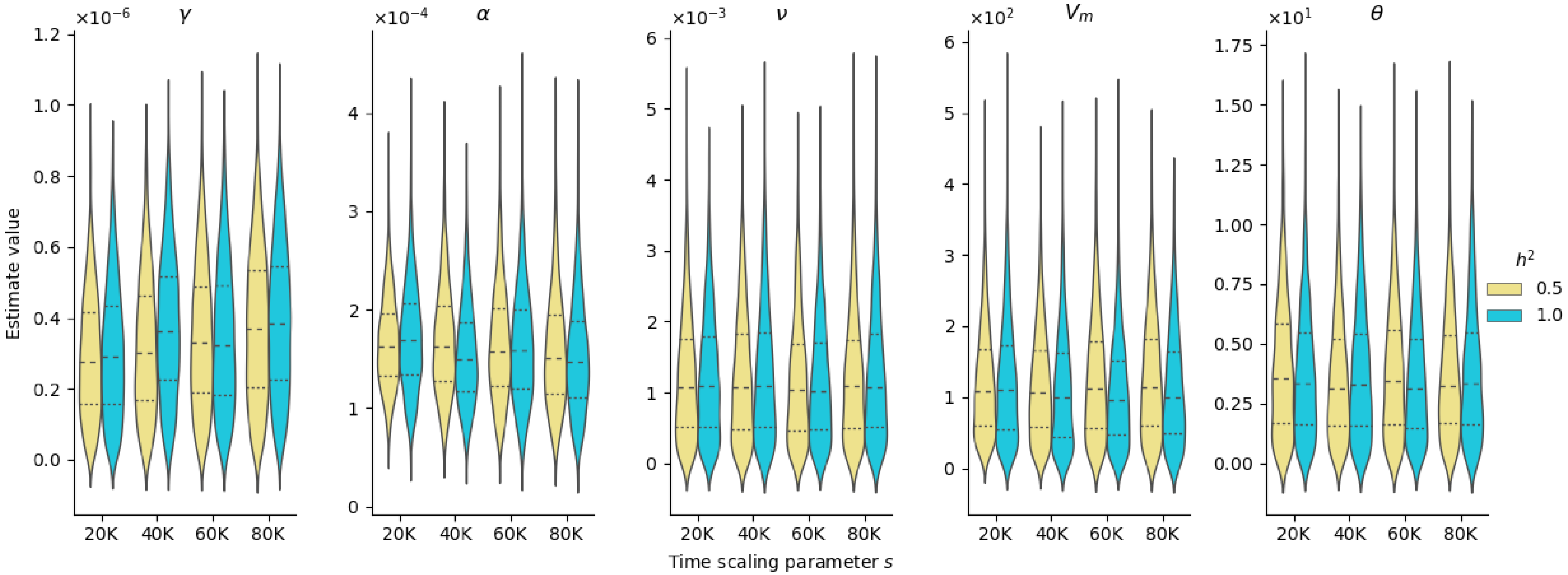

Fig. S62 – Parameter inference under 8 scenarios on the untransformed body length to test the influence of the number of time steps and heritability on parameter estimation using PICs only as the summary statistic. The three dashed lines in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

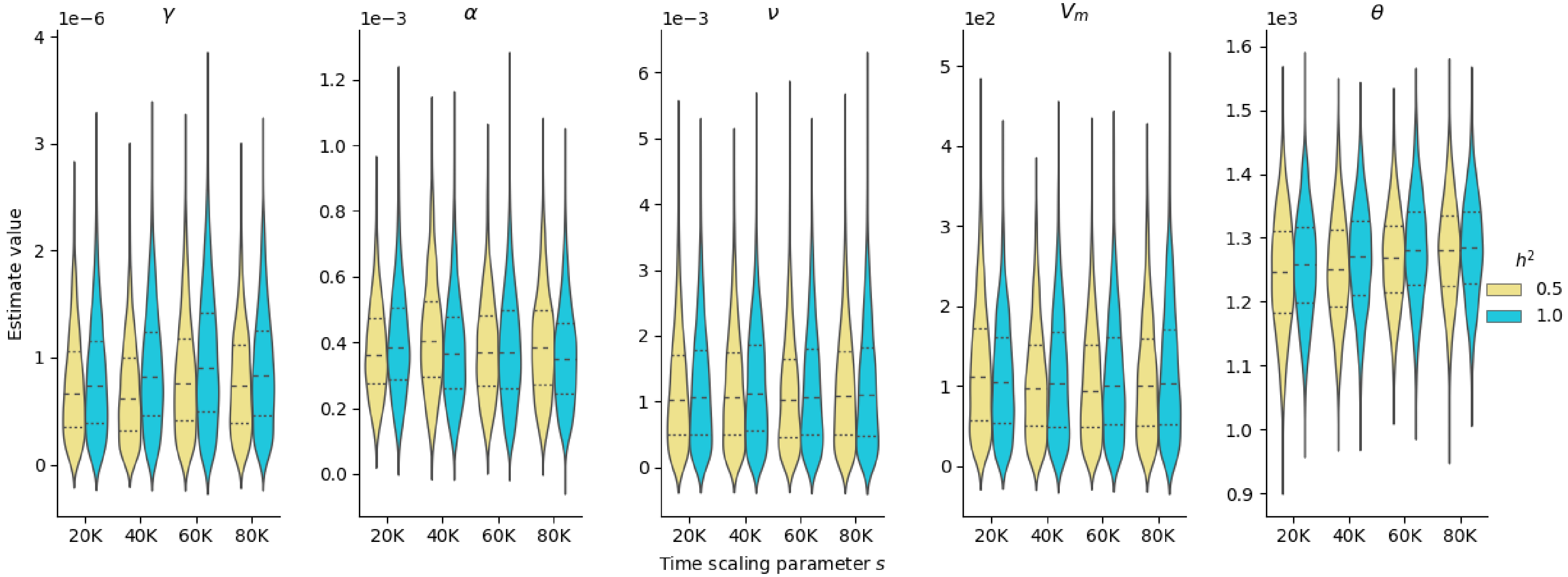

Fig. S63 – Parameter inference under 8 scenarios on the untransformed body length to test the influence of the number of time steps and heritability on parameter estimation using UMTD+PICs as the summary statistic. The three dashed lines in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

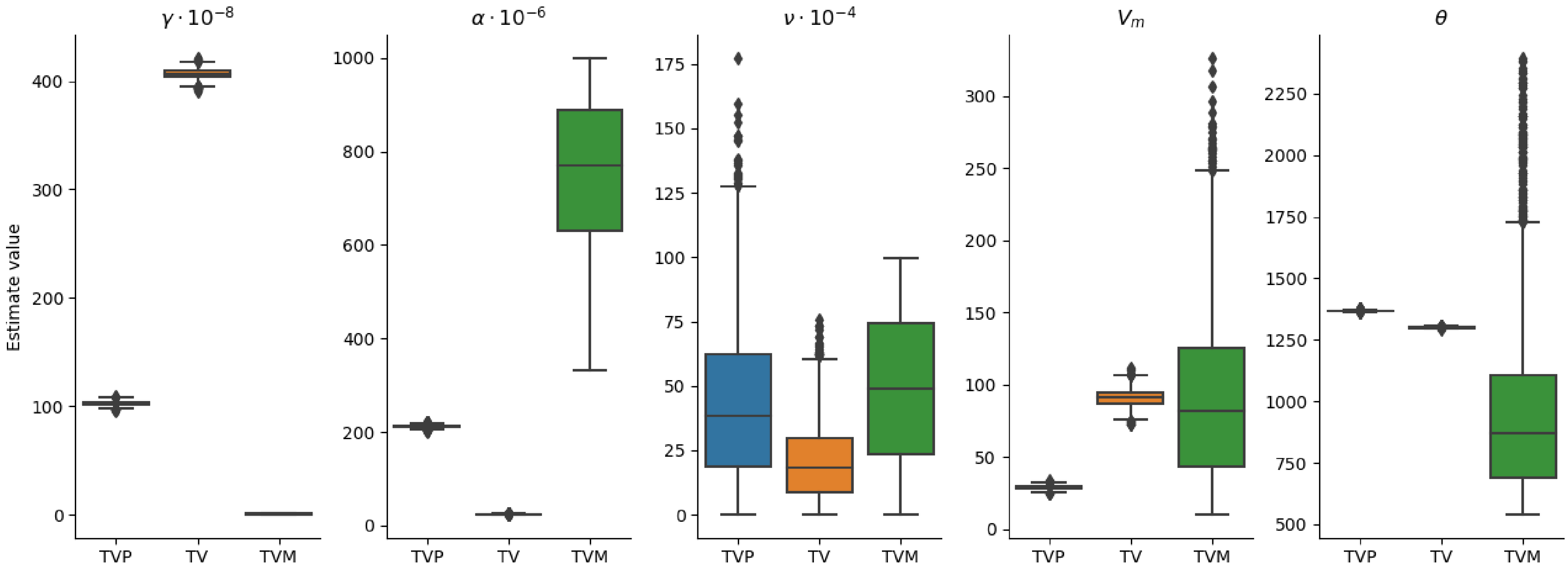

Fig. S64 – Parameter inference comparison among the three models using SMTD as the summary statistic on the untransformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

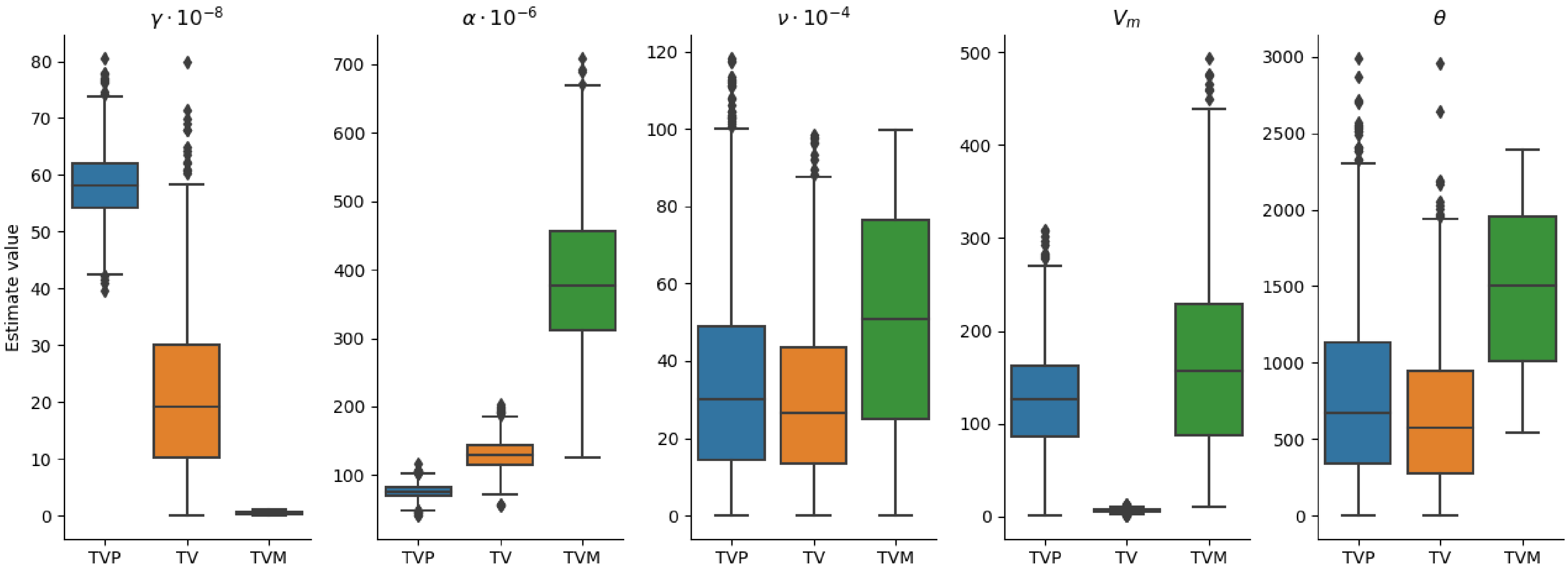

Fig. S65 – Parameter inference comparison among the three models using PICs only as the summary statistic on the untransformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

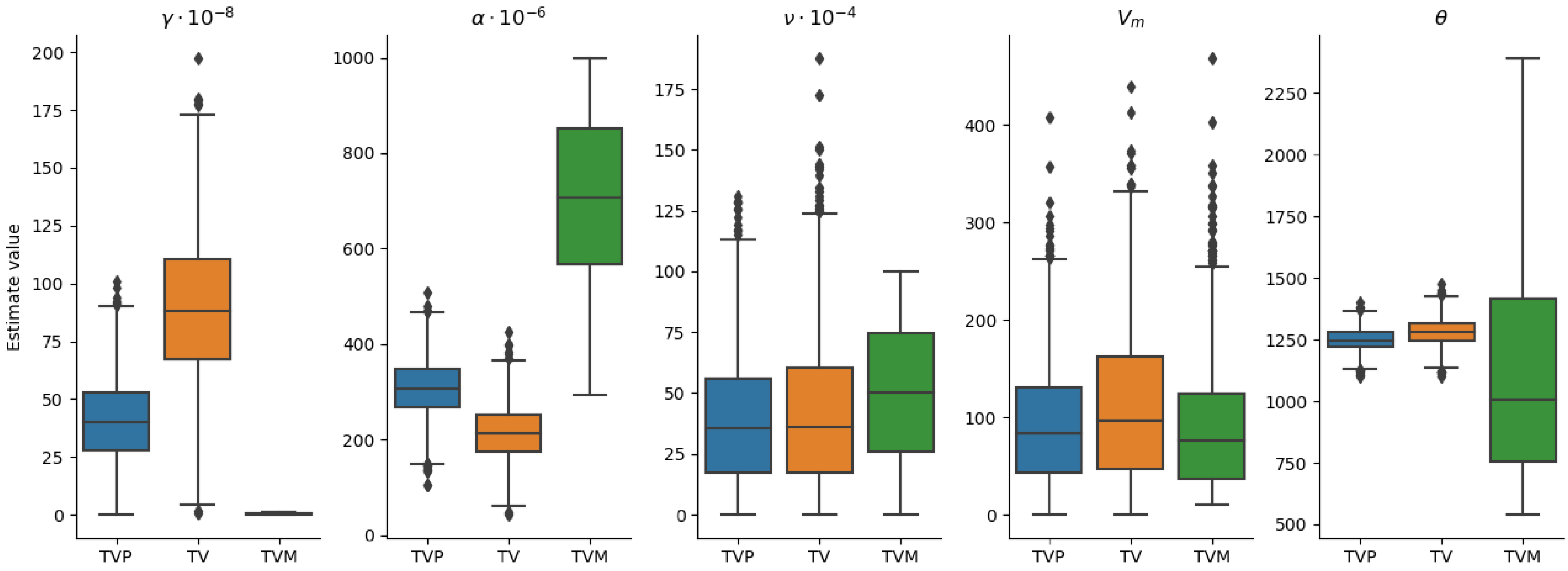

Fig. S66 – Parameter inference comparison among the three models using UMTD+PICs as the summary statistic on the untransformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

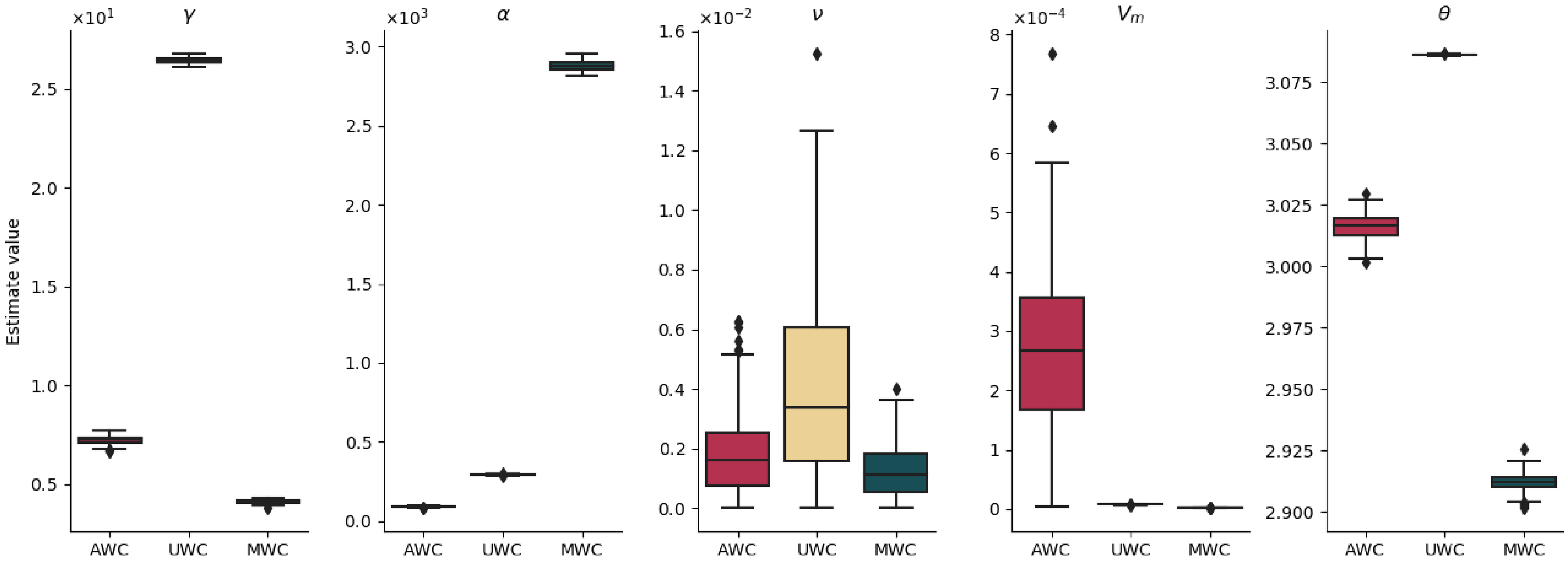

Fig. S67 – Parameter inference comparison among the three models using SMTD as the summary statistic on the log-transformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

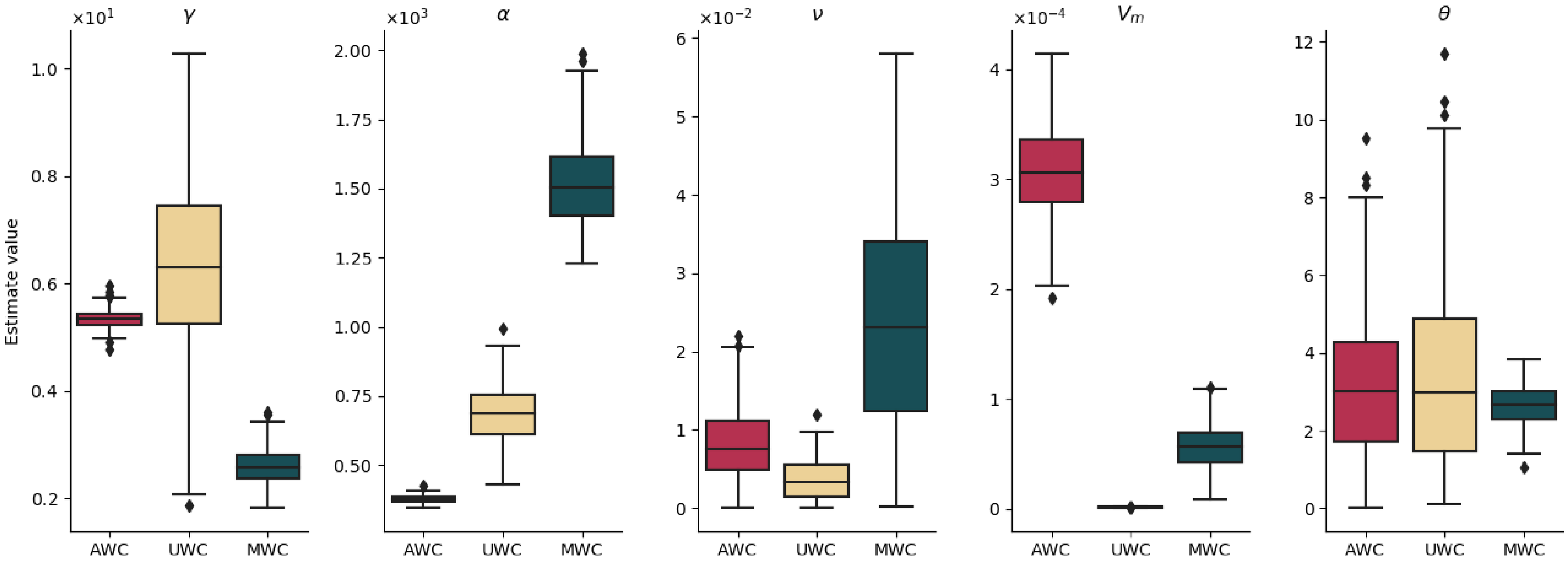

Fig. S68 – Parameter inference comparison among the three models using PICs only as the summary statistic on the log-transformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

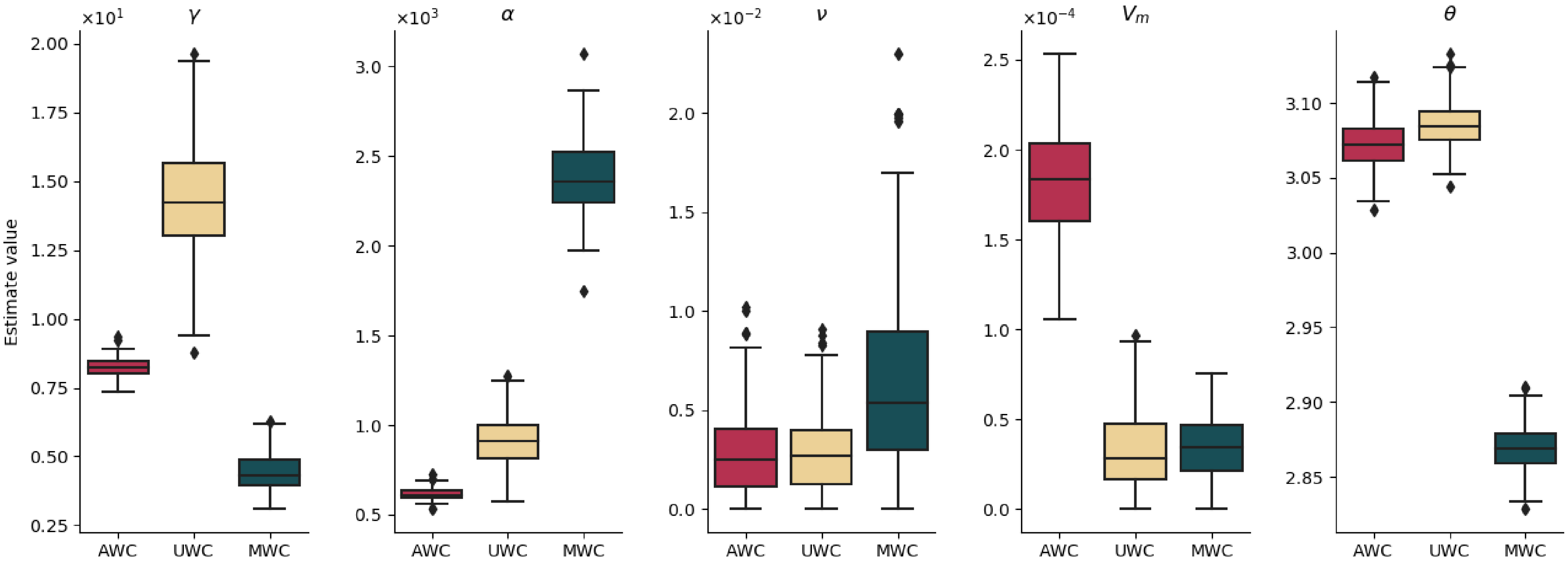

Fig. S69 – Parameter inference comparison among the three models using UMTD+PICs as the summary statistic on the log-transformed body length. The boxes in the violin plot are 1st quantile, median and 3rd quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

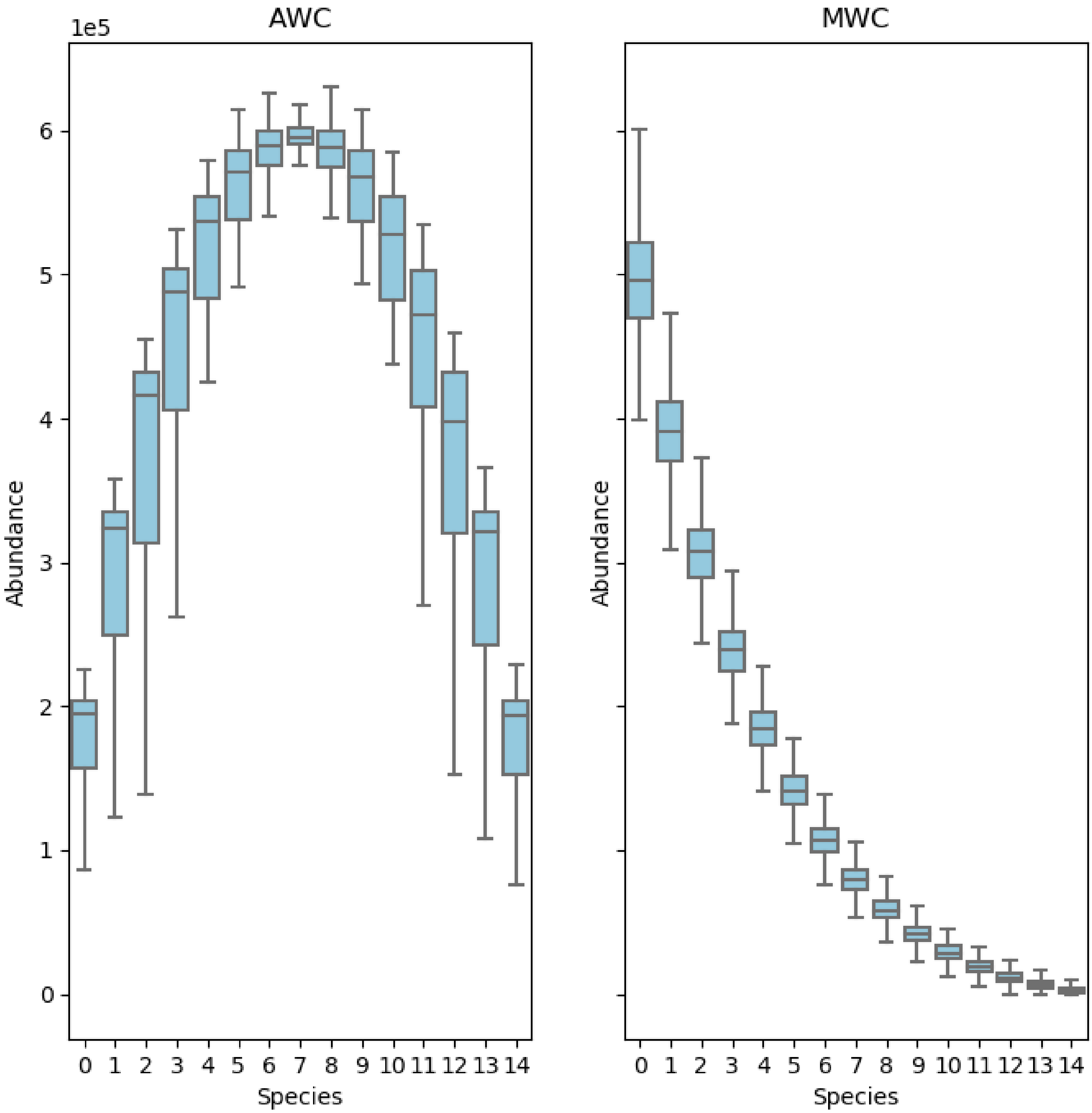

Fig. S70 – Distribution of the prediction of the abundance of species under the AWC model and the MWC model across 1000 simulations using estimated parameters from the UMTD+PICs test. x-axis denotes the species index ranked by the body size sorting from small to large. The abundance distribution under the AWC model shows a symmetric distribution while the MWC model produces a decreasing pattern with increase of the body size.

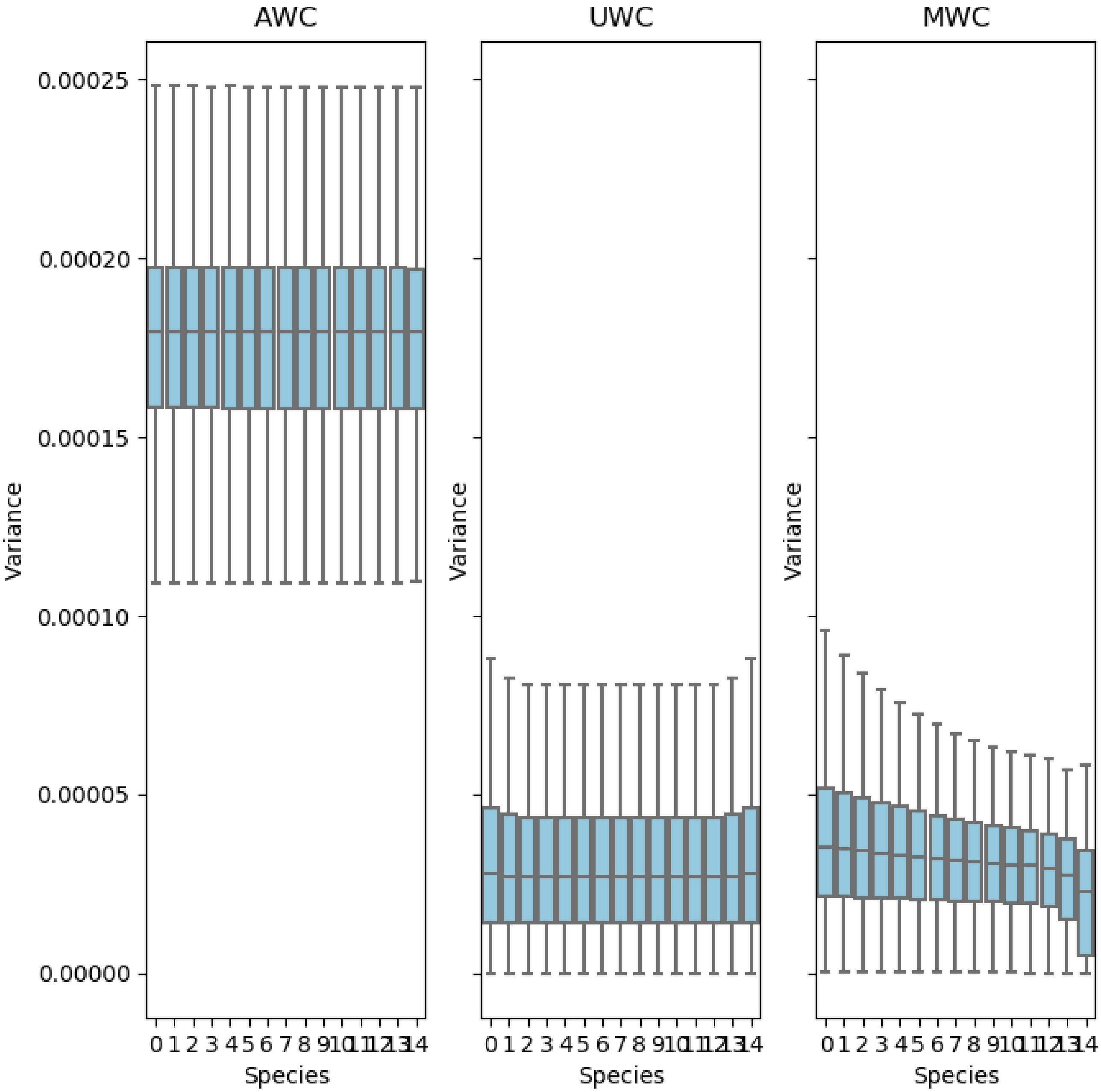

Fig. S71 – Distribution of the prediction of the trait variance of species under the AWC model, the UWC model and the MWC model across 1000 simulations using estimated parameters from the UMTD+PICs test. x-axis denotes the species index ranked by the body size sorting from small to large. The abundance distribution under the AWC model shows a symmetric distribution while the MWC model produces a decreasing pattern with increase of the body size.

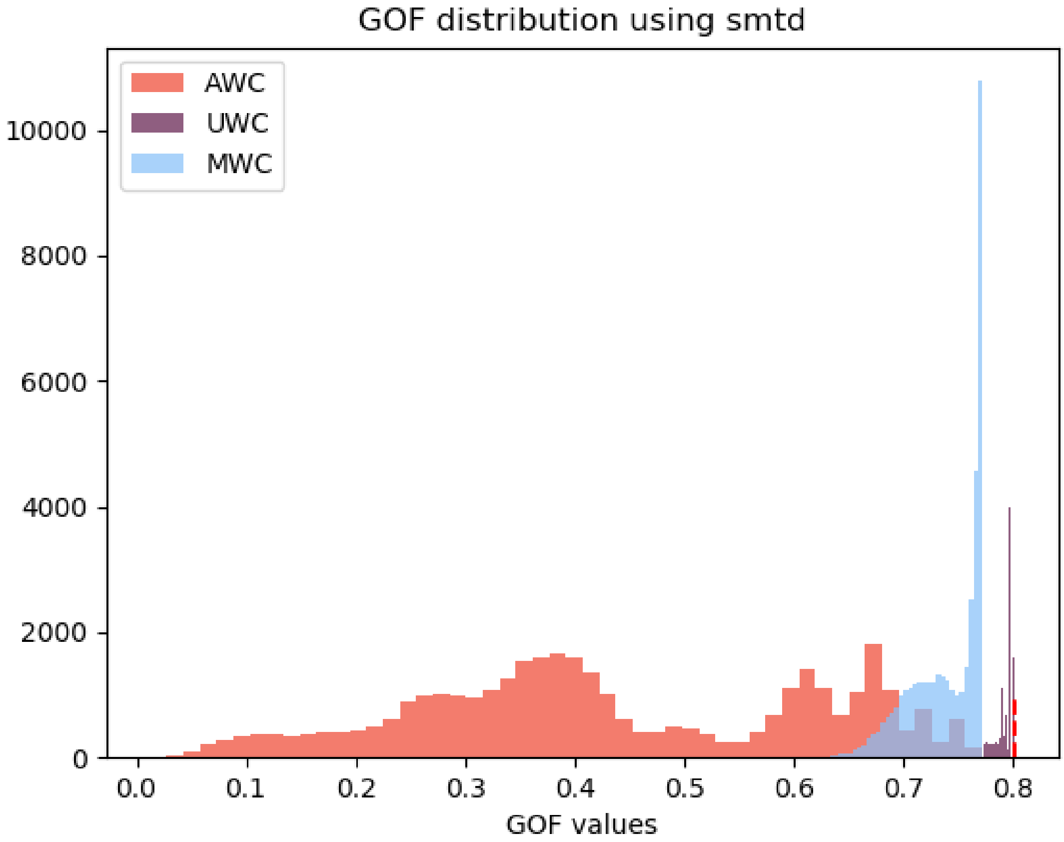

Fig. S72 – The goodness-of-fit distribution of the three models using the summary statistic SMTD. Higher values on GOF value axis denotes a better fit to data. The red dashed line denotes the 5% best GOF-values.

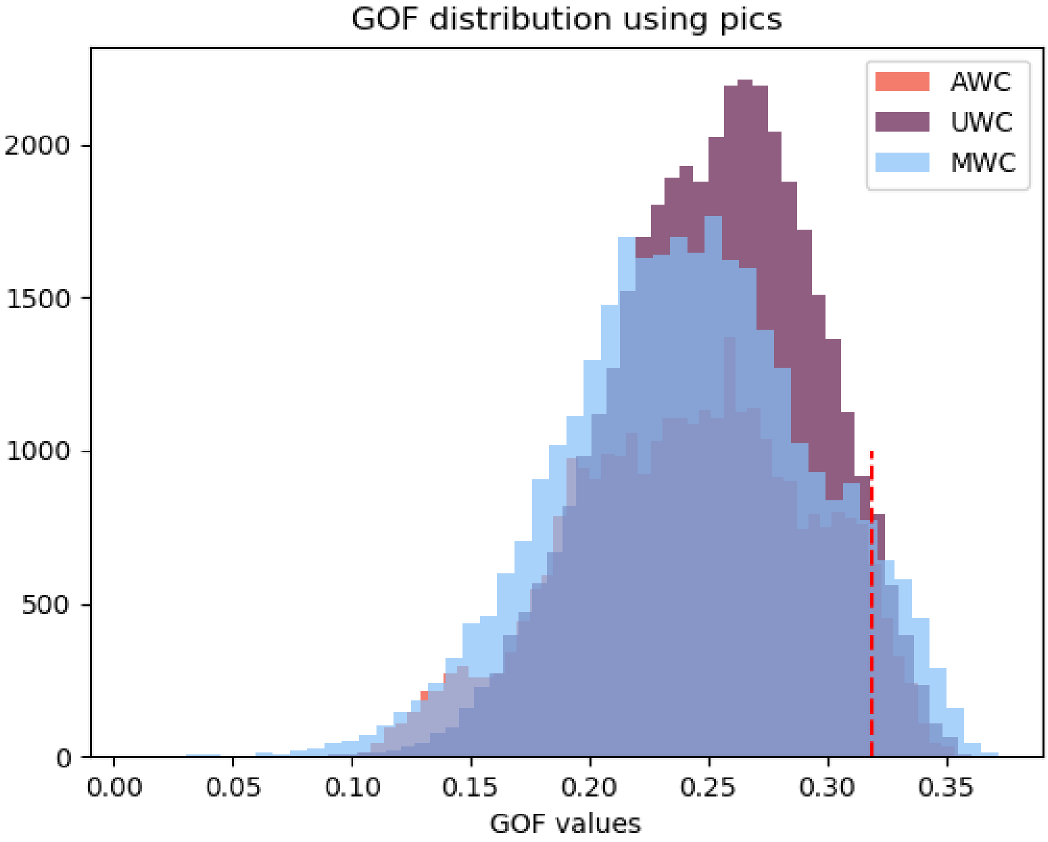

Fig. S73 – The goodness-of-fit distribution of the three models using the summary statistic PICs. Higher values on GOF value axis denotes a better fit to data. The red dashed line denotes the 5% best GOF-values.

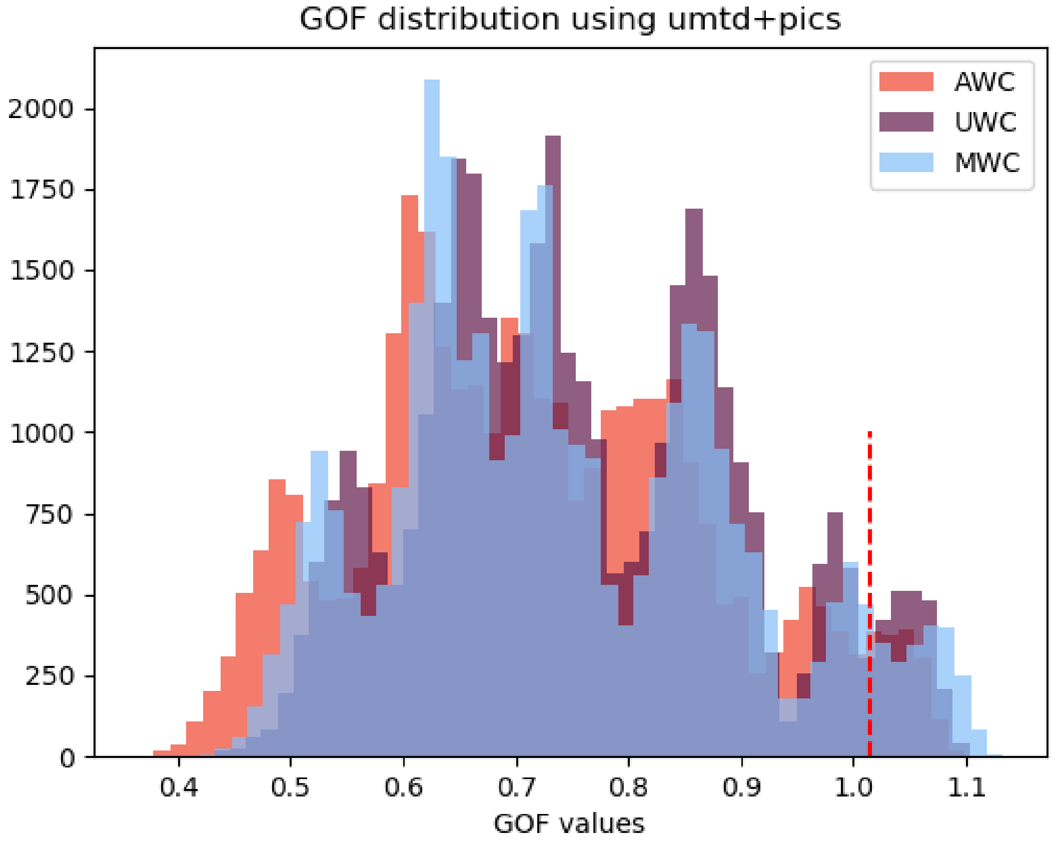

Fig. S74 – The goodness-of-fit distribution of the three models using the summary statistic UMTD+PICs. Higher values on GOF value axis denotes a better fit to data. The red dashed line denotes the 5% best GOF-values.

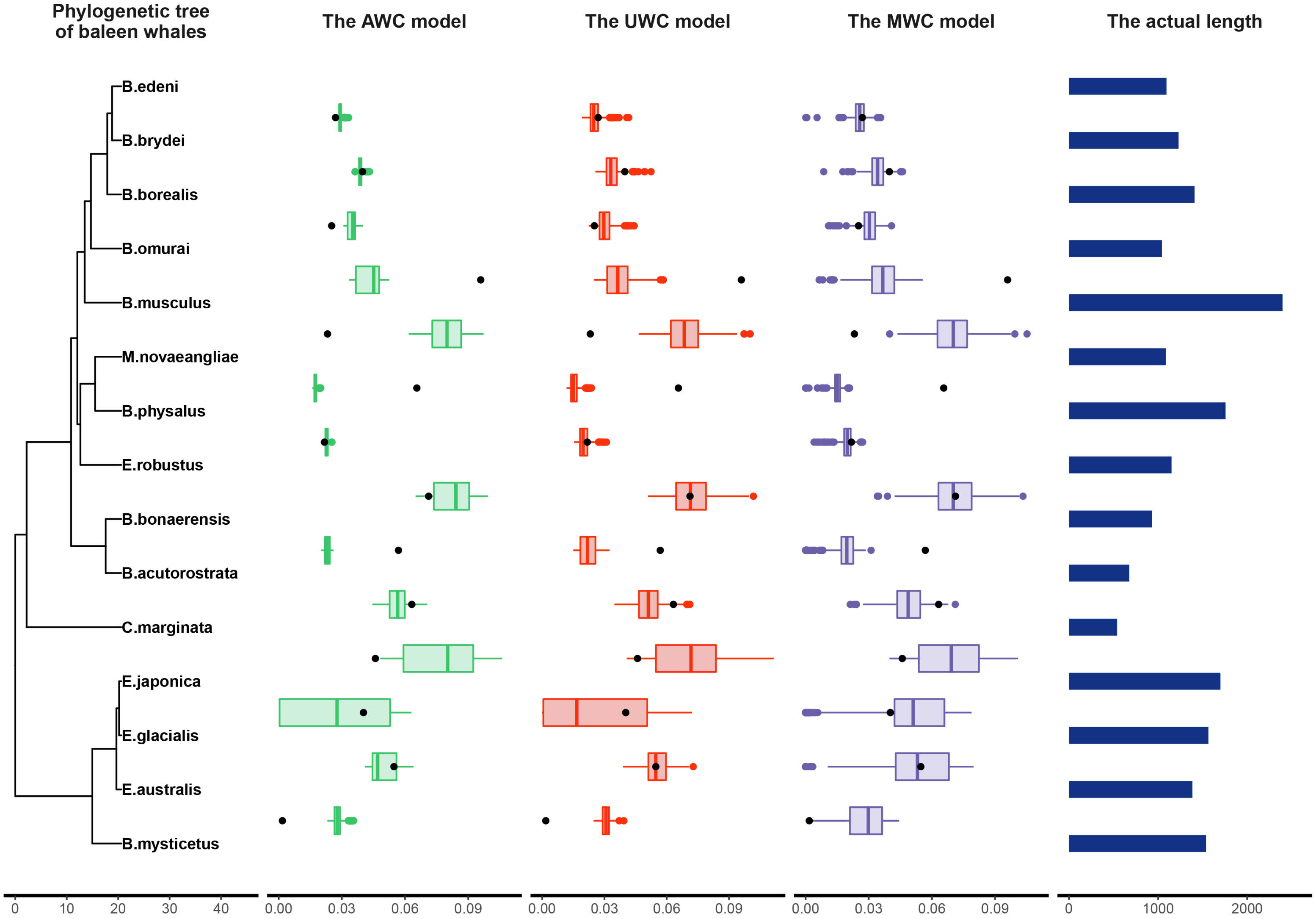

Fig. S75 – Prediction of the phylogenetic independence contrasts (PICs) of baleen whales, simulated using the parameters estimated with PICs only from the log transformed data. The time scaling parameter s is $20,000$, corresponding to 50 years per generation and heritability $h^2 = 1$. The phylogeny is the reconstructed tree of the Mysticeti with the x-axis in units of million years. The box plots show the distributions of PICs across 1000 simulations under the three models against the true data (the black dots) with the x-axis denoting the absolute phylogenetic independence contrasts. The right panel denotes the untransformed body length of the species.

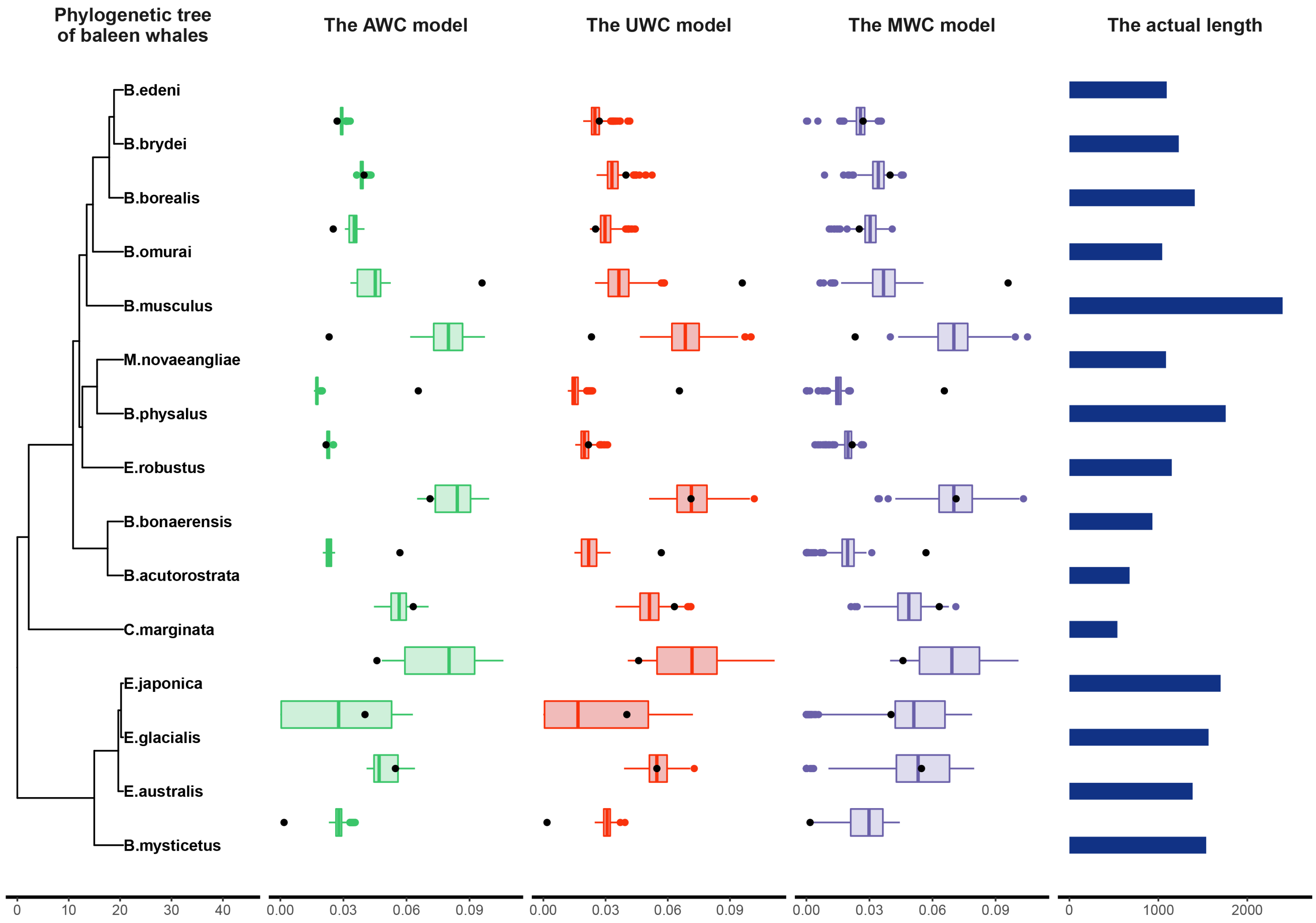

Fig. S76 – Prediction of the phylogenetic independence contrasts (PICs) of baleen whales, simulated using the parameters estimated with SMTD from the log transformed data. The time scaling parameter s is $20,000$, corresponding to 50 years per generation and heritability $h^2 = 1$. The phylogeny is the reconstructed tree of the Mysticeti with the x-axis in units of million years. The box plots show the distributions of PICs across 1000 simulations under the three models against the true data (the black dots) with the x-axis denoting the absolute phylogenetic independence contrasts. The right panel denotes the untransformed body length of the species.