

## Inferring the Effect of Species Interactions on Trait Evolution

LIANG XU\*, SANDER VAN DOORN, HANNO HILDENBRANDT, AND RAMPAL S. ETIENNE

Faculty of Science and Engineering, Groningen Institute for Evolutionary Life Sciences, University of Groningen, PO Box 11103, Groningen 9700 CC, The Netherlands

\*Correspondence to be sent to: Faculty of Science and Engineering, Groningen Institute for Evolutionary Life Sciences, University of Groningen, PO Box 11103, Groningen 9700 CC, The Netherlands;  
E-mail: [liang.xu@rug.nl](mailto:liang.xu@rug.nl)

Received 7 February 2020; reviews returned 1 September 2020; accepted 8 September 2020

Associate Editor: Bryan Carstens

**Abstract.**—Models of trait evolution form an important part of macroevolutionary biology. The Brownian motion model and Ornstein–Uhlenbeck models have become classic (null) models of character evolution, in which species evolve independently. Recently, models incorporating species interactions have been developed, particularly involving competition where abiotic factors pull species toward an optimal trait value and competitive interactions drive the trait values apart. However, these models assume a fitness function rather than derive it from population dynamics and they do not consider dynamics of the trait variance. Here, we develop a general coherent trait evolution framework where the fitness function is based on a model of population dynamics, and therefore it can, in principle, accommodate any type of species interaction. We illustrate our framework with a model of abundance-dependent competitive interactions against a macroevolutionary background encoded in a phylogenetic tree. We develop an inference tool based on Approximate Bayesian Computation and test it on simulated data (of traits at the tips). We find that inference performs well when the diversity predicted by the parameters equals the number of species in the phylogeny. We then fit the model to empirical data of baleen whale body lengths, using three different summary statistics, and compare it to a model without population dynamics and a model where competition depends on the total metabolic rate of the competitors. We show that the unweighted model performs best for the least informative summary statistic, while the model with competition weighted by the total metabolic rate fits the data slightly better than the other two models for the two more informative summary statistics. Regardless of the summary statistic used, the three models substantially differ in their predictions of the abundance distribution. Therefore, data on abundance distributions will allow us to better distinguish the models from one another, and infer the nature of species interactions. Thus, our framework provides a conceptual approach to reveal species interactions underlying trait evolution and identifies the data needed to do so in practice. [Approximate Bayesian computation; competition; phylogeny; population dynamics; simulations; species interaction; trait evolution.]

While it is generally acknowledged that “nothing in evolution or ecology makes sense except in the light of the other.” (Pelletier et al. 2009), how exactly this mutual interaction can be understood is an area of active research (Schoener 2011). We are starting to understand how ecology, that is, species interactions and population dynamics, depends on evolutionary history (Schoener 2011) and how evolutionary history, such as morphological trait evolution and phylogenetic information, explains ecological processes and patterns (Ives and Godfray 2006; Rezende et al. 2007; Rafferty and Ives 2013; Pigot and Etienne 2015; Clarke et al. 2017). However, the relative roles of biotic and abiotic factors in driving evolution remain elusive. This is in part because establishing a more explicit and mechanistic connection between the processes of trait evolution and community dynamics remains a difficult task (Narwani et al. 2015). More specifically, we do not fully understand how population dynamics influence trait evolution and vice versa.

Species tend to evolve towards a phenotype that best utilizes the most abundant resource (Darwin 1859), but also tend to diverge, in order to avoid competition with other species feeding on the same resources. The balance between these opposing tendencies is governed by two factors, niche width and population size, which ultimately determine how many species coexist in a

given environment and how they partition the available ecological niches among them. Larger populations exert a larger impact on other species competing with them in the same part of niche space. Conversely, how much competition is experienced by a species ultimately determines its population size and may lead to a shift in niche space.

Classic methods attempting to infer the relative roles of the forces shaping trait evolution from trait distribution data use models that describe trait dynamics with Brownian motion or Ornstein–Uhlenbeck processes (Raup et al. 1973; Hansen and Martins 1996; Theodore Garland et al. 2000), with the latter describing evolution towards an optimum. These models assume independent evolution for each species and thus do not account for species interactions (Pennell and Harmon 2013). Recently, several phylogenetic comparative tools have been developed to study how the abiotic environment imposing a trait optimum and competition among species jointly drive trait dynamics. Specifically, (Nuismer and Harmon, 2015) derived an analytical likelihood formula based on Lande’s work (Lande 1976) to investigate how phylogenetic relationships influence trait evolution and rates of interaction among species. This method was extended to incorporate biogeographical factors (Drury et al. 2016), and generalized to other ecological interactions (Manceau et al. 2017). However, none of these studies

took into account population dynamics thereby ignoring eco-evolutionary feedbacks of population size on trait evolution. Moreover, to obtain a tractable expression for the likelihood, they relied on a linear approximation of the fitness gradient function that has the following consequence: the more different species are, the more they repel each other. This is unrealistic because we expect that for species diverging more than their niche width, the rate of trait divergence will decrease.

Here, we present a general framework of trait evolution that can accommodate, in principle, any type of species interaction. We do this by defining the fitness function (which determines the direction in which traits will evolve) in terms of a population dynamics model. Species interactions ranging from predation and parasitism to mutualism can then be accounted for in the population dynamics model. We illustrate our framework with an application to competitive interactions and show that we resolve the above-mentioned drawbacks of pre-existing methods (Nuismer and Harmon 2015; Drury et al. 2016, 2017). Particularly, our model predicts an intermediate trait distance between species where repulsion between them is most intense, which reflects a nonlinear relationship between competition and trait similarity. Furthermore, our model allows trait evolution to depend on the intraspecific variance in traits. And last but not least, we allow for population size to affect trait evolution, and we refer to this model as the abundance-weighted competition (AWC) model. We assume that the traits in our model do not drive population divergence (speciation), but we let species evolve along a given phylogenetic tree. The model is not amenable to analytical treatment, so we use Approximate Bayesian Computation to investigate parameter inference. We find that our method is capable of recovering the generating parameters from simulated data sets under certain conditions; if these conditions are not satisfied, very different parameter sets can yield very similar patterns, so we expect that no method will be able to infer the parameters under these conditions. We then illustrate our approach using data on the evolution of total body length in baleen whales (Mysticeti) and compare our model with two variants: one that ignores the effect of population dynamics on trait evolution, that is, a model with unweighted competition (the UWC model) and one that assumes that competition is weighted by total metabolic rate (abundance multiplied by per capita metabolic rate, which is a power function of body size and hence of abundance) rather than just abundance (i.e., a linear function of abundance), which we refer to as the metabolic-rate-weighted competition (the MWC model). We find that competition played an important role in shaping trait distribution patterns in baleen whales. We discuss how our model sets the stage for a new generation of models that allow determining the relative roles of various biotic and abiotic drivers of trait evolution.

## MODEL DESCRIPTION AND ANALYSIS

### *Trait Evolution and Population Dynamics on Trees*

We consider  $n$  species competing for the same spectrum of resources with a fixed and unimodal distribution (Mahler et al. 2013).

We define fitness for an individual of species  $i$  with trait  $z_{i,t}$  at the generation  $t$  through the per capita growth function

$$\omega(z_{i,t}) = \frac{N_{i,t+1} f_{i,t+1}(z_{i,t})}{N_{i,t} f_{i,t}(z_{i,t})}, \quad (1)$$

where  $N_{i,t}$  denotes the population size at the  $t$ th generation and  $f_{i,t}$  and  $f_{i,t+1}$  are the densities of traits of the two generations, respectively. After applying simplifying assumptions such as weak selection, we can derive a model on the species level that describes the dynamics of the population size  $N$ , of the mean species trait value  $\mu$  and of the intraspecific variance of each species' trait  $V$  in terms of the fitness function (see the [Supplementary material](#) for the full derivation available on Dryad at <https://doi.org/10.5061/dryad.905qfttj4>) (Harmon et al. 2019):

$$N_{i,t+1} = N_{i,t} \cdot \omega(\mu_{i,t}) \quad (2)$$

$$\mu_{i,t+1} = \mu_{i,t} + h^2 \cdot V_{i,t} \cdot \frac{\omega'(\mu_{i,t})}{\omega(\mu_{i,t})} \quad (3)$$

$$V_{i,t+1} = (1 - \frac{1}{2}h^2)V_{i,t} + \frac{1}{2}h^2 \cdot V_{i,t}^2 \cdot \frac{\omega''(\mu_{i,t})}{\omega(\mu_{i,t})} + \frac{2N_{i,t}V_m}{1+4N_{i,t}V} \cdot h^2, \quad (4)$$

where  $N_{i,t}$ ,  $\mu_{i,t}$ , and  $V_{i,t}$  denote the population size, trait mean, and trait variance of species  $i$  at the  $t$ th generation. Our framework is amenable to all sorts of species interactions through the fitness function  $\omega(\mu_{i,t})$  which can be a function of the traits and abundances of other species as well as of the abiotic environment. In the model,  $\omega'(\mu_{i,t})$  and  $\omega''(\mu_{i,t})$  are the first and second derivatives of the fitness function with respect to the trait value, evaluated at the species mean of the trait  $\mu_{i,t}$ . The first derivative of the fitness function is proportional to the rate at which the mean trait in the population evolves uphill on the fitness landscape. The second derivative of the fitness function determines whether selection is stabilizing or disruptive, which influences the variance of the trait and, thereby, the rate of evolution. Eq. 2, which is essentially Lande's formula (Eq. 7 in Lande 1976), links the population size to a user-specified fitness function. In the literature, the trait variance is usually assumed to be constant (Lande 1976; Nuismer and Harmon 2015; Drury et al. 2016, 2017), but this may be an oversimplification (Barnett and Simpson 1955; Van Valen 1969). Therefore, we allow variance dynamics (Eq. 4) in trait evolution which consists of three components that capture a reduction of variance due to sexual reproduction ( $(1 - \frac{1}{2}h^2)V_{i,t}$ ), the effect of selection ( $\frac{1}{2}h^2 \cdot V_{i,t}^2 \cdot \frac{\omega''(\mu_{i,t})}{\omega(\mu_{i,t})}$ ), and

an inflow of new variance (Kimura and Crow 1964) due to segregation and mutation ( $\frac{2N_{i,t}vV_m}{1+4N_{i,t}v} \cdot h^2$ ). Here,  $h^2$  is the heritability of the trait, that is, the degree of phenotypic resemblance between parents and offspring (Falconer and Mackay 1996),  $v$  is the average rate of mutation of the alleles existing in a diploid population.  $V_m$  represents the maximum variance of the trait that can be supported by mutation in an infinitely large population, in the absence of selection on the trait. In the simulation study, we set  $V_m$  equal to 1 and the environmental contribution to phenotypic variance to 0, such that trait heritability  $h^2$  is equal to 1. In the empirical application, we estimate  $V_m$  as a free parameter, and we consider two distinct values of heritability  $h^2$ .

We allow for stochastic evolutionary trait change due to drift by adding a noise term  $\eta_{i,t}$  to Eq. 3 (Lenormand et al. 2009; Nuismer and Harmon 2015) that follows a normal distribution with mean 0 and a variance that is inversely proportional to the effective population size (which we set equal to the actual population size  $N_{i,t}$ ), that is,  $\eta_{i,t} \sim N(0, \frac{1}{2} \frac{V_{i,t}}{N_{i,t}})$ . We incorporate demographic stochasticity by drawing species abundances from a zero-truncated Poisson distribution with a mean determined by Eq. 2. Hence, we do not allow for extinction due to demographic stochasticity, but species can become extinct if the phylogeny tells us so (see below).

We define the fitness of the mean phenotype  $\mu_{i,t}$  at the  $t$ th generation using a Ricker-type form of discrete-time population dynamics (see the [Supplementary material](#) available on Dryad):

$$\omega(\mu_{i,t}) = \frac{N_{i,t+1}}{N_{i,t}} := R(\mu_{i,t}) e^{-\beta(\vec{\mu}, \vec{N})/\beta_0}. \quad (5)$$

Here,  $R(\mu_{i,t})$  is the growth factor that depends on the trait value and the parameter  $\theta$  that represents the optimum trait value favored by abiotic stabilizing selection as follows:

$$R(\mu_{i,t}) = R_0 e^{-\gamma(\theta - \mu_{i,t})^2}, \quad (6)$$

where  $R_0$  is the optimal growth factor and  $\gamma$  determines the strength of stabilizing selection towards the optimum. Furthermore, the function  $\beta$  in Eq. 5 quantifies the intensity of competition. Assuming a Gaussian competition kernel, we define  $\beta$  as

$$\beta(\vec{\mu}, \vec{N}) = \sum_{j=1}^n (e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2} N_{j,t}). \quad (7)$$

Equation 7 states that the strength of competition between two species with trait means  $\mu_{i,t}$  and  $\mu_{j,t}$  increases with similarity in these trait means, and that the effect of competition on species  $i$  increases with population size of species  $j$ . The competition coefficient  $\alpha$  scales the strength of the interaction and determines the effective interaction length. Finally, the

parameter  $\beta_0$  in Eq. 5 has a similar interpretation as an individual-scale carrying capacity of each species (Abrams 2001), as it sets the scale at which competitive interactions start to strongly impact the growth of the population. Because Eq. 5 is an increasing function of  $\beta_0$ , the ecological equilibrium  $\omega(\mu_{i,t}) = 1$  is reached at a carrying capacity set by the equilibrium condition  $\beta = \beta_0 \cdot \ln R$  where environmental stabilizing selection and competition balance each other.

Inserting our fitness function Eq. 5 in Eq. 2, 3, 4 and adding stochasticity leads to

$$N_{i,t+1} \sim \text{Pois} \left( N_{i,t} R_0 e^{-\gamma(\theta - \mu_{i,t})^2} \cdot e^{-\sum_{j=1}^n (e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2} N_{j,t})/\beta_0} \right) \quad (8)$$

$$\mu_{i,t+1} = \mu_{i,t} + h^2 \cdot V_{i,t} (2\gamma(\theta - \mu_{i,t}) + \Omega) + \eta_{i,t} \quad (9)$$

$$V_{i,t+1} = \left( 1 - \frac{1}{2} h^2 \right) V_{i,t} + \frac{2N_{i,t}vV_m}{1+4N_{i,t}v} \cdot h^2 + \frac{1}{2} h^2 \cdot V_{i,t}^2 \left[ 2\gamma(-1+2\gamma\theta^2) + \frac{\partial \Omega}{\partial \mu_{i,t}} - (2\gamma(2\theta - \mu_{i,t}) + \Omega) \cdot (2\gamma\mu_{i,t} - \Omega) \right], \quad (10)$$

where  $\Omega = \frac{2\alpha}{\beta_0} \sum_j (\mu_{i,t} - \mu_{j,t}) e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2} N_{j,t}$  and  $\text{Pois}(\cdot)$  denotes the zero-truncated Poisson distribution. Equations 2–4 advances Nuismer and Harmon (2015)'s model by relaxing the simplification of a linear species interaction, as in Drury et al. (2017)'s model. The nonlinear pairwise competitive repulsion,  $\alpha(\mu_{i,t} - \mu_{j,t}) e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2}$  as a function of the trait difference  $\mu_{i,t} - \mu_{j,t}$ , is illustrated in Figure 1 for several values of  $\alpha$ . Moreover, in contrast to Drury's models (Drury et al. 2016, 2017) it takes into account the abundance of competitors in the community as a weight in the competition kernel, and includes the dynamics of trait variance. Furthermore, the full derivation of our model is based on a coherent definition of fitness. We consider trait and population dynamics along given phylogenetic trees. The phylogenetic trees can be either reconstructed trees, containing only extant species, or full trees, containing extinct species as well. We assume that data may be available on the final trait distribution and species abundances at the tips of the tree. We used simulations to explore these distributions. We initialized the simulations with two ancestral species of identical trait means equal to 0 and variances equal to 0.5. The initial population sizes were drawn from a normal distribution with a mean of 500 individuals and a variance of 100. Without loss of generality, we assumed zero as the trait optimum set by the environment. To be in line with the previous models of phenotype evolution along a given phylogeny, we assumed that at the branching points of the tree the two daughter species inherit the trait value from their parent. The

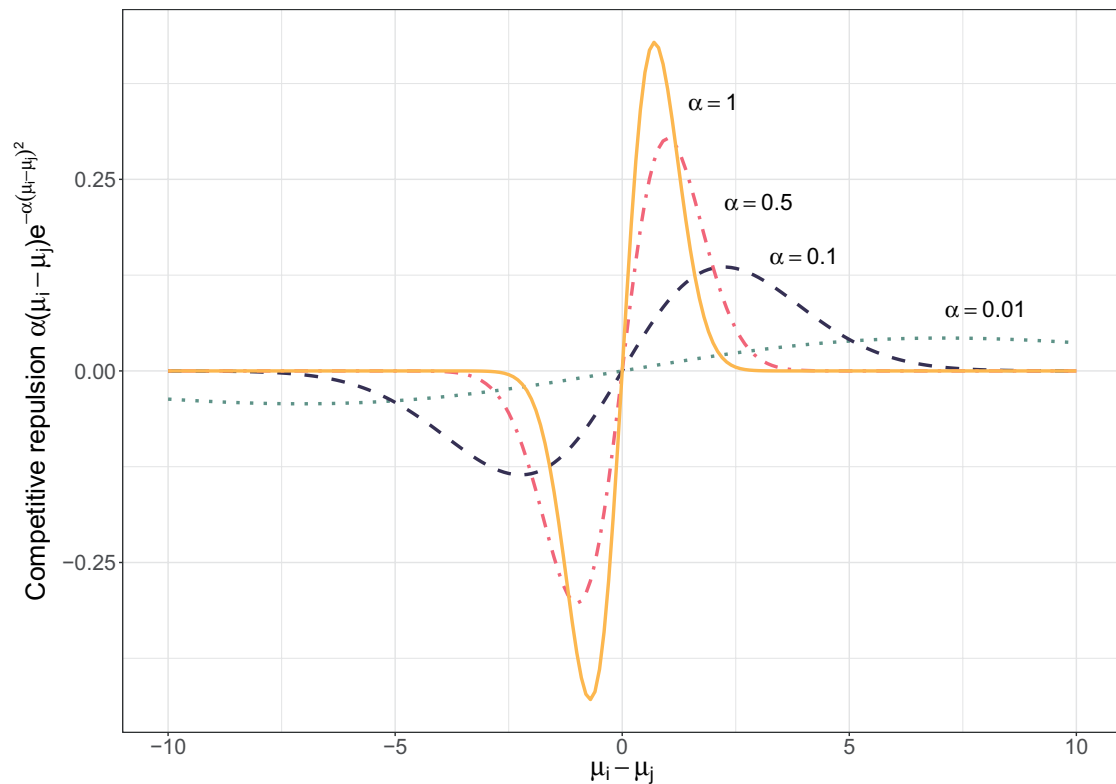


FIGURE 1. The nonlinear strength of competitive repulsion of two species with identical population size that follows from our model. When the traits of two competitors are very similar, they experience intense competition but little directional repulsion. With increasing difference in traits, the repulsion force increases. Eventually, with a further increase in trait difference, repulsion decreases again because competition is avoided. Different  $\alpha$  values cause different shapes of the strength of repulsion. A large  $\alpha$  implies strong repulsion when competitors are very similar in traits but this competitive strength drops quickly, implying a small competitive interaction distance. In contrast, a small  $\alpha$  implies a large competitive interaction distance but mild repulsion.

abundance of the mother lineage is divided over the two daughter species following a binomial distribution that is truncated at the bottom and the top, so that the daughter species always have positive abundance. When a species goes extinct according to the phylogeny, we just remove the species from the simulation but keep its trait means and variances and abundance stored for the time they were extant.

#### Parameter inference using ABC-SMC

The complexity of our model precludes analytical approaches to fit the model to data. Hence, we developed an inference framework using Approximate Bayesian Computation with Sequential Monte Carlo (ABC-SMC), which is a genetic algorithm that has computational advantages over Approximate Bayesian Computation with Markov Chain Monte Carlo because it allows parallelization, and it shows efficient convergence in high dimensional parameter space (Sunnåker et al. 2013). 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 many data sets. We then evaluate the similarity

of the simulated data to the empirical data (measured by one or more summary statistics). This similarity is the goodness-of-fit (GOF). The GOFs for all these data sets are used as weights to sample parameter sets in the next iteration (generation in ABC-SMC jargon), with some random noise added to it. After a few iterations, the parameter sets will form the posterior distribution. The details of the algorithm, including the computation of the GOF, can be found in the [Supplementary material](#) available on Dryad.

The choice of an efficient summary statistic is crucial to evaluate the similarity between simulated and empirical data. In the simulation study, we use the Euclidean distance between simulated and observed trait values. Because of the stochasticity of the trait change after speciation, traits of a focal species can differ substantially across replicate simulations. However, the difference in trait values between species that are adjacent in trait space regardless of species identities reflects the true strength of environmental stabilizing selection and competition. Therefore, we do not label the species in our simulation and sort both the empirical traits and the simulated traits in an increasing order before computing the Euclidean distance of these two vectors. We refer to this summary statistic as the sorted mean trait distance (SMTD). We also compute the Euclidean distance of



the variance vectors corresponding to the reordered trait means. In principle, we can also add summary statistics based on abundance data and intraspecific trait variances (again using Euclidean distances between simulated and observed values), but we do not do so here because such data (for entire populations) is often unavailable empirically (as in our empirical example of baleen whales). This does not mean that abundances have no effect: according to our model they affect trait evolution and hence the species' mean trait values.

Choosing the Euclidean distance of sorted traits may not be the best way to fit our model to empirical data, because we ignore information on the empirical order of the trait values across the phylogenetic tree. Therefore, in the empirical study (see below), we considered phylogenetic independent contrasts (PICs) (Felsenstein 1985) as an alternative set of summary statistics. The PICs are designed to transform the original  $n$  traits of species to  $n-1$  independently and identically distributed contrasts between pairs of related species or estimated ancestral nodes (Garland 2005). Because the PICs have one dimension less than the trait data, we combined the PICs with the unsorted mean trait distance (UMTD) to obtain a third set of summary statistics, referred to as UMTD+PICs. We compared results between the three sets of summary statistics.

### Simulation Setup

To assess the behavior of our model, we first simulated data for known parameter sets and explored whether the parameter values can be correctly inferred. We considered six different values (0, 0.001, 0.01, 0.1, 0.5, 1) for both the stabilizing selection coefficient  $\gamma$  and the competition coefficient  $\alpha$  leading to a total of 36 parameter combinations for a given phylogenetic tree. We set  $R_0 = e$  (i.e., the mathematical constant 2.7183),  $\beta_0 = 10^9$  and the mutation rate  $v = 10^{-11}$  for all simulations. We focused on the inference of three parameters, namely  $\gamma$ ,  $\alpha$ , and  $v$ .

To study how the phylogenetic information influences the evolution process, we generated several phylogenetic trees, including extinct branches, under the diversity-dependent diversification model (Etienne et al. 2011) for various parameter settings of this macroevolutionary model (see Table 1). In addition, to mimic the fact that in practice complete trees with extinct species are often not available, we reconstructed phylogenies of only extant species by pruning the extinct species. This means that we generated the trait data under the full tree but estimated the parameters of our trait evolution model using only the reconstructed tree. Comparing this inference to inference using the complete tree informs us to what extent the loss of information of extinct species affects parameter estimation.

The ratio of the time scale of trait evolution and population dynamics to the time scale set by the phylogeny (i.e., the number of time-steps of trait and population size dynamics in each unit of time of the

phylogeny, which can be interpreted as the number of generations per unit of time in the phylogeny, usually million years) is a crucial factor, as it determines whether trait and population dynamics can reach equilibrium before a new speciation (or extinction) event disrupts it. We denote this ratio in our model by the time scaling parameter  $s$ . For instance, given a phylogenetic tree with a crown age of 15 million years, trait and population dynamics involves  $15 \times s$  time steps. The value of  $s$  may influence our parameter estimates. So to assess how not exactly knowing the true number of time steps (i.e., the number of generations in a million years) affects parameter inference, we generated data under  $s=10,000$  and then ran our inference algorithm under  $s=10,000$  and  $s=20,000$  and compared their performance in parameter estimation (see Table S1 of the [supplementary material](#) available on Dryad).

In summary, we generated a total of 14 phylogenetic trees and pruned these trees when extinction rates were nonzero, resulting in 22 trees in total (see Table S1 of the [supplementary material](#) available on Dryad). We designed 30 scenarios to investigate the influence of tree size, speciation rate, extinction rate, removal of extinct species and number of time steps (see Table 1). We simulated our model for 36 parameter combinations for each scenario. We applied our inference algorithm on the simulated data and examined if the generating parameters could be recovered correctly. In the inference process, we set 30 iterations and 20 000 particles for each iteration. For the analysis of a single scenario, we exploited a cluster of 36 high performance computers with 32 threads running on each computer. Each parameter combination for each scenario analysis took between 2 and 80 h, depending on the number of evolutionary events and tree size of the specific scenario. All the code is available on Github ([https://github.com/xl0418/The\\_trait\\_population\\_coevolution\\_model\\_code](https://github.com/xl0418/The_trait_population_coevolution_model_code)).

To contrast our model with a trait evolution model where abundance does not affect trait evolution, we defined a model in which the competition kernel does not depend on species abundance; we call this model the unweighted competition (UWC) model (see Eqs S30–S31 of the [Supplementary Material](#) available on Dryad). The UWC model is similar to Drury et al.'s nonlinear extension (Drury et al. 2017) of Nuismer and Harmon's model (Nuismer and Harmon 2015). However, it differs in the competition kernel, that is, from the population dynamics model it follows that pairwise competition is described as  $(\mu_i - \mu_j) \cdot e^{-\alpha(\mu_i - \mu_j)^2}$  instead of Drury et al.'s choice of  $\text{sign}(\mu_i - \mu_j) \cdot e^{-\alpha(\mu_i - \mu_j)^2}$  (Eq. 1 in Drury et al. 2017 where  $\text{sign}(a-b)=1$  when  $a>b$  while  $\text{sign}(a-b)=-1$  when  $a<b$ ). We compared the simulated trait trajectories under the two models in the simulation study. We explored three values of the time scaling parameter, that is,  $s=100, 1000, 10,000$ , to assess whether the resulting trait patterns of the two models differ. The choice of  $s=100$  corresponding to 10,000 years per generation may be absurd. However, we used this

TABLE 1. The scenarios of the simulated phylogenies

Scenarios	Trees	$\lambda$	$\mu$	$K$	Pruned	Time scales
1	1	0.4	0	10	No	10,000
2	2	0.4	0	30	No	10,000
3	3	0.4	0	100	No	10,000
4	4	0.4	0.2	10	No	10,000
5	5	0.4	0.2	30	No	10,000
6	6	0.4	0.2	100	No	10,000
7	7	0.8	0	10	No	10,000
8	8	0.8	0	30	No	10,000
9	9	0.8	0	100	No	10,000
10	10	0.8	0.2	10	No	10,000
11	11	0.8	0.2	30	No	10,000
12	12	0.8	0.2	100	No	10,000
13	13	0.8	0.4	100	No	10,000
14	14	0.8	0.6	100	No	10,000
15	4	0.4	0.2	10	Yes	10,000
16	5	0.4	0.2	30	Yes	10,000
17	6	0.4	0.2	100	Yes	10,000
18	10	0.8	0.2	10	Yes	10,000
19	11	0.8	0.2	30	Yes	10,000
20	12	0.8	0.2	100	Yes	10,000
21	13	0.8	0.4	100	Yes	10,000
22	14	0.8	0.6	100	Yes	10,000
23	9	0.8	0	100	No	10,000, 20,000
24	12	0.8	0.2	100	No	10,000, 20,000
25	13	0.8	0.4	100	No	10,000, 20,000
26	14	0.8	0.6	100	No	10,000, 20,000
27	9	0.8	0	100	No	20,000
28	12	0.8	0.2	100	No	20,000
29	13	0.8	0.4	100	No	20,000
30	14	0.8	0.6	100	No	20,000

The experimental setup for testing the influence of phylogenetic information. The first 14 scenarios are generated under various diversification rates (speciation rate  $\lambda$  and extinction rate  $\mu$ ) and clade-specific carrying capacities  $K$ . Pruning these trees from extinct species results in the Scenarios 15–22 (only for nonzero extinction rates). Scenarios 23–26 are designed for studying the effect of the rate of evolution. The observations are generated under a time scaling parameter  $s$  of 10,000 (microevolutionary time steps per unit of macroevolutionary time) while the algorithm uses  $s = 20,000$ . For Scenarios 27–30,  $s = 20,000$  is used in both data generation and parameter inference.

value to examine how different values of  $s$  influence the behavior of the model. We emphasize that the choice of competition kernel in the UWC model (and in Drury et al.'s model) does not follow from a coherent fitness definition derived from population dynamics.

#### *Applying the model to baleen whale body size evolution*

Baleen whales represent the largest extant animal species and are distributed globally. They are filter-feeders on small fish and crustaceans. Body mass is an ideal trait that responds both to the abiotic factors (Smith et al. 2010) and biotic competitors but measurements of body mass are rarely available. However, data on total length are available. It has been shown that whale total length scales with body mass raised to a power of  $\frac{1}{3}$  (Lockyer 1976). So we used the total length as a proxy for body mass (Slater et al. 2017). We log-transformed (base 10) the body length, because the log scale is a more natural scale on which evolution takes place (Gingerich 2019). We fitted our model to mean trait data given a reconstructed phylogeny with 15 extant species (Slater et al. 2017). We did not use abundance or trait variance data, as they were not available.

We designed eight scenarios to fully assess the effects of environmental stabilizing selection and competition: four values of the time scaling parameter  $s$  (20,000, 40,000, 60,000, and 80,000) corresponding to four reasonable generation times (50, 25, 16.7, 12.5 years/generation, respectively) and two heritability values ( $h^2 = 0.5, 1$ ). In contrast to the simulation study, we also estimated the variance due to mutation and segregation,  $V_m$ , and the trait optimum,  $\theta$ . The remaining parameter settings were identical to the simulation study. In the ABC-SMC algorithm, we set 40,000 particles for each iteration and in total 40 iterations for each scenario (which are both more than in the simulation study because we are estimating more parameters).

We developed one more model for comparison with the AWC model and the UWC model. This new model, which we call the metabolism weighted competition (MWC) model, assumes that competition depends on total metabolic rate, in which abundance is multiplied by the per capita metabolic rate, which depends on body length (see Eqs S32–S33 and S35–S37 of the [Supplementary Material](#) available on Dryad). That is, the pairwise competition is  $e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2} B_{j,t}$  instead of  $e^{-\alpha(\mu_{i,t} - \mu_{j,t})^2} N_{j,t}$  (as used in the AWC model), where

$B_{j,t}$  is the total metabolic rate of species  $j$  at the  $t$ th generation. Because the logarithms of body length and body mass of whales are strongly correlated with a slope of  $\frac{1}{3}$  (Lockyer 1976) and per capita metabolic rate has a scaling with body mass of  $\frac{3}{4}$  (Brody and Procter 1932; Brody 1945; Kleiber 1947; Etienne et al. 2006), the total metabolic rate depends on body length and abundance as follows:  $B_i = N_i \cdot B_0 \cdot \mu_i^{9/4}$ . Here,  $B_0$  is a basal metabolic rate per kg (BMR/kg) that is assumed to be constant across whale species, and therefore drops out of our equations because only the relative metabolic rate matters. Thus, large-bodied species have more competitive power than small-bodied species. For the two additional models (UWC and MWC models) we again estimated five parameters ( $\gamma, \alpha, v, V_m, \theta$ ), but we considered only one scenario of heritability and time scaling ( $h^2 = 1; s = 20,000$ ), because the analyses are computationally demanding, and because we found that the scenarios were similarly supported for the AWC model (see Results section).

We used three alternative sets of summary statistics, that is, the sorted mean trait distance (SMTD), the PICs only, and the unsorted mean trait distance with the phylogenetic independent contrasts (UMTD+PICs). To compare the goodness-of-fit of the eight scenarios (for the AWC model) among each other, we took the simulations with the 5% highest GOF-values across all scenarios and computed the percentage of simulations represented by each scenario in these 5% best fitting simulations as a measure of the support of that scenario (Toni et al. 2009). We did this for each of the three sets of summary statistics. For comparing the three models we used the exact same procedure; support of a model is thus measured by its representation among the 5% best GOF-values across all three models. Lastly, because the estimates converged well, for each model we used the mean of the parameter estimates to generate 1000 data sets to compare the predicted PICs with the empirical observations.

## RESULTS

Incorporating population dynamics in trait evolution captures the gradual divergence in traits while trait evolution without population dynamics leads to fast branching in traits after speciation (see Figs. S2–S16 of the [Supplementary Material](#) available on Dryad). This is due to the fact that speciation splits the parent population into two daughter populations that are less abundant. As a consequence, the repulsion due to competition between the two daughter species is not as strong as that between two fully developed populations. This pattern is especially significant when trait and population dynamics are slow (e.g., due to a long generation time) relative to macroevolutionary dynamics, for example,  $s = 100$ . Branching in traits under the UWC model is observed substantially earlier than that under the AWC model. With sufficiently large  $s$  the two models tend to

result in similar trait distribution at the tips, indicating that equilibrium is reached.

The patterns of traits and abundances under the AWC model strongly depend on the combination of the stabilizing selection coefficient  $\gamma$  and the competition coefficient  $\alpha$ . When the competition coefficient  $\alpha$  is smaller than  $\gamma$ , the plots show highly compressed traits at the optimum trait value ( $\theta = 0$ ) (Fig. 2 and Figs. S17–S18 of the [Supplementary Material](#) available on Dryad). The traits are most diverse when the competition coefficient is moderate, due to the nonlinear nature of competition (Fig. 1). This is more pronounced when environmental stabilizing selection is weak. We also find that when environmental stabilizing selection is absent ( $\gamma = 0$ ), the species at the edge of the trait spectrum have larger population sizes than species with intermediate traits (see Figs. S19–S21 of the [Supplementary Material](#) available on Dryad). However, for all the nonzero values of the stabilizing selection coefficient  $\gamma$  we explored, the species with intermediate trait values (i.e., closer to the optimum) are more abundant than species with low or high trait values (Figs. S19–S21 of the [Supplementary Material](#) available on Dryad). Species with intermediate trait values also had larger trait variance than those with low or high trait values (Figs. S22–S24 of the [Supplementary Material](#) available on Dryad).

We found that when  $\alpha$  is larger than  $\gamma$ , but not too large, the traits are informative for parameter inference because the traits are sufficiently separated, and we will call these parameter values the informative parameter domain and otherwise the uninformative parameter domain. Looking more closely, we found that the most accurate inference occurs when the quantity  $\sqrt{\frac{\alpha}{\gamma}}$  is equal to the number of tips in the phylogeny. When this criterion is met for the generating parameter values, the ratios  $\frac{\sqrt{\hat{\alpha}/\hat{\gamma}}}{\text{Richness}}$  calculated using the estimates ( $\hat{\alpha}, \hat{\gamma}$ ) are also observed to be close to 1 and these ratios have small variance (see Figs. S25–S27 of the [Supplementary Material](#) available on Dryad). Therefore, we developed a metric  $m$  to measure this quantity given by  $m = |\text{mean}(\frac{\sqrt{\hat{\alpha}/\hat{\gamma}}}{\text{Richness}}) - 1| + \text{sd}(\frac{\sqrt{\hat{\alpha}/\hat{\gamma}}}{\text{Richness}})$ , where  $\text{sd}(\cdot)$  stands for the standard deviation. When  $m$  is close to 0, the parameter inference is reliable (see Figs. S25–S27 of the [Supplementary Material](#) available on Dryad). For example, in Scenario 1 where the tree has 10 tips (see Fig. 2), the estimates for  $\gamma$  and  $\alpha$  are expected to be most accurate for generating parameter values of  $\gamma = 0.01$  and  $\alpha = 1$ , and indeed they are. In Scenario 4 where the tree has 5 tips in Figure S28 of the [Supplementary Material](#) available on Dryad the most accurate inference is found when the generating parameter values are  $\gamma = 0.01$  and  $\alpha = 0.5$ .

*Influence of phylogenetic information.*—Tree size has a large impact on the estimation of the competition coefficient and mutation rate, but not on the estimation of the stabilizing selection coefficient (Fig. 3 and

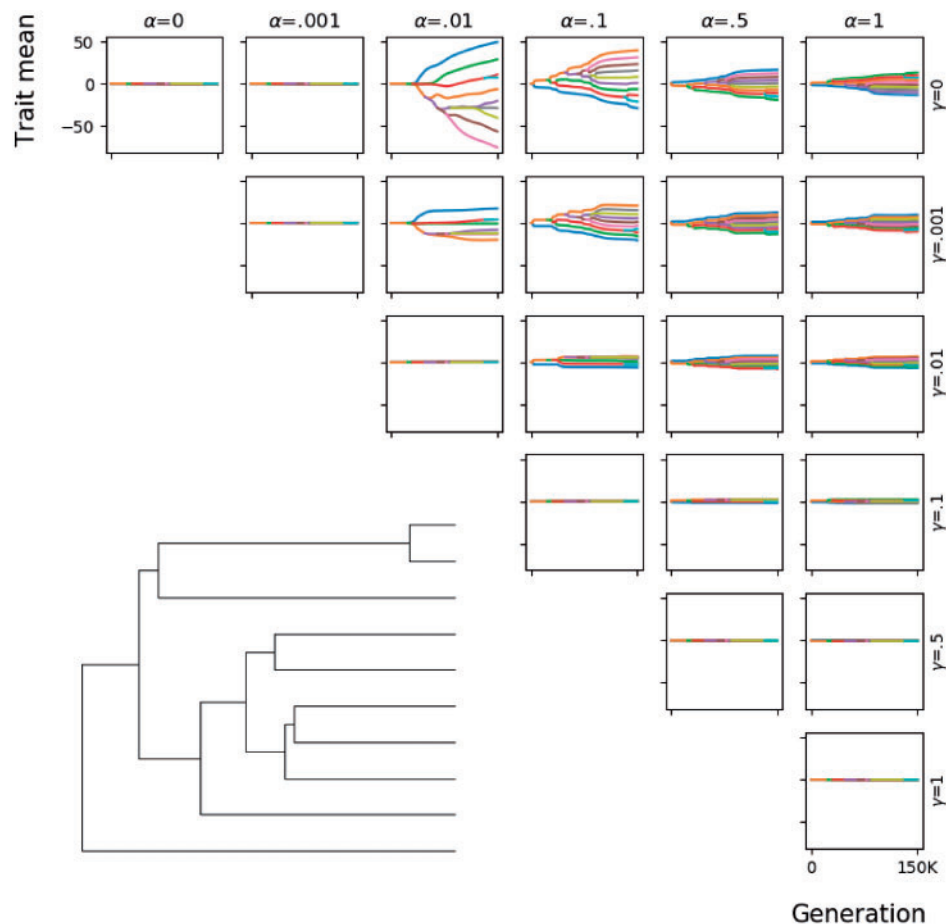


FIGURE 2. Parameter inference for the 36 parameter combinations under Scenarios 1, 2, and 3 (Table 1) (the corresponding phylogenetic trees are shown at the bottom left). The dashed lines in three colors denote three parameter values used to generate the data. Note that the scale of mutation rate is  $10^{-11}$ . 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  $\lambda = 0.4$ ,  $\mu = 0$  with a crown age of 15 myr. The clade-level carrying capacity is  $K = 10, 30, 100$  for trees 1, 2, 3, respectively. Data are shown only for the informative parameter domain ( $\gamma \leq \alpha$ ); when  $\gamma > \alpha$ , the parameter estimation shows major bias and substantial variance.

Figs. S28–S30 of the [Supplementary Material](#) available on Dryad). In general, when comparing across the scenarios with different carrying capacities  $K$  of the diversity-dependent diversification model, large trees tend to cause large variation in parameter estimates. For these large trees, the variance becomes even larger when environmental stabilizing selection becomes stronger. For very small trees (e.g., Scenarios 1, 4, 7, 10 with carrying capacity of 10), the variance of the estimates first declines and then increases with increasing  $\alpha$  while keeping  $\gamma$  fixed (see Fig. 3 in the main text and Figs. S28–S30 of the [Supplementary Material](#) available on Dryad). The same pattern is also found when increasing  $\gamma$  and fixing  $\alpha$ . The estimates of  $\nu$  show a peculiar pattern that depends on the tree size. For small trees the estimates of  $\nu$  are a bit overestimated but with a low variance. For large trees, the median of the estimates is close to the true value but the variance is large.

A high speciation rate tends to bias the estimation of  $\alpha$  and  $\nu$  and increases the variance of the estimates for the scenarios with nonzero extinction while for the pure birth scenarios it improves the estimation (see Figs. S31–S33 of the [Supplementary Material](#) available on Dryad for the scenarios with no extinction and Figs. S34–S36 of the [Supplementary Material](#) available on Dryad for the scenarios with extinction). For the estimation of  $\gamma$ , there is no substantial effect of the speciation and extinction rates. However, the estimation is likely affected by the range of the resulting traits, e.g. a small range of the resulting traits leads to inferring large  $\gamma$ . The estimation of the competition coefficient and the mutation rate is likely to be affected by large speciation and extinction rate because the frequent speciation and extinction prevent the trait evolution from reaching equilibrium.

Generally, extinction may cause more variance in the estimates of  $\alpha$  but the median of the estimates resembles



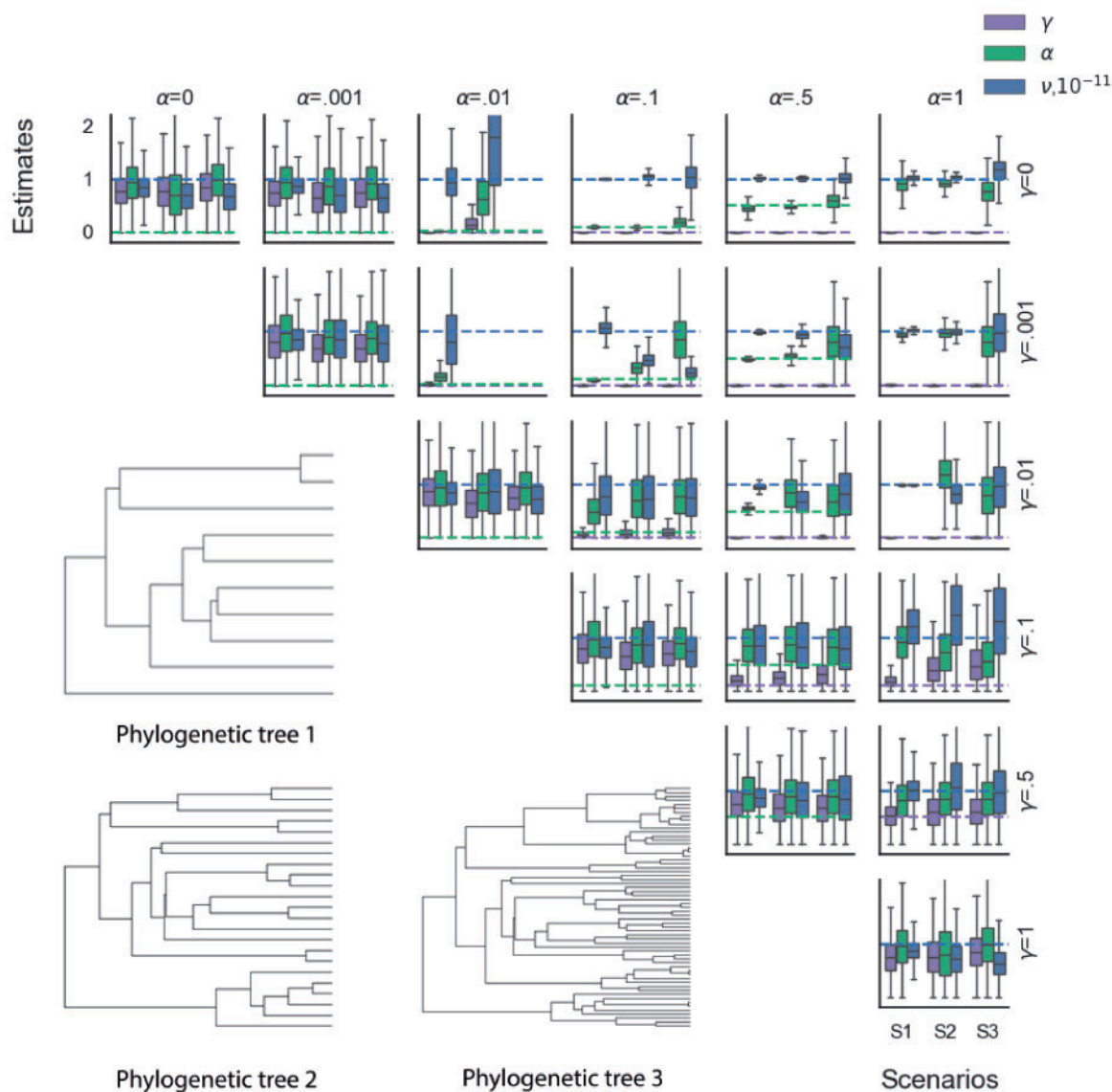


FIGURE 3. Parameter inference for the 36 parameter combinations under Scenarios 1, 2, and 3 (Table 1) (the corresponding phylogenetic trees are shown at the bottom left). The dashed lines in three colors denote three parameter values used to generate the data. Note that the scale of mutation rate is  $10^{-11}$ . 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  $\lambda = 0.4$ ,  $\mu = 0$  with a crown age of 15 myr. The clade-level carrying capacity is  $K = 10, 30, 100$  for trees 1, 2, 3, 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.

the true values for the trees that are not too small (see Figs. S37–S42 of the [Supplementary Material](#) available on Dryad). For small trees, for example Scenarios 4 and 10 (see Fig. S34 of the [Supplementary Material](#) available on Dryad),  $\alpha$  is greatly underestimated when there is no stabilizing selection and the generating competition coefficient is large ( $\alpha = 1$ ), while  $\gamma$  is well estimated except when stabilizing selection is weak and the competition coefficient is large for the scenario with very large extinction rate ( $\mu = 0.6$ ) (see Scenario 14 in Fig. S42 of the [Supplementary Material](#) available on Dryad). Although the estimates of  $\nu$  show an increasing variance when the

extinction rate increases, the true value is still recovered on average. Interestingly, in the cases where  $\gamma$  and  $\alpha$  are both large ( $\gamma = 0.5$ ,  $\alpha = 1$ ) a nonzero extinction rate greatly improves the estimation of all parameters when using the full tree (Scenarios 4–6 and 10–14).

*Influence of reconstructing phylogenetic trees.*—Surprisingly, from the overall comparison among the full-tree scenarios and the corresponding pruned-tree scenarios, there is no substantial difference in parameter estimates for most of the parameter combinations in the informative parameter domain (see Figs. S43–S50

of the [Supplementary Material](#) available on Dryad). Only a slight underestimation of  $\alpha$  is observed for the pruned-tree scenarios compared to the full-tree scenarios. The estimation of  $\gamma$  is only substantially biased for the scenarios using extremely large extinction rate (Scenarios 14 and 22 with extinction rate  $\mu=0.6$ , see [Fig. S50](#) of the [Supplementary Material](#) available on Dryad) and small stabilizing selection coefficient. There seems to be no clear influence of pruning extinct species on the estimate of  $v$ . The overall relative insensitivity of the parameter estimates to phylogenetic reconstruction for only extant species may be due to the fact that species extinction will almost immediately lead to another species occupying its niche.

*Influence of the time scaling parameter.*—Choosing a larger value of the time scaling parameter ( $s=20000$ ) in estimation than the one used to generate the data ( $s=10000$ ) does not lead to substantial difference in estimation for most of the parameter combinations (see [Figs. S51–S54](#) of the [Supplementary Material](#) available on Dryad). The competition coefficient is only slightly underestimated for the scenarios with large  $\alpha$  except for one special case when  $\gamma=0$  and  $\alpha=1$  in the pure birth tree (Scenario 23, see [Fig. S51](#) of the [Supplementary Material](#) available on Dryad). The estimation of  $\gamma$  is accurate in the informative parameter domain. The estimates of  $v$  show a somewhat larger variance but still match the true value in most of the parameter combinations.

A larger value of the time scaling parameter used both in generating the data and in the parameter inference generally does not improve the performance of parameter inference ([Figs. S55–S58](#) of the [Supplementary Material](#) available on Dryad). A significant improvement only appears in the estimation of  $\alpha$  when  $\gamma=0$  for the scenarios with small or zero extinction rates. The stabilizing selection coefficient  $\gamma$  is normally equally well estimated for the scenarios with small extinction rates for both time scaling parameters. By contrast, for the scenarios that have large extinction rates, the parameter inference is worst when stabilizing selection is weak. When stabilizing selection becomes strong, parameter estimation for all three variables is substantially improved.

#### *Environmental stabilizing selection and competition in baleen whales*

In all scenarios, the estimates of  $\gamma$  are much smaller than the estimates of  $\alpha$ . The square root of the ratio of  $\alpha$  and  $\gamma$  is close to the clade size of the baleen whale phylogeny with the  $m$  values being around 1 (in the range of (1, 1.4) for SMTD, (0.82, 1.7) for PICs and (1.02, 1.6) for UMTD+PICs). By comparing with scenario 2 which has similar tree size (see [Fig. S26](#) of the [Supplementary Material](#) available on Dryad), this suggests that parameter inference is reliable. Heritability tends to affect the inference of the environmental stabilizing selection coefficient  $\gamma$ , and the competition

coefficient  $\alpha$  but not so much the mutation rate  $v$ , the maximum variance by mutation  $V_m$  and the optimum trait value  $\theta$  (see [Fig. 4](#)). With a larger heritability ( $h^2=1$ ), the estimates of  $\gamma$  and  $\alpha$  are smaller and less variable than for  $h^2=0.5$  across all three summary statistics. Alternative sets of summary statistics lead to similar results in the estimation of the parameters  $v$ ,  $V_m$  and  $\theta$ , but for  $\gamma$  and  $\alpha$ , the inferences using PICs and UMTD+PICs differ from those using SMTD particularly when  $h^2=0.5$  (see [Fig. 4](#) for SMTD, [Fig. S59](#) of the [Supplementary Material](#) available on Dryad for PICs and [Fig. S60](#) of the [Supplementary Material](#) available on Dryad for UMTD+PICs). The estimated  $\gamma$  is smallest, around 6, when using PICs. The value increases when the summary statistics include absolute trait information (SMTD and UMTD+PICs), reaching 9 for the algorithm using SMTD only. In contrast, the estimations for  $\alpha$  using SMTD are much smaller than using the other two summary statistics. The mutation rate  $v$  is consistently inferred to be 0.001 except when using PICs, in which increasing  $s$  to 80,000 increases the estimate of  $v$  to 0.002. We did not find substantial differences in the estimates under different time scaling parameters, except that with only PICs as the set of summary statistics, the predicted  $\alpha$  decreases when increasing the time scaling parameter and/or heritability. The estimation of  $\theta$  is estimated to be around 3.05 across all three summary statistics, although a larger time scaling parameter leads to a larger variance when using only PICs. Results based on the untransformed total length can be found in [Figures S61–S63](#) of the [Supplementary Material](#) available on Dryad.

Comparing the support of the scenarios across  $h^2$  and  $s$  (computed from the GOF-values) tells us that when using SMTD as summary statistic  $h^2=0.5$  better fits the data than  $h^2=1$  while all values of  $s$  are similarly supported. Using PICs and UMTD+PICs as summary statistics leads to equally good fits for different  $h^2$  and  $s$  (see [Fig. 5](#)).

We note that the quantitative values of the estimates for stabilizing selection and competition coefficients are not comparable among the three models of trait evolution (AWC, UWC and MWC) because these models assume different factors affecting competition. Nevertheless, the estimations all suggest a small environmental stabilizing selection coefficient and a large competition coefficient (see [Figs. S67–S69](#) of the [Supplementary Material](#) available on Dryad for the three sets of summary statistics). When using the summary statistics SMTD, the best fitted value of the optimum trait  $\theta$  is close to the mean of the extant species traits (3.087) for the AWC model and the UWC model, and around 2.9 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 in estimations of  $\theta$  in all three models. The abundance distribution substantially differs for the three models ([Fig. S70](#) of the

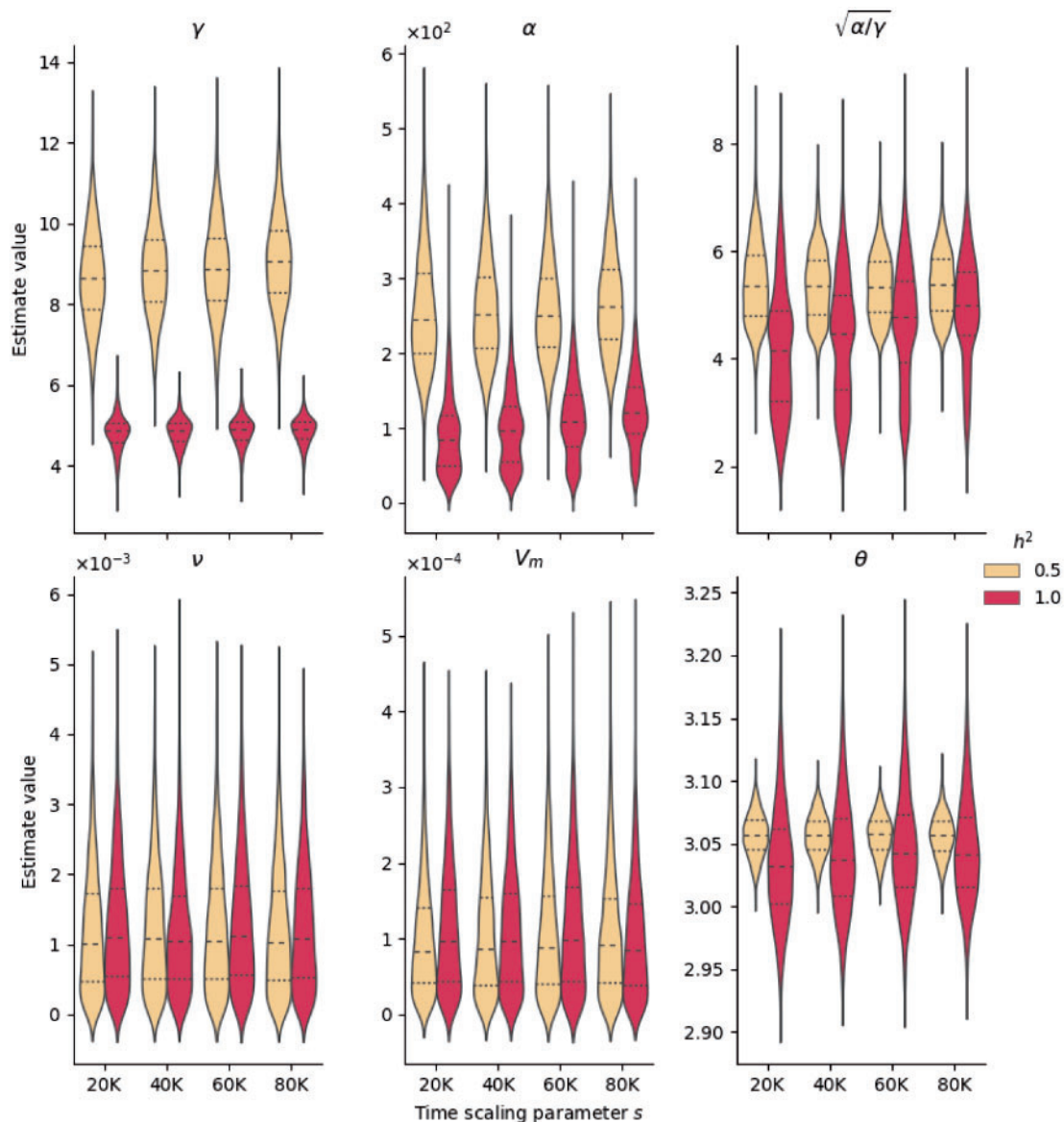


FIGURE 4. Parameter inference under eight parameter settings (i.e. all combinations of  $s = 20,000, 40,000, 60,000, 80,000$  and  $h^2 = 0.5, 1$ ) to test the influence of the number of time steps and heritability on parameter estimation using SMTD as the summary statistic (for the results using PICs and UMTD+PICs, see Figs. S59 and S60 of the [Supplementary Material](#) available on Dryad). The three dashed lines in the violin plot are 25th percentile, median and 75th percentile quantile of the samples in the last iteration of the ABC algorithm that produce the 5% best fits to the baleen whale data.

[Supplementary Material](#) available on Dryad) regardless of the choice of summary statistics. The AWC model generates a symmetric unimodal abundance distribution around the optimum trait value. For the MWC model, the abundance distribution shows a decrease of abundance with increasing body length. The prediction of the intraspecific variance also differs substantially among the three models (see Fig. S71 of the [Supplementary Material](#) available on Dryad). The AWC model produces higher intraspecific variance than the UWC and MWC models. Comparing the supports of the three models we find that the UWC model is favored when using SMTD as summary statistic while all three models show a similar

fit (with the MWC model slightly favored) when using the other two (more informative) summary statistics (Fig. 5 and see Figs. S72–S74 of the [Supplementary Material](#) available on Dryad for the GOF distributions of the three models for the three summary statistics).

Using the estimated parameters, we generated 1000 replicate trait data sets for the three models from the posterior distributions and compared the predicted phylogenetic independent contrasts to the empirical data (see Fig. 6 for the results using UMTD+PICs and Figs. S75–S76 of the [Supplementary material](#) available on Dryad for the results using PICs and SMTD). We observe a good fit for most of the contrasts but substantial

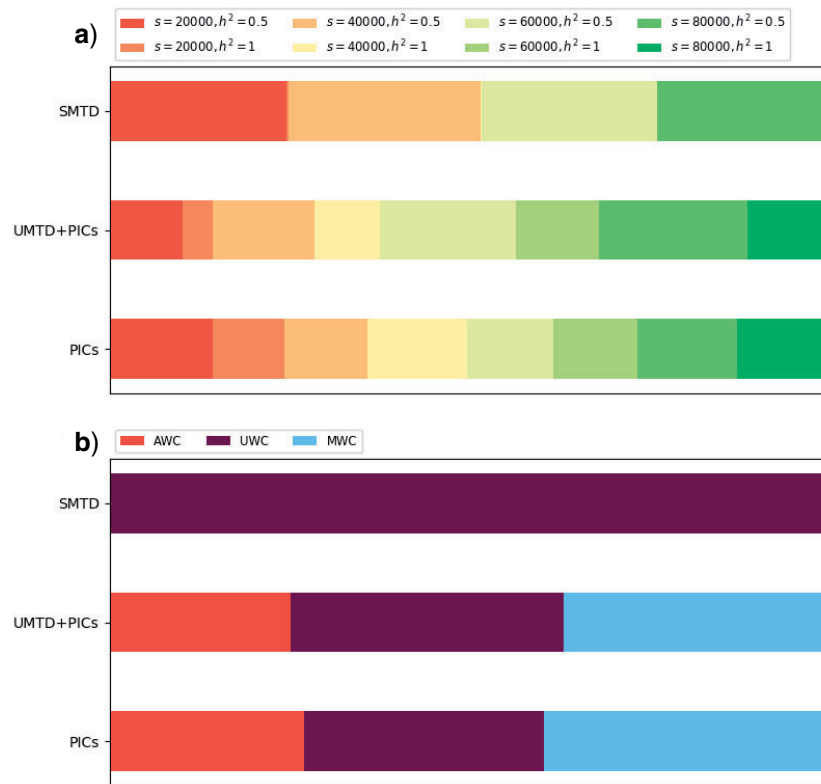


FIGURE 5. Model comparison. a) Support of various  $h^2$  and  $s$  for the three summary statistics under the AWC model. The width of the bar denotes the support of the corresponding parameter combination. b) Support across the three candidate models (AWC, UWC, and MWC) for the three summary statistics. The height of the bar denotes the support of the corresponding model.

differences for some pairs of species and clades. For example, the contrasts between the pairs of *B.musculus* and its daughter clade, *M.novaeangliae* and *B.physalus* are underestimated.

## DISCUSSION

We have developed a novel, coherent, framework to study how species interactions affect trait evolution. We have illustrated it with a model of trait evolution and population dynamics to explore how environmental stabilizing selection and competition shape trait patterns. We employed Approximate Bayesian Computation with a Sequential Monte Carlo algorithm to infer the parameters of interest and measure the performance of our method for simulation data. Our analysis reveals that the trait data at the present is generally informative for parameter inference, particularly when the number of species allowed by the strength of stabilizing selection and competition is similar to the number of species present in the phylogeny. Otherwise, our inference approach is limited due to an uninformative trait pattern that is highly compressed to the optimum trait. Our empirical study shows a small environmental stabilizing selection coefficient and a relatively large competition coefficient, suggesting a rapid repulsion among whale species.

Our model advances previous studies (Nuismer and Harmon 2015; Drury et al. 2016, 2017; Manceau et al. 2017) in three ways. First, our model is based on an explicit definition of fitness based on population dynamics, instead of chosen ad hoc. This property makes our model more coherent. Second, we have relaxed the assumption of a linear competition kernel, and thereby capture a realistic mechanism in species interactions: when two populations have very similar trait values, the substantial overlap in their intraspecific trait distributions neutralizes directional selection although competition is very intense there. Consequently, the traits are diverging very slowly. However, once stochastic mutation produces imbalance, competition leads to character displacement. Thus, the force of directional selection dramatically increases at the onset of a division in traits. When species interactions decrease due to increasing dissimilarity, the force of directional selection drops. This process continues until the two populations evolve apart in traits to the extent that competition is balanced by stabilizing attraction. The third advantage of our model is that we have made population size a factor in the force of competition (which seems more realistic), and modeled the population dynamics.

Our simulation study reveals that phylogenetic information, that is, tree size, branching times and extinct branches, is informative for the inference of stabilizing selection but carries little information



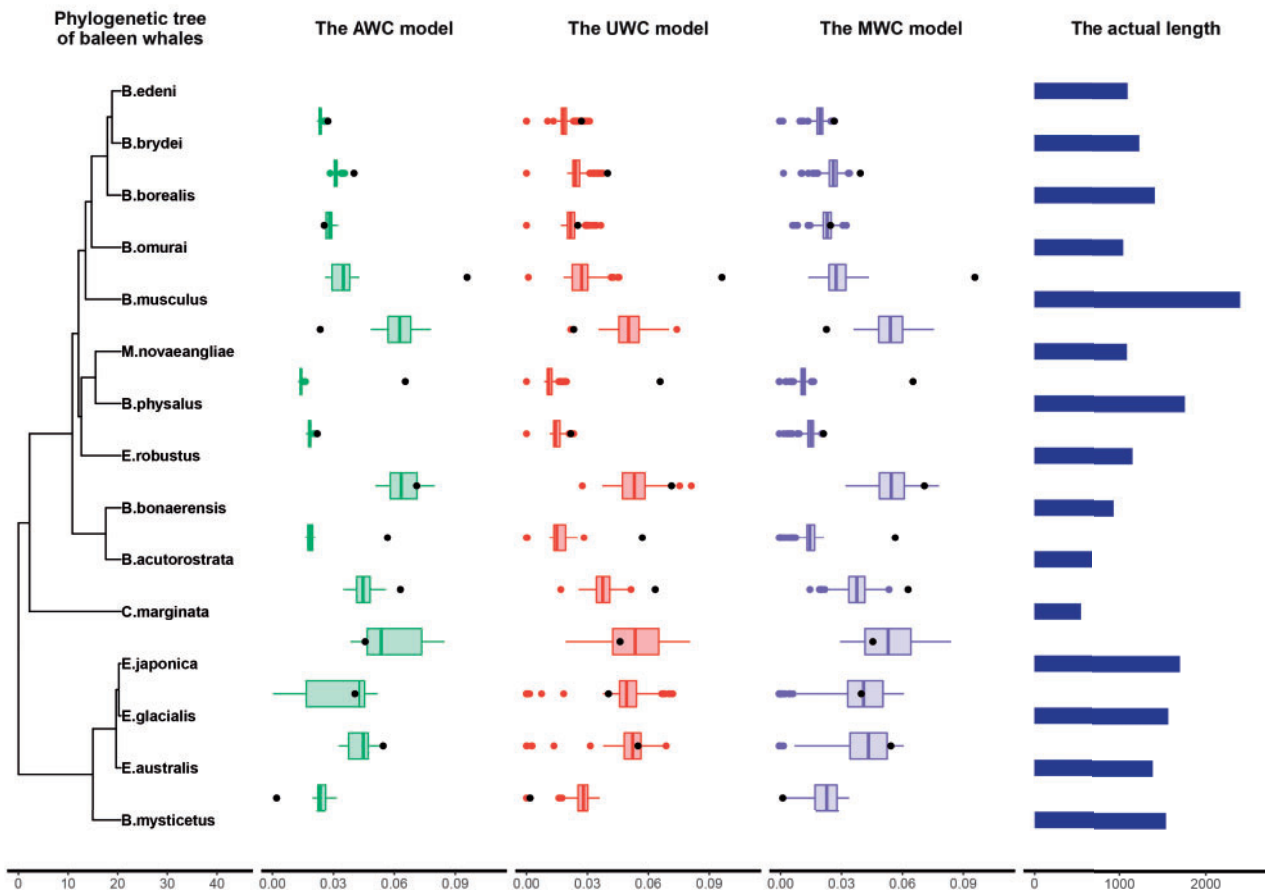


FIGURE 6. Prediction of the phylogenetic independent contrasts (PICs) of baleen whale log-transformed body length, simulated using the parameters estimated using UMTD+PICs. The time scaling parameter  $s$  is 20,000, corresponding to 50 years per generation and heritability  $h^2 = 1$ . The phylogeny is the reconstructed tree of the Mysticeti with the x-axis in units of million years (Slater et al. 2017). 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 bars on the right show the untransformed body length of the species.

on species interactions in phylogenies with nonzero extinction rate (Nuismer and Harmon 2015). Large trees show more variance in estimations. On the one hand, this is because large trees have more speciation events that may result in species branching before they can evolve to equilibrium. On the other hand, the potential mismatch between the species carrying capacity  $K$  used to generate phylogenetic trees and the number of coexisting species that is allowed by the strength of environmental stabilizing selection and competition could play a role in the performance of the inference method. Large environmental stabilizing selection results in a narrow width of natural resources, which may sustain only a few species. However, our simulation conditioning on a prescribed individual carrying capacity and a given phylogenetic tree forces more species to survive than stabilizing selection would naturally allow. This conflict may produce more variance in parameter estimation. High speciation rate generally improves parameter inference in phylogenies resulting from pure birth process, because the community can reach the clade-level carrying capacity faster, allowing more time to

form a specific pattern in traits. Conversely, extinction prevents the community from reaching the carrying capacity and generates more macroevolutionary events, allowing less time for trait evolution. Thus, whether parameter inference is accurate depends on the time scaling parameter ( $s$ ). Generally our study suggests that if  $s$  is high, extinction rate has little influence on the behavior of the parameter inference. If, conversely, extinction rate is so high that the trait and population dynamics are not fast enough to fill the trait niches left by extinct species, considerable bias in parameter inference is expected.

Reconstructing trees by pruning extinct species has a similar effect on parameter inference as using full trees if environmental stabilizing selection is present and the number of time steps is large. One might expect that extinction leaves gaps in the trait distribution suggesting a wider effective competition zone corresponding to a smaller competition coefficient  $\alpha$ , which is observed in our results. However, any possible bias that extinction may cause will disappear if given enough time steps because extant species eventually take over empty niches

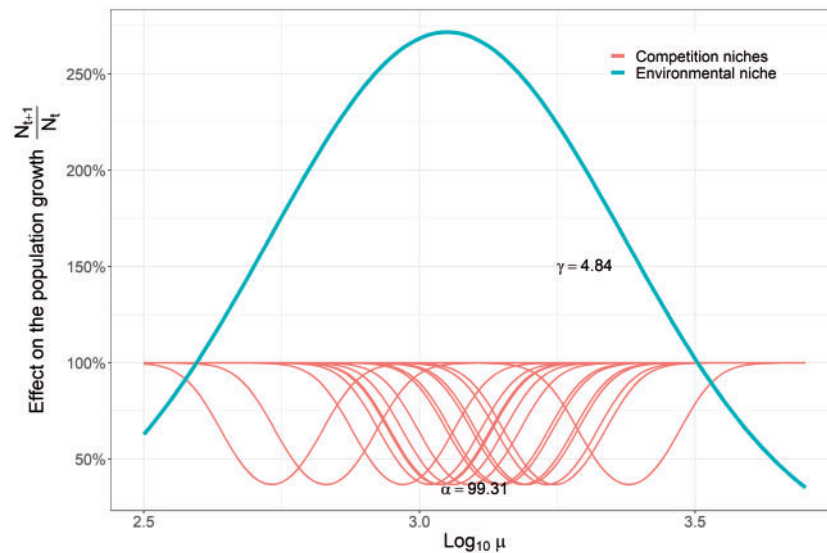


FIGURE 7. Illustration of the effect of the environmental stabilizing selection coefficient  $\gamma$  and the competition coefficient  $\alpha$  on the population growth rate in baleen whales. The estimates are obtained from the scenario under the AWC model with  $s = 20,000$ ,  $h^2 = 1$ . The two coefficients  $\gamma$  and  $\alpha$  are estimated as 4.84 and 99.3, respectively, with the optimum  $\theta$  being estimated as 3.05. The green curve corresponds to the term  $e^{-\gamma(\theta-\mu)^2}$  that refers to the breadth of the environmental resource with the optimum trait  $\theta$ . The red curve corresponds to the term  $e^{-\alpha(\mu_i-\mu)^2}$  (see Eqs. 5, 6, and 7).

left by species that went extinct. Thus, if trait evolution is fast enough, the trait pattern of extant species only can still inform about species interactions. That is to say, under these conditions phylogenetic information is not necessarily needed for the inference of species interactions. This conclusion is consistent with Nuismer and Harmon's finding for the phenotype differences model (Nuismer and Harmon 2015). This result increases the robustness of our inference approach when applied to empirical data, where we often have only molecular phylogenies and morphological trait data for extant species.

Our ABC approach can be used in principle to estimate all parameters, including  $s$  and  $h^2$ . However, we did not do so for the simulation study because we wanted to focus on the ecological parameters, and we did not do so for the baleen whales because of the small size of the data set. Moreover, these parameters can often be determined independently: for the baleen whales we chose four different values of  $s$  that correspond to plausible generation times, and two values of  $h^2$  that span the range of plausible heritability values. We found that none of the results depended much on the value of  $s$ , and  $\sqrt{\frac{\alpha}{\gamma}}$  was hardly affected by the value of  $h^2$ . For the estimated parameters, we developed a simple metric  $m$  to determine whether inference is reliable. This metric can be easily computed from the posterior distribution of the parameter values; the larger it is, the less reliable the results are. We consider values of 1 as we found in the baleen whales to be quite reliable.

In our example of baleen whales we find that  $\alpha > \gamma$  suggesting strong competition, but we must

interpret this result with caution. A large value of the competition coefficient  $\alpha$  does not immediately imply strong competition because of the shape of  $\Omega$  (see Fig. 1) where larger  $\alpha$  not only increases the intensity of the repulsion between two species with different traits, but at the same time reduces the interaction distance, that is, species with very similar trait values experience stronger interaction, but species with dissimilar trait values experience less interaction. How should we then interpret the parameter values? As an example, for the fit with the AWC model with  $s = 20,000$ ,  $h^2 = 1$  and using the STMD summary statistic,  $\theta$  is estimated as 3.05 (Fig. 7). This means that the optimum body length for baleen whales is  $10^{3.05} = 1122$  cm. The estimated value for  $\gamma$  is 4.84. This means that the growth rate of species with a body length differing by  $\sqrt{\frac{1}{\gamma}}$  on a base-10 logarithmic scale is a factor  $\frac{1}{e} = 0.36$  of the optimum growth rate obtained at  $\theta = 3.05$ . This corresponds to body lengths of 398 cm and 3162 cm. The estimated value for  $\alpha$  is 99.3 which means that species body lengths being apart by  $\sqrt{\frac{1}{2\alpha}}$  on a base-10 logarithmic scale from a competitor's mean body length experience the highest repulsion from that competitor. For example, *B. bonaerensis* with trait mean  $10^{2.97} = 933$  cm will be repelled most strongly by species with trait means 792 cm and 1099 cm.

We developed a metric  $m$  that can be applied to the ABC output to determine whether our parameter estimation is reliable. We showed that when the generating parameter values yield values of  $\frac{\sqrt{\hat{\alpha}/\hat{\gamma}}}{\text{Richness}}$  close to 1, parameter estimations yielding an  $m$ -value

close to 0 signal reliability of the inference, whereas parameter values with  $m$ -values far from 0 should not be trusted. It is possible that when the generating values do not satisfy that  $\frac{\sqrt{\hat{\alpha}/\hat{\gamma}}}{\text{Richness}}$  is close to 1, estimates yield  $m$ -values close to 0 as well, and hence one might falsely conclude that the estimation is reliable. However, we believe that such a scenario would not occur in practice if the basic assumptions of the model are satisfied: the richness observed in the phylogeny will have to be similar to the richness dictated by the parameters  $\alpha$  and  $\gamma$ , otherwise the species will have gone extinct, or the phylogeny does not reflect local diversification, thus violating a key model assumption.

Arguably more important than the parameter estimates is the model fit and comparison. Our simulation study demonstrates the ability of our method to recover the generating parameters when  $\sqrt{\frac{\alpha}{\gamma}}$  is close to the number of species at the tips of the phylogeny, while our empirical study among three models reveals that different processes can produce similar patterns in traits but result in substantial differences in population sizes. Thus, the difference in the predictions of population sizes for different trait values seems informative for distinguishing between mechanisms. This underscores the importance of implementing population dynamics in trait evolution. Because the fitness function (Eq. 1) can be altered to accommodate any type of species interactions, using our approach for model selection allows one to unravel underlying species interactions.

We find some differences between observed and predicted contrasts (regardless of the set of summary statistics used in ABC), showing a limitation in the ability of our model to predict body lengths. That is, some close relatives show larger contrasts than predicted by our model. This suggests that (1) other traits than body length also determine the strength of competitive interactions, (2) there are multiple optimal body lengths set by the abiotic environment, and/or (3) some of the baleen whales explore different niche dimensions (e.g., they utilize different types of resources), that is, they evolve to adapt to different types of resources to avoid competition rather than experience character displacement to utilize a different part of the current resource spectrum. Our model assumes only one type of resource with only one optimum trait, so it cannot capture these more complex scenarios. For example, the largest underestimation of the contrasts is found in the pair of *B. musculus* and its daughter clade and the pair of *B. physalus* and its sister lineages which seems to imply that the two largest species found niches that favor gigantism (Slater et al. 2017). This mismatch does not mean that competition has not played a role in baleen whale trait evolution: competition may have led to diversification in different niche dimensions. Both explanations involve multiple traits, and hence extending our model to incorporate evolution of multiple traits presents an interesting avenue for further research. There are also other reasons why our

model does not provide a perfect fit. For instance, even though whales are widely distributed, they do not co-occur everywhere and hence direct interactions, as assumed in our model, may be absent. This can be remedied by formulating a spatially explicit model (see e.g. Drury et al. 2017; Manceau et al. 2017; Xu and Etienne 2018). Furthermore, sexual dimorphism occurs in baleen whales, but this difference is relatively small compared to the length difference between species (5%, Ralls and Mesnick 2009). On the data side, there is also considerable uncertainty in phylogenetic reconstruction and trait means, which may have affected our results. Detailed analysis of the effect of such errors can be performed in specific case studies.

With the three models we studied in this paper we have introduced additional mechanisms in the competition kernel. The formulation of the pairwise competition can be written in the general form  $e^{-\alpha(\mu_{i,t}-\mu_{j,t})^2}F(N_{j,t},\mu_{j,t})$ , where  $F(N_{j,t},\mu_{j,t})$  describes the mechanism of interest. For example, the UWC model assumes equal competitive power ( $F(N_{j,t},\mu_{j,t})=1$ ) to all species while the AWC model assumes an abundance-dependent competition ( $F(N_{j,t},\mu_{j,t})=N_{j,t}$ ) and the MWC model sets a metabolism-dependent competition ( $F(N_{j,t},\mu_{j,t})=B_{j,t}$ ). The AWC model agrees with the intuition that the species that can utilize the most available natural resource is the most abundant even if competition is most intense there. The MWC model captures the fact that the relationship between body length and abundance is generally negative (Damuth 1981; Peters and Wassenberg 1983; Peters and Raelson 1984; Damuth 1987). Our general formulation (Eqs 2–4) allows for other ways of weighing the effect of competitive interactions on trait evolution.

Our model can be further extended in various directions. For example, our model currently only incorporates demographic stochasticity; the noise term decreases rapidly with population size. One can easily add environmental noise, which is independent of population size. Other possible extensions would be to replace the single optimum of a single trait by multiple optima. Furthermore, the optimum trait value,  $\theta$ , may be made time-dependent as the abiotic environment (e.g., temperature, resource availability) has obviously not been constant over macroevolutionary time scales. It may even vary on much shorter time scales. For example, Darwin finches on the Galápagos show different beak sizes in different years due to the changing seed size distribution (Lack 1947; Grant and Grant 2006). Another extension of our model would be to model the effect of trait evolution on diversification, along the lines of Aristide and Morlon (2019) who integrated trait evolution in a macroevolutionary model to study the effect of competition on biodiversity and phenotypic diversity (but without population dynamics). Insights from adaptive dynamics on evolutionary branching may be useful here (Geritz et al. 1998). Finally, our model and the ABC-approach also allow for joint inference of

macroevolutionary (speciation and extinction) dynamics and microevolutionary (trait) dynamics.

In summary, we have presented a general framework to model trait evolution subject to species interactions and provided an inference tool to perform model selection based on trait data with or without abundance data. We have shown that the availability of abundance data may be crucial to distinguish between different models. The availability of data on intraspecific trait variation may further aid model selection. Our approach thus represents an important step in the eco-evolutionary synthesis and testing this on empirical data.

#### FUNDING

We thank the Netherlands Organization (NWO) for financial support through a VICI grant awarded to RSE and the China Scholarship Council for financial support of LX.

#### ACKNOWLEDGMENTS

We thank Jonathon Drury, Luke Harmon, and Per Palsbøll for discussions. Furthermore, we thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

#### REFERENCES

- Abrams P.A. 2001. A world without competition. *Nature* 412:858–859.
- Aristide L., Morlon H. 2019. Understanding the effect of competition during evolutionary radiations: an integrated model of phenotypic and species diversification. *Ecol. Lett.* 22:2006–2017.
- Barnett S.A., Simpson G.G. 1955. The major features of evolution, Vol. 46. New York: Columbia University Press.
- Brody S. 1945. Bioenergetics and growth, with special reference to the efficiency complex in domestic animals. New York: Reinhold.
- Brody S., Procter R. 1932. Relation between basal metabolism and mature body weight in different species of mammals. *Univ. Mo. Agr. Exp. Sta. Res. Bull.*
- Clarke M., Thomas G.H., Freckleton R.P. 2017. Trait evolution in adaptive radiations: modeling and measuring interspecific competition on phylogenies. *Am. Nat.* 189:121–137.
- Damuth J. 1981. Body size in mammals. *Nature* 290:699.
- Damuth J. 1987. Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy-use. *Biol. J. Linnean Soc.* 31:193–246.
- Darwin C.R. 1859. On the origin of species by means of natural selection; or the preservation of favoured races in the struggle for life, Vol. 5. London, UK: John Murray.
- Drury J., Clavel J., Manceau M., Morlon H. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Syst. Biol.* 33:700–710.
- Drury J.P., Grether G.F., Garland T., Morlon H., Garland Jr. T., Morlon H., Garland T., Morlon H., Garland Jr. T., Morlon H., Garland T., Morlon H., Garland Jr. T., Morlon H. 2017. An assessment of phylogenetic tools for analyzing the interplay between interspecific interactions and phenotypic evolution. *Syst. Biol.* 67:413–427.
- Etienne R.S., Apol M.E.F., Olff H. 2006. Demystifying the West, Brown & Enquist model of the allometry of metabolism. *Funct. Ecol.* 20:394–399.
- Etienne R. S., Haegeman B., Stadler T., Aze T., Pearson P.N., Purvis A., Phillimore A.B. 2011. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* 279:1300–1309.
- Falconer D.S., Mackay T.F. 1996. Introduction to quantitative genetics. 4th ed. Pearson: Harlow Pearson, Prentice Hall.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15.
- Garland T. 2005. Phylogenetic approaches in comparative physiology. *J. Exp. Biol.* 208:3015–3035.
- Geritz S., Kisdi E., Meszena G., Metz J. 1998. Evolutionary singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol.* 12:35–57.
- Gingerich P.D. 2019. Rates of evolution: a quantitative synthesis. Cambridge and New York: Cambridge University Press.
- Grant P.R., Grant B.R. 2006. Evolution of character displacement in Darwin's finches. *Science* 313:224–226.
- Hansen T.F., Martins E.P. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404.
- Harmon L.J., Andreazzi C.S., Débarre F., Drury J., Goldberg E.E., Martins A.B., Melián C.J., Narwani A., Nuismer S.L., Pennell M.W., Rudman S.M., Seehausen O., Silvestro D., Weber M., Matthews B. 2019. Detecting the macroevolutionary signal of species interactions. *J. Evol. Biol.* 32:769–782.
- Ives A.R., Godfray H.C.J. 2006. Phylogenetic analysis of trophic associations. *Am. Nat.* 168.
- Kimura M., Crow J.F. 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49:725–738.
- Kleiber M. 1947. Body size and metabolic rate. *Physiol. Rev.* 27:511–541.
- Lack D. 1947. Darwin's finches. Cambridge: Cambridge University Press.
- Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Source: Evolution* 30:314–334.
- Lenormand T., Roze D., Rousset F. 2009. Stochasticity in evolution. *Trends Ecol. Evol.* 24:157–165: E1–E14.
- Lockyer C. 1976. Body weights of some species of large whales. *CES J. Marine Sci.* 36:259–273.
- Mahler D.L., Ingram T., Revell L.J., Losos J.B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* 341:292–295.
- Manceau M., Lambert A., Morlon H. 2017. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Syst. Biol.* 6:51–568.
- Narwani A., Matthews B., Fox J., Venail P. 2015. Using phylogenetics in community assembly and ecosystem functioning research. *Funct. Ecol.* 29:589–591.
- Nuismer S.L., Harmon L.J. 2015. Predicting rates of interspecific interaction from phylogenetic trees. *Ecol. Lett.* 18:17–27.
- Pelletier F., Garant D., Hendry A.P. 2009. Eco-evolutionary dynamics. *Philos. Trans. R. Soc. B* 364:1483–1489.
- Pennell M.W., Harmon L.J. 2013. An integrative view of phylogenetic comparative methods: connections to population genetics, community ecology, and paleobiology. *Ann. N. Y. Acad. Sci.* 1289:90–105.
- Peters R.H., Raelson J.V. 1984. Relations between individual size and mammalian population density. *Am. Nat.* 124:498–517.
- Peters R.H., Wassenberg K. 1983. The effect of body size on animal abundance. *Oecologia* 60, 89–96.
- Pigot A.L., Etienne R.S. 2015. A new dynamic null model for phylogenetic community structure. *Ecol. Lett.* 18:153–163.
- Rafferty N.E., Ives A.R. 2013. Phylogenetic trait-based analyses of ecological networks. *Ecology* 94:2321–2333.
- Ralls K., Mesnick S. 2009. Sexual dimorphism. In: Perrin W., Würsig B., Thewissen J.G.M., editors. *Encyclopedia of marine mammals*. 2nd ed. Amsterdam, Boston: Academic Press.
- Raup D.M., Gould S.J., Schopf T.J.M., Simberloff D. 1973. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* 81:525–542.
- Rezende E.L., Jordano P., Bascompte J. 2007. Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos* 116:1919–1929.



- Schoener T.W. 2011. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* 331:426–429.
- Slater G.J., Goldbogen J.A., Pyenson N.D. 2017. Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proc. R. Soc. B* 284:20170546.
- Smith F.A., Boyer A.G., Brown J.H., Costa D.P., Dayan T., Ernest M., Evans A.R., Fortelius M., Gittleman J.L., Marcus J., Harding L.E., Lintulaakso K., Lyons S.K., McCain C., Okie J.G., Saarinen J.J., Sibly R.M., Stephens P.R., Theodor J., Uhen M.D. 2010. Supporting online material for the evolution of maximum body size of terrestrial mammals. *Science* 1216: 1216–1220.
- Sunnåker M., Busetto A.G., Numminen E., Corander J., Foll M., Dessimoz C. 2013. Approximate Bayesian computation. *PLoS Comput. Biol.* 9(1):e1002803. doi:10.1371/journal.pcbi.1002803.
- Theodore Garland J., Ives A.R. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. *Am. Nat.* 155:346.
- Toni T., Welch D., Strelkowa N., Ipsen A., Stumpf M.P. 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* 6:187–202.
- Van Valen L. 1969. Variation genetics of extinct animals. *Am. Nat.* 103:193–224.
- Xu L., Etienne R.S. 2018. Detecting local diversity-dependence in diversification. *Evolution* 72:1–12.