REVIEW

Methods in Ecology and Evolution   BRITISH ECOLOGICAL SOCIETY

Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC)

Mikael Pontarp[1,2,3] (iD) | Åke Brännström[4,5] | Owen L. Petchey[2] (iD)

[1]Department of Biology, Lund University, Lund, Sweden

[2]Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

[3]Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden

[4]Department of Mathematics and Mathematical Statistics, Umeå University, Umeå, Sweden

[5]Evolution and Ecology Program, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria

Correspondence
Mikael Pontarp
Email: mikael.pontarp@biol.lu.se

Handling Editor: Timothée Poisot

Abstract

1. Statistical techniques exist for inferring community assembly processes from community patterns. Habitat filtering, competition, and biogeographical effects have, for example, been inferred from signals in phenotypic and phylogenetic data. The usefulness of current inference techniques is, however, debated as a mechanistic and causal link between process and pattern is often lacking, and evolutionary processes and trophic interactions are ignored.

2. Here, we revisit the current knowledge on community assembly across scales and, in line with several reviews that have outlined challenges associated with current inference techniques, we identify a discrepancy between the current paradigm of eco-evolutionary community assembly and current inference techniques that focus mainly on competition and habitat filtering. We argue that trait-based dynamic eco-evolutionary models in combination with recently developed model fitting and model evaluation techniques can provide avenues for more accurate, reliable, and inclusive inference. To exemplify, we implement a trait-based, spatially explicit eco-evolutionary model and discuss steps of model modification, fitting, and evaluation as an iterative approach enabling inference from diverse data sources.

3. Through a case study on inference of prey and predator niche width in an eco-evolutionary context, we demonstrate how inclusive and mechanistic approaches—eco-evolutionary modelling and Approximate Bayesian Computation (ABC)—can enable inference of assembly processes that have been largely neglected by traditional techniques despite the ubiquity of such processes.

4. Much literature points to the limitations of current inference techniques, but concrete solutions to such limitations are few. Many of the challenges associated with novel inference techniques are, however, already to some extent resolved in other fields and thus ready to be put into action in a more formal way for inferring processes of community assembly from signals in various data sources.

KEYWORDS
biogeography, community assembly, community structure, ecology, evolution, process inference

1 | INTRODUCTION

Community assembly processes are difficult to observe in the field and revealing processes using manipulative experiments is not always feasible. Consequently, there is a considerable need to infer processes from observations, such as trait distributions, species distributions, abundances, and phylogenies (Cadotte et al., 2010; Mouquet et al., 2012; Pausas & Verdu, 2010; Ricklefs & Travis, 1980).

Current techniques that aim to infer assembly processes from community data have limitations. Most methods consider one or a few processes, although community assembly occurs via multiple processes including eco-evolutionary feedbacks (Leibold, Economo, & Peres-Neto, 2010). Patterns observed in nature may be consistent with multiple explanations (Vellend, 2010) and current techniques may thus fail to provide accurate inference, particularly if evolutionary processes and trophic interactions that are known to be important for macroevolution are poorly integrated (Pausas & Verdu, 2010; Pontarp & Petchey, 2016). Fundamental assumptions (e.g., that competition will result in overdispersed trait distributions), on which current inference techniques often rely, have also been questioned (Mayfield & Levine, 2010). Existing inference techniques and their shortcomings are covered in several reviews (Adler, Fajardo, Kleinhesselink, & Kraft, 2013; Cadotte et al., 2010; Cavender-Bares, Kozak, Fine, & Kembel, 2009; Emerson & Gillespie, 2008; Mouquet et al., 2012; Pausas & Verdu, 2010; Vamosi, Heard, Vamosi, & Webb, 2009). Here, we also outline the most relevant features of some of the common inference techniques (Tables 1–2 and Appendix S1), but our major aim is to motivate transformation and improvement of the practice of inferring process from pattern in ecology and evolutionary biology.

Such transformation, already underway, involves models of community assembly and we highlight specific components including mechanistic modelling, parameter estimation, and model selection (see also Cabral, Valente, & Hartig, 2017; Csillery, Blum, Gaggiotti, & Francois, 2010; van der Plas et al., 2015). We present a trait-based and spatially explicit dynamic eco-evolutionary community model of adaptive radiations that includes intra- and interspecific competition, trophic interactions, dispersal as well as trait evolution within trophic levels and co-evolution among trophic levels. We use this model in a case study to illustrate how model modification (including or excluding processes), Approximate Bayesian Computation (ABC) statistics, and multiple data sources can be used for inference. We focus on predator–prey interactions and co-evolution processes that are important in structuring communities but are largely overlooked in traditionally inference techniques (Mouquet et al., 2012).

2 | THE CASE FOR INCLUSIVE AND MECHANISTIC PROCESS INFERENCE

Current inference techniques have both been praised and criticized, and calls for more inclusive and mechanistic approaches have been made due to challenges associated with: (a) basic assumptions on which the methods rely, (b) quantification of processes acting in concert on different spatiotemporal scales, and (c) the identification of particular mechanisms that link process and pattern (Table 1, Appendix S1). We now describe each category of challenges in turn.

2.1 | Basic assumptions of classical approaches

The most common inference methods involve analysis of trait or phylogenetic community patterns, to find signals consistent with community assembly processes. Community overdispersion or underdispersion has been determined by comparing the trait/phylogenetic distribution of the focal community with that of randomly

TABLE 1 Methods that infer assembly processes from community patterns (column 1) and the information that they consider (columns 2–5)

Method	Input data					Inference of processes			
	Phenotype	Phylogeny	Abundance	Environment	Space	Eco.	Evo.	Bio.geo	Reference example
Analysis of phenotypic structure	✓				✓*	✓			Petchey and Gaston (2006)
Analysis of phylogenetic structure	✓*	✓	✓+		✓*	✓			Webb et al. (2002)
Fourth corner	✓	✓	✓	✓	✓*	✓	✓+		Dray and Legendre (2008)
Variance partitioning			✓	✓	✓	✓		✓	Borcard, Legendre and Drapeau (1992)
Co-occurrence analysis			✓			✓			(Harris, 2016)

Tick mark denotes which data/processes are considered explicitly. Superscript denotes data/processes that are implicitly considered (*) or included in extensions of the basic method (+). "Eco" processes include habitat filtering and competitive exclusion; "Evo" includes the evolution of phenotypes, for example, via character displacement; "Bio. geog." includes dispersal limitation.

TABLE 2 Challenges and limitations associated with current inference techniques and methods categorized into three overarching categories (column 1)

Challenge category	Challenge/Limitation	Description	Reference example
Basic assumptions of the methods	Identifying traits that can be used as a proxy for niche	Phenotypic inference techniques assume that trait(s) can be used as a proxy for niche. Identifying the traits that define an organism's niche and thus drive the assembly processes can, however, be difficult and must be supported by expert knowledge of organisms' natural histories. Appropriate weighting of traits can also be difficult to establish	Petchey and Gaston (2006)
	Identifying the relationship between niche, traits, and phylogeny	Phylogenetic inference techniques assume a mapping between relatedness and niche. The interpretation of phylogenetic patterns is contingent on trait evolution and the distribution of traits over the phylogeny. The degree of community clustering given by a certain process will, in other words, be contingent on whether the niche is conserved or labile.	Wiens et al. (2010)
	Assuming a fixed species pool and no evolution or no explicit space	The current theory is mainly focused on local community assembly from a fixed regional species pool and assumes that dispersal is not limiting. This implies that local community composition results only from local processes (e.g., habitat filtering and competition)	Pausas and Verdu (2010)
	Assuming that only similar species exclude each other due to competition.	A current paradigm in ecology states that species that are closely related, share traits, and thus also have similar niches cannot co-exist due to competitive exclusion. Recent studies have, however, showed that competition can sometimes eliminate more different and less related species	Mayfield and Levine (2010)
Method specificity and scale-dependent processes	Single process inference	Most methods only allow for inference on the net effect of processes and do not partition the relative importance of ecological, evolutionary, and spatial processes acting in concert	Petchey (2007)
	Evolution is ignored	Evolutionary contingencies are largely ignored in current inference techniques. This is limiting, in particular when analysing phylogenetic patterns	Emerson and Gillespie (2008)
	Scale-dependent processes	Different traits may be affected by different processes at different temporal and spatial scales.	Trisos, Petchey and Tobias (2014)
Lack of mechanistic links between process and pattern	Unknown mechanisms	Methods that take environmental factors and explicit spatial components into account (e.g., variance partitioning) are mainly phenomenological or statistical. A mechanistic understanding of the causal link between pattern and process is thus often lacking	Gotelli et al. (2009)
	Trophic interactions are often ignored	Even though it is well known that trophic interactions can structure communities, trophic processes are largely ignored when patterns are interpreted	Mouquet et al. (2012)

assembled null communities and interpreted as competitive exclusion or habitat filtering respectively (Webb, Ackerly, McPeek, & Donoghue, 2002). Other methods interpret correlation between traits and environmental variables as a signal of habitat filtering (Legendre, Galzin, & HarmelinVivien, 1997; Mouquet et al., 2012). Furthermore, variance partitioning aims to decompose variation in community composition into fractions determined by environmental factors and spatial location to reveal environmental filtering and spatial contingencies (Borcard, Legendre, & Drapeau, 1992). A fundamental challenge associated with the methods listed above is that a particular trait or set of traits cannot always be assumed to reflect the ecological niche (Trisos, Petchey, & Tobias, 2014). Conclusion based on null models, to which patterns are contrasted, is also criticized (Mittelbach & Schemske, 2015). Even common assumptions about competitive exclusion are contradicted by studies showing that competition can eliminate more different and less related taxa (Mayfield & Levine, 2010)•

2.2 | Processes acting in concert

Other challenges are associated with the complex nature of community ecology (Appendix S2). Communities are structured through multiple processes, acting on different spatiotemporal scales, and different processes can give similar patterns (Vellend, 2010). Existing methods, however, often only infer the net effect of processes or the dominant process. This introduces challenges even in the simplest case when processes on local geographical and ecological temporal scale are considered. The communities may be structured through a combination of environmental filtering and competition (Kraft, Valencia, & Ackerly, 2008), but the relative strength of these processes varies continuously with abiotic and biotic variables (Pontarp, Ripa, & Lundberg, 2012; Trisos et al., 2014). The focus of current inference methods on habitat filtering and competition is also surprising as both empirical (Alto, Malicoate, Elliott, & Taylor, 2012) and theoretical (Pontarp & Petchey, 2016) studies show that trophic

interactions also structure communities. Furthermore, spatial structure can facilitate coexistence of similar species and high dispersal may lead to increased co-occurrence of species. Current methods do, however, rarely consider spatiotemporal dynamics. Finally, current process inference methods focus largely on ecological processes (Leibold et al., 2010) although the importance of evolutionary processes is well known (Cortez & Ellner, 2010; Leibold et al., 2010).

2.3 | Lack of mechanistic inference

The final major challenge associated with current inference techniques that emerge from the literature is their inability to provide explicit information about the mechanistic link between process and pattern (Adler et al., 2013). Lack of information about the underlying mechanisms can lead to misinterpretations of patterns. For example, trophic interactions and evolution are seldom considered in process inference (Pausas & Verdu, 2010) despite evidence of major effects on both trait distributions and phylogenetic patterns (Pontarp & Petchey, 2016, 2018).

3 | IMPLEMENTING MECHANISTIC AND INCLUSIVE APPROACHES

In line with advances in other fields, such as macroecology (Cabral et al., 2017; D'Amen, Rahbek, Zimmermann, & Guisan, 2015) and ecological forecasting (Niu et al., 2014; Urban et al., 2016), more general and flexible methodological inference frameworks can be adopted. Here, we present a generic macroevolutionary modelling framework coupled with Bayesian statistical methods for model

fitting and model improvement (Fig. 1). The framework makes use of *a priori* knowledge for both model construction and parameter estimation. Evolutionary processes can be turned on or off, competitive and/or predator–prey interactions can be included, different types of data can be utilized, and the level of mechanistic detail can be adjusted as needed.

3.1 | Developing an eco-evolutionary model for inference

A community model for inference needs to be flexible, include multiple processes, and output multiple types of data. With this in mind, we base our model on the generalized Lotka–Volterra (GLV) equations (Case, 2000) extended into geographical space (Fig. 2, Appendix S3). This model thus includes ecological and spatial processes and we combine available microevolutionary theory and simulations of macroevolutionary patterns in the form of adaptive radiations.

For simplicity, we assume that space is represented by three patches R, G, and B denoted with subscripts $_R$, $_G$, and $_B$ below. Dispersal between any two patches is assumed symmetrical and occurs at given per capita rates (see Appendix S3 for implementation). The dynamics of n prey populations and m predator populations in patch R are given by:

$$\frac{dV_{i,R}}{V_{i,R}dt} = r(1 - \sum_{j=1}^{n} \frac{\alpha(u_i,u_j)V_{j,R}}{K(u_i,u_{opt})}) - \sum_{k=1}^{m} a\left(u_i,z_k\right)P_{k,R} + m_{R,G}\left(V_{i,G} - V_{i,R}\right) + m_{R,B}\left(V_{i,B} - V_{i,R}\right) \quad (1)$$

$$\frac{dP_{k,R}}{P_{k,R}dt} = -d + c\sum_{i=1}^{n} a\left(u_i,z_k\right)V_{i,R} + m_{R,G}\left(P_{i,G} - P_{i,R}\right) + m_{R,B}\left(P_{i,B} - P_{i,R}\right) \quad (2)$$
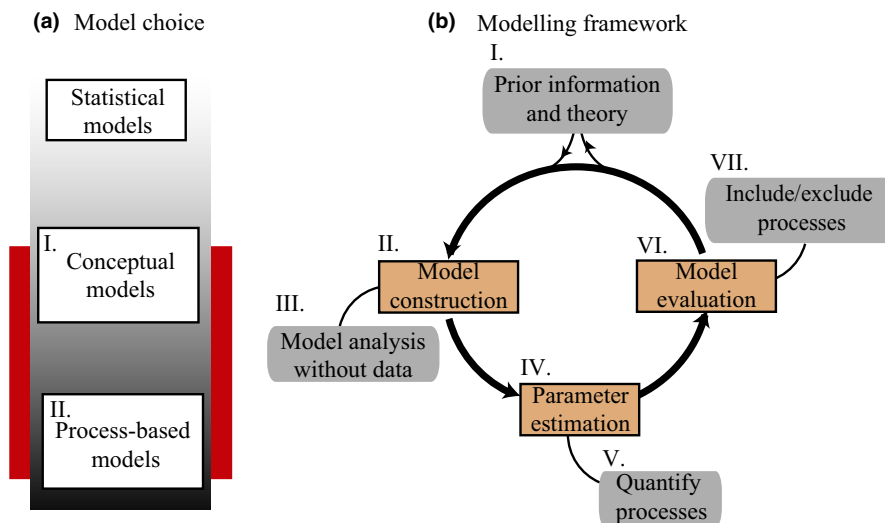

(a) Model choice

(b) Modelling framework

FIGURE 1 The proposed process inference approach including (1) model construction, (2) model fitting and parameter estimation, and (3) model selection. Dynamic community models are suitable as they are often based on simple population dynamical models but can be extended to include mechanisms through trait-based dynamics and population- or individual-based implementation (red section in a). Prior knowledge informs model construction and implementation (b, I, II). Theoretical model investigation can identify different processes that may give rise to similar patterns and thus may be difficult to distinguish between (b, III). Parameter estimation provides quantitative information on the processes that are modelled given the data (b, IV, V) and the model selection procedure guide the model construction and inclusion or exclusion of particular processes (b, VI, VIII)
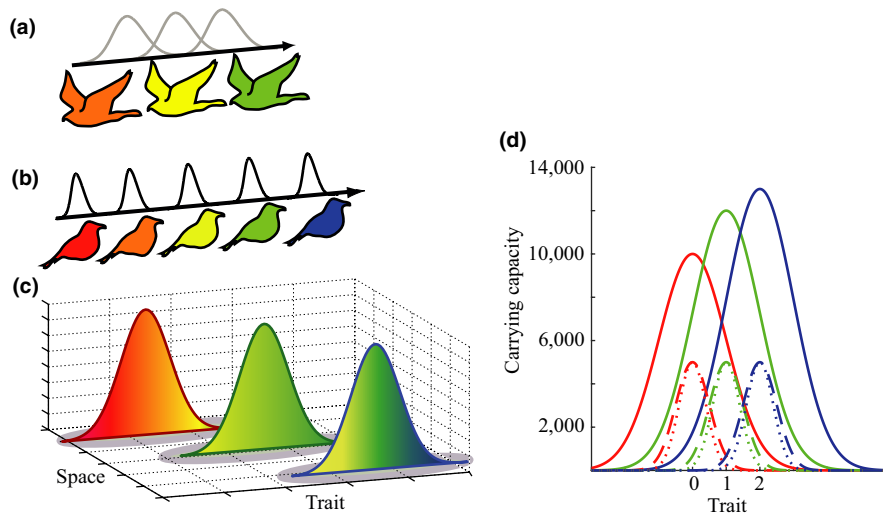
FIGURE 2 Model illustration (a–c) and initial conditions for the most complex model scenario presented this paper (d). Predators (a) and prey (b) can coexist and disperse between three habitats, defined by their resource distributions (c). Species and resources are distributed in trait space (colour gradient) and consumption is dictated by consumer-resource trait matching. As an example, red prey is optimized for utilizing red parts of the resource distribution and green predators are optimized to consume green prey. Competition between species is dictated by their niche width (black and grey kernels). Large overlap between niche kernels indicates high competition and predation pressure respectively. In the most complex case, we initiate the model with three habitats each with their own resource distribution (solid red, green and blue lines in d). Maximum carrying capacity in each habitat is set to 10,000, 12,000, and 13,000 and the peak of the distributions is situated at trait value 0, 1, and 2 respectively. We initiate the system with three prey and three predators (one in each habitat) with trait values equal to the resource distribution peaks. Prey and predators have the same trait value, but we set niche widths for prey (dashed lines) to be slightly wider than predator niche width (dotted lines). Colour coding in (d) denotes habitats and niche kernels are coded according to where the species occurs initially. Note that the y-axis has no direct association with the niche kernels in d

for $i = 1$ to n and $k = 1$ to m. Here, $V_{i,s}$ and $P_{k,s}$ denote prey and predator population size in patch s, respectively, with s referring to patch R, G, or B. The equations for the other two patches are similar. The parameters r and d are the intrinsic prey growth rate and the predator mortality rate, respectively. Here, we assume that all patches are similar, but variation can be implemented through species- and/or patch-specific growth and death (formulated as $r_{i,s}$ and $d_{k,s}$).

The trait-dependent functions on the right-hand side of Equations 1 and 2 are given by:

$$K\left(u_i, u_{\mathrm{opt}}\right) = K_0 e^{-\frac{\left(u_{\mathrm{opt}} - u_i\right)^2}{2\sigma_K^2}}, \tag{3}$$

$$\alpha\left(u_i, u_j\right) = e^{-\frac{\left(u_i - u_j\right)^2}{2\sigma_a^2}}, \tag{4}$$

and

$$\left(u_i, z_k\right) = b_{\max} e^{-\frac{\left(u_i - z_k\right)^2}{2\sigma_a^2}}. \tag{5}$$

where $K(u_i, u_{\mathrm{opt}})$ represents the carrying capacity for a monomorphic population i of prey individuals with trait value u_i in a habitat characterized by a resource distribution with its peak resource availability at the point u_{opt}. Resource availability declines symmetrically as u deviates from u_{opt} according to σ_K. The interaction, $\alpha(u_i, u_j)$, between prey populations i and j is modelled with a Gaussian function where σ_α determines the degree of competition between populations with traits u_i and u_j and can thus be viewed

as the niche width of the prey. Equation 5 models the interaction, $a(u_i, z_k)$, between a focal predator population k with trait value z and a prey population i with trait value u. The parameter $b_{\max}$ denotes the maximum attack rate obtained when $u_i = z_k$ and this rate then falls of symmetrically as u_i deviates from z_k according to σ_a. Similar to the σ_α parameter, σ_a can be viewed as the niche width of the predator. Again parameters $b_{\max}$ and K_0 can be made patch- and species-specific if needed.

The model is flexible enough to model newly established communities of organisms with low evolutionary potential by seeding only one patch with interacting populations and by omitting dispersal. With this being said, for more complex systems, explicit space and dispersal may be required. For old communities or fast-evolving organisms (e.g., microbes), implementing evolutionary process is desirable. The model is thus able to capture a range of scenarios ranging from ecological assembly across space to macroevolutionary processes where, for example, ecological opportunity is followed by adaptive radiations. For this, we use an adaptive dynamics approach (Brännström, Johansson, & von Festenberg, 2013; Geritz, Kisdi, Meszena, & Metz, 1998; Metz, Geritz, Meszena, Jacobs, & Van Heerwaarden, 1996) (see also Appendix S3). With full complexity, the simulation model includes intra- and interspecific competition, trophic interactions, dispersal, and trait evolution, and can exhibit evolutionary branching (Fig. 2). The model outputs population dynamics, equilibrium population sizes, and trait distributions for each evolutionary step (Fig. 3). By assigning a species identity to

individuals, for example, using a trait-based species definition (see also Pontarp, Ripa, & Lundberg, 2015; Pontarp et al., 2012), we can follow trait evolution, diversity, and phylogenetic and phenotypic community structure throughout evolutionary history (see also Pontarp & Petchey, 2018). Code for the model implementation is provided at https://zenodo.org/badge/latestdoi/158187026 and a guide to model modification is found in Appendix S5.

3.2 | Quantifying assembly processes through parameter estimation and model selection

While classical parameter estimation methods assign each parameter a given value or range of values based on the data provided, Bayesian methods additionally make use of prior knowledge of reasonable parameter values to arrive at parameter estimates. This knowledge takes the form of a prior probability of a given model and prior distribution for each model parameter. The data are then used to determine corresponding posterior distributions that reflect the best estimate of each parameter value and model

selection is made through Bayes factors (Appendix S4). The uncertainty associated with the distributions provides information about model sensitivity or lack of signal in data for a given parameter while comparisons of estimates among parameters help reveal the relative strength and importance of different processes. Furthermore, by each time evaluating an increasingly complicated model (Fig. 1), it is possible to circumvent potential problems of using an overly complex model.

There are several benefits associated with using the proposed approach (Fig. 1). First, and before the models are fitted to real data, a theoretical model investigation can identify different processes that may give rise to similar patterns. In this case, it will be difficult to distinguish between those processes and additional information or experimentation may be needed for successful inference. Second, it is possible to evaluate the model and inference techniques by simulating data and then using the suggested inference method to retrieve known parameters. The amount of prior information about the system necessary for correct inference can be evaluated. If processes cannot be inferred in simulated data,
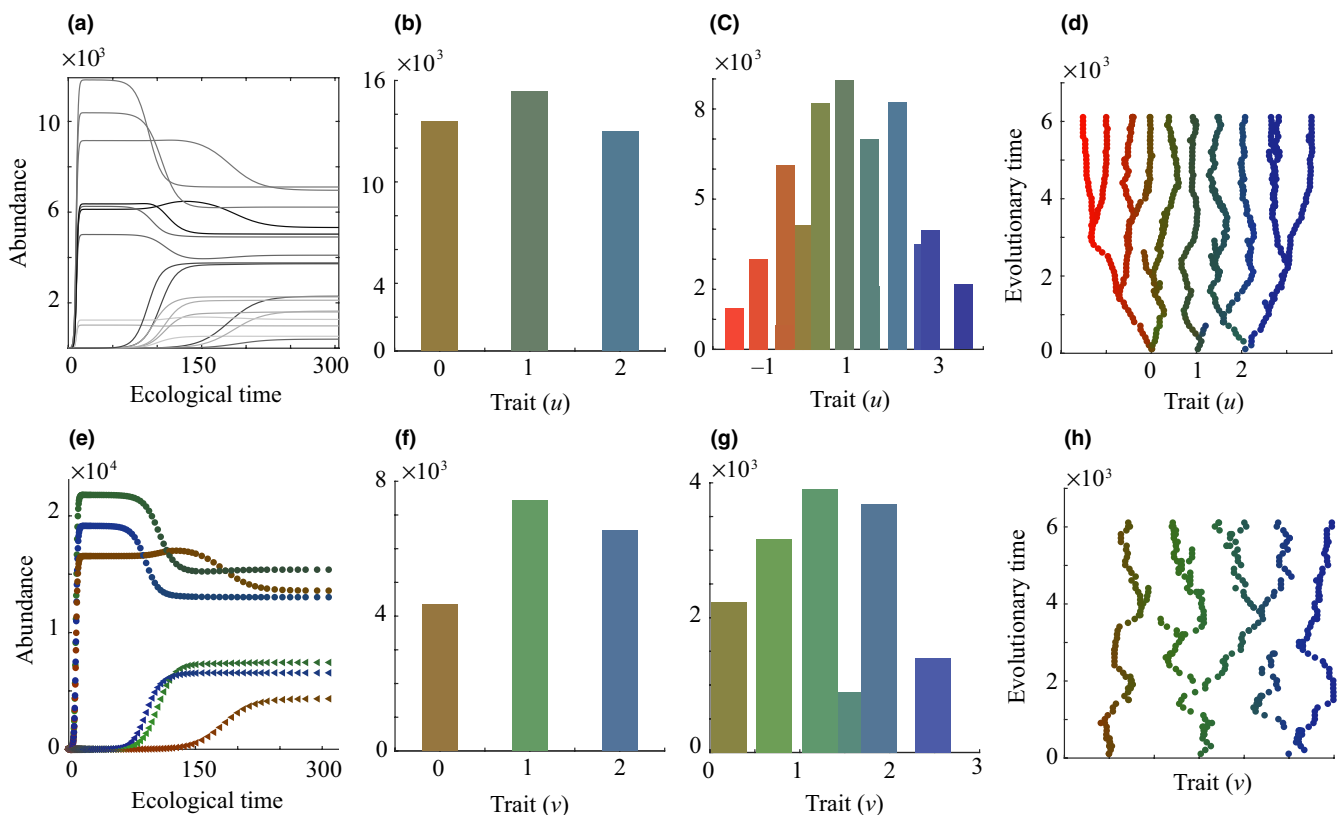

FIGURE 3 Representative data outputs from the model, including time series (a, e), trait distributions (n, c, f, g), and adaptive radiations (d, h) for prey (a, e, b, c, d) and predators (a, e, f, g, h). We compute ecological dynamics and equilibrium for each subpopulation in each habitat (a) and for each prey (circles) and predator (triangles) population across habitats (e). Trait distributions at equilibrium for the initiated community are shown for the initiated community of three prey (b) and three predators (f) distributed in space. The spatial distribution and trait distributions of prey (c) and predators (g) evolved through adaptive radiations of co-evolving prey (d) and predators (h). Colour coding in (e) illustrates spatial distribution. Pure red, green, and blue denote sole occupancy in habitat A, B, and C, respectively. Occupancy in multiple habitats is illustrated as a mix of colours proportional to the spatial distribution. Parameters for this simulation are $K_{0,A} = 10,000$; $K_{0,B} = 12,000$; $K_{0,C} = 13,000$; $u_{opt,A} = 0$; $u_{opt,B} = 1$; $u_{opt,C} = 2$; $\sigma_K = 1$; $\sigma_\alpha = 0.5$; $\sigma_a = 0.4$; $r = 1$; $d = -0.2$; $cp = 0.3$; $b_{max} = 0.0001$; $M = 0.05$; $P_{mut,prey} = 0.01$; $P_{mut,pred} = 0.1$; $\sigma_{mut,prey} = 0.02$; $\sigma_{mut,pred} = 0.03$, with initial conditions $u_1 = 0$; $u_2 = 2$; $u_3 = 3$; $v_1 = 0$; $v_2 = 2$; $v_3 = 3$

even though the model that underlies the patterns is known, correct inference on non-simulated (real) data is unlikely. Third, while fitting models to data (Fig. 1), one can evaluate a model that includes fewer ecological processes against models that include more processes and thus guiding the inclusion or exclusion of particular processes.

Approximate Bayesian Computation (ABC) is a parameter estimation and model fitting technique (Beaumont, 2010; Csillery et al., 2010; Sisson, Fan, & Tanaka, 2007) with promise for overcoming the difficulty of fitting complex models to diverse data. ABC takes priors for each model parameter as input, data are simulated with the model many times given parameters that are each time drawn randomly from the priors, and then, the distance (often Euclidean distance) between model output and observed data is evaluated through a set of summary statistics. Parameter combinations that render small distance (defined through some user-defined threshold) constitute the posterior distributions. The mean, mode, variance, and other metrics of the posteriors define the actual parameter estimates. The simplest way of generating posterior distributions is through a simple rejection algorithm where parameters are drawn from a constant prior, but more powerful global optimization techniques such as Kalman filters (Kalman, 1960), Markov chain Monte Carlo (Gao et al., 2011), or Sequential Monte Carlo (Sisson et al., 2007) are commonly used. The model selection procedure is based on the same general principles, except that the summary statistics is defined somewhat differently (Prangle, Fearnhead, Cox, Biggs, & French, 2014). The output from the ABC model selection is focused on acceptance/rejection ratio between models rather than posterior parameter distributions (Liepe et al., 2014; Toni, Welch, Strelkowa, Ipsen, & Stumpf, 2009).

3.3 | Inferring competitive and trophic interactions—an illustrative case study

We asked whether it is possible to use the framework, model, and tools presented above to discriminate and quantify competition and trophic interactions from data on trait distributions, phylogenetic data, and species abundances. The answer to this question is non-trivial for two reasons. First, competition and predation have been shown to have similar effects on trait distributions which may limit inference of the two processes (Pontarp & Petchey, 2016). Second, as far as we know, no one has considered predatory effects on prey community structure in an inference framework before.

We generate our own observed, "empirical", data through simulations (Zurell et al., 2010) (Fig. 4). We start simulations with monomorphic population(s) at the centre of the resource distribution and we simulate the eco-evolutionary assembly of four local consumer communities for 5,000 evolutionary time steps. Our simulated data are thus representing a macroevolutionary scenario where ecological opportunity facilitates diversification through adaptive radiations. Two of the observed communities includes competitive prey only (no predators) and we set prey niche width $\sigma_\alpha = 0.1$ and $\sigma_\alpha = 0.3$ for the two communities respectively. In our simulations, we also set the width of the resource distribution $\sigma_K = 1$ which means that diverse communities can evolve (Geritz et al., 1998). Furthermore, we simulate the assembly of two communities of co-evolving predators and prey. We use the same values for σ_α as above and we set predator niche width $\sigma_a = 0.7$. Constant model parameters for all simulations were $K_0 = 10,000$; $\sigma_K = 1$; $r = 1$; $d = 0.2$; c = 0.3; $b_{max} = 0.0001$. Mutation size for both predators and prey was defined as a random trait value drawn from a normal distribution with mean equal to the
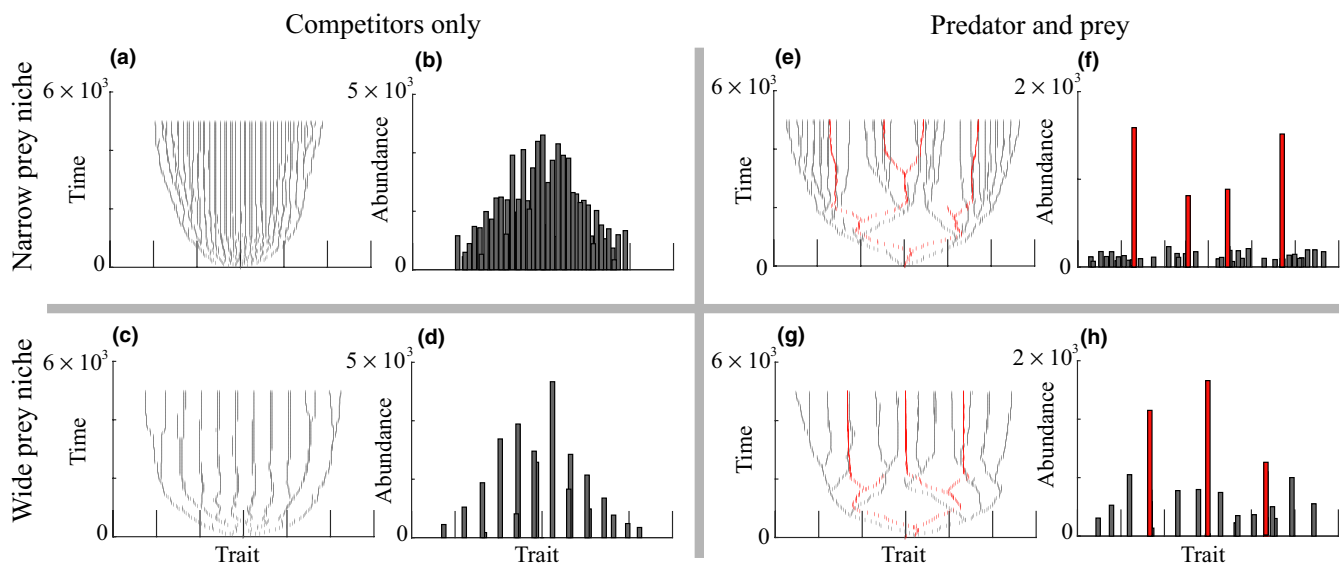

FIGURE 4 Four simulated communities used as "observed" data in our illustrative case study. Radiations and trait distributions for prey (grey) and predators (red). Two of the communities contain competitive consumers only (a–d) with low (0.1) niche width (a,b) and high (0.3) niche width (c,d). Two communities are assembled through co-evolving competitive consumers and predators (e–h). The competitors in the predator–prey communities have the same high (e, h) and low (g, h) niche widths. The predators have niche width

trait value of the mutating population and a variance (σ_{mut}) equal to 0.02 and mutation probability for the prey and predators were kept constant at 0.01. We will refer to data, including species richness, species abundance, trait distributions, and phylogenetic structure, from these four communities as "observed data" from now on.

To infer the processes that underpin patterns in observed data, we consider prior information about the observed system and available theory while constructing the candidate model scenarios (Fig. 1). For this case, we assume that the system is known well enough for us to focus on eco-evolutionary predator–prey interactions and we use established predator–prey models in our simulations. To illustrate how model modification can facilitate process inference, we generate model output for two candidate models: one that includes consumer competition only and one that includes competition and predation. We also implement the model such that multiple data types are produced and we generate model output given parameters from wide non-informative prior distributions. We infer the processes for each of our four observed communities by selecting the best model and by approximating the posterior parameter distribution for the parameters that are relevant for prey competition (prey niche width) and predator niche width, using ABC and a rejection algorithm (Beaumont, 2010; Sisson et al., 2007). As summary statistics for the ABC, we utilize as much of the available data as possible. We

use well-known community structure metrics in the field of inference, including the number of prey species, mean abundance, the width of the trait distribution, mean trait distance, mean nearest trait distance, net relatedness index, and nearest phylogenetic taxon index (Appendix S4). The output from the ABC analysis constitutes the distance between simulated and observed data, evaluated as the Euclidean distance between summary statistics, as a function of model structure and model parameters (Figs S1–S3, Appendix S4). The Euclidean distance approximates the likelihood for a given model and parameter set (Beaumont, 2010; Liepe et al., 2014; Prangle et al., 2014), and hence also the posterior parameter distribution (Fig. 5).

We find that the minimum Euclidean distance between the observed data and the correct model is always smaller than the distance obtained by fitting the wrong model, and the acceptance ratio provided by the rejection algorithm is always larger for the correct model compared to the wrong model (Figs S1–S3, Appendix S4). The correct model is thus selected for each of the four observed communities, suggesting that prey only communities can indeed be distinguished from predator–prey systems. This is encouraging as previous studies (Pausas & Verdu, 2010; Pontarp & Petchey, 2016; Schluter, 1984) have raised concerns about inferring predator–prey interactions, using traditional techniques. The parameter estimates were also largely correct for the best
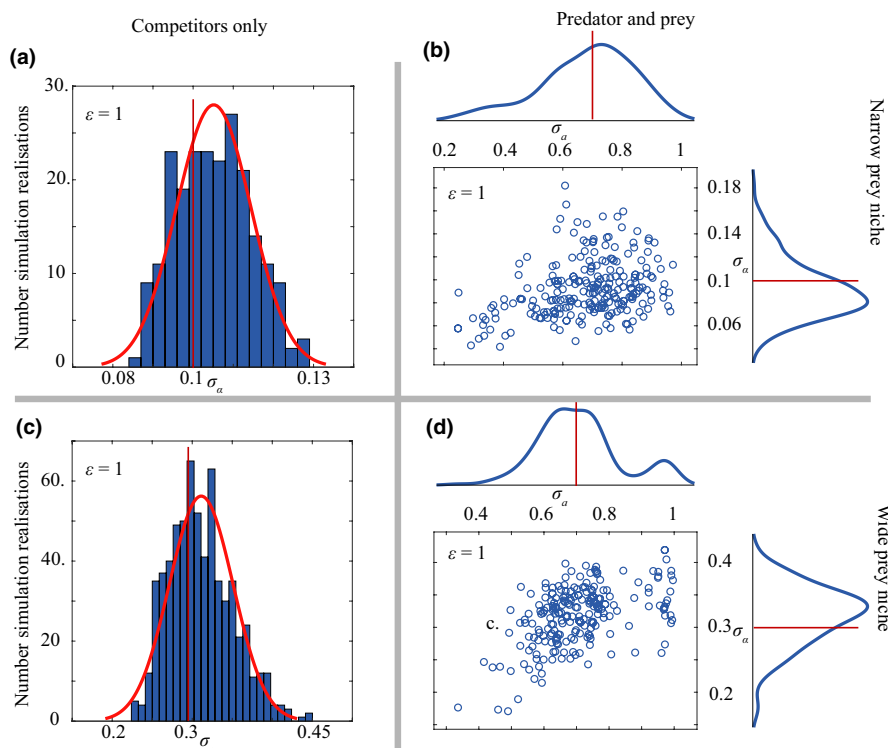

FIGURE 5 Posterior distributions for each of the prey only, predator–prey, narrow prey niche width, and wide prey niche width scenarios, when the correct models were fitted to observed data. Prey niche width was estimated for the prey only scenario (a,c); hence, the one-dimensional posterior distribution plotted as a histogram with superimposed fitted normal density (red curve) for the number of simulation realizations that did fall within Euclidean threshold value (ε) equal to 1. In predator–prey scenarios, both predator and prey niche width were estimated, hence the two-dimensional posterior distribution (b,d). The posterior is shown as a scatter plots with the parameter combinations that did fall within ε =1 together with marginal distributions illustrated as blue curves. Red lines denote the correct parameter values in true data

performing models (Fig. 5), showing that quantitative measures of the strength of competition and predation can be obtained in a scenario described by our case study. This is an advantage over more traditional inference where quantitative or relative measures of processes are rarely presented.

4 | DISCUSSION

Ecological communities are complex, with diverse processes and actors (e.g., Urban, De Meester, Vellend, Stoks, & Vanoverbeke, 2012; Vellend, 2010) and it is clear that several of the current inference techniques are too simplistic (Adler et al., 2013; Cadotte et al., 2010; Cavender-Bares et al., 2009; Emerson & Gillespie, 2008; Mouquet et al., 2012; Pausas & Verdu, 2010; Vamosi et al., 2009). A novel, more mechanistic, more inclusive, and more unified approach, like the one proposed here, for future assembly-process inference techniques is desirable as this will allow for a causal and quantitative link between multiple processes and community patterns.

Although the approach has not been synthesized for process inference explicitly before, inference does seem to be moving in the proposed direction (e.g., van der Plas et al., 2015). Several modelling approaches (Harris, 2016; Jabot & Bascompte, 2012; May, Giladi, Ristow, Ziv, & Jeltsch, 2013) and statistical techniques can be used, but the trade-off in model complexity and tractability needs to be considered. We present dynamical eco-evolutionary modelling in combination with ABC as particularly suitable. Dynamic modelling is simple in the sense that it is based on simple, often phenomenological population dynamical models (Brännström et al., 2012), but they can be extended to include detailed mechanisms through the inclusion of trait-based dynamics and through a population- or individual-based implementation. A major advantage is a clear link between ecological and microevolutionary processes and macroevolutionary patterns. The trade-off between realism, computational costs, and model tractability can be monitored and controlled as the models gain in complexity. By iterating over model construction and model evaluation, each time evaluating different models that include/exclude particular assembly processes, a type of robustness check can be accomplished and the best model for the data identified. Similarly, by evaluating increasingly complex versions of the models, the optimal model for inference about the study system can be found (see details in Appendix S4). Any model that includes several processes will, however, tend to be complex and computationally costly. Likelihood functions are often intractable, leading to the need for fitting and model selection techniques like ABC.

In our case study, we illustrate how dynamical modelling and ABC can be used for inference, through model selection and parameter estimation. We test multiple models against simulated data and the correct models were selected for our observed data, indicating that the different processes can be distinguished between with a simple rejection algorithm (Sisson et al., 2007) and without explicit summary statistics optimization (Prangle et al., 2014). Our case study emphasizes the importance to include the correct processes and thus to test different models. For example, a predator–prey system that is evaluated by a prey only model (Fig. S4c, Appendix S4) can lead to seemingly good estimates of prey niche width, but the overall model fit compared to the true model is relatively poor with overestimation of the niche width parameter. These results are promising and they indicate that our suggested approach is working under the ideal conditions associated with our case study. It will be intriguing to see how the approach will perform in practice.

Our case study also emphasizes the possibility and value of taking prior information and multiple data sources into account. The fact that we are using both phylogenetic and trait distribution data seems to be one of the reasons why we are able to distinguish between predator and prey processes as such signals can be diffuse in trait data only (Pontarp & Petchey, 2016). This also highlights the empirical side of inference, namely data to which the models are fitted. As noted above, data inform the models and are thus imperative for the model fitting. Data also dictate model construction as the model output needs to be comparable with data (e.g., diversity, size distributions, or phylogenetic patterns). In our case study, we are thus not only illustrating the technical aspects of novel inference but also provide an indication of what type of data may be useful in the future for inference to become more inclusive and detailed. Furthermore, data provide the knowledge of the natural history of the study system that also informs model construction. A priori information of a particular system can narrow down the priors for ABC and thus facilitate parameter estimation by reducing the parameter space that needs to be searched in the optimization procedure. For certain systems, reasonable parameter values may already be available in the literature and it might be possible to measure some parameters in independent studies. With this in mind, we envision that process inference will continue to move away from simple statistical and non-mechanistic inference techniques for approaches with a constant flow of information between experimental and field data, model construction, parameter estimation, and model selection. This will come with technical challenges as well as increased demands on data, computational power, and experimental progress. Many of these difficulties are, however, already identified and to some extent resolved.

AUTHORS' CONTRIBUTIONS

DATA ACCESSIBILITY

The model and the ABC analysis presented in the main text are implemented as MATLAB (version R2017b) code. All code is publicly available at https://zenodo.org/badge/latestdoi/158187026 and described in Appendix S5. No empirical or other previously published data were used.

ORCID

Mikael Pontarp https://orcid.org/0000-0001-9045-7829

Owen L. Petchey https://orcid.org/0000-0002-7724-1633

REFERENCES

Adler, P. B., Fajardo, A., Kleinhesselink, A. R., & Kraft, N. J. B. (2013). Trait-based tests of coexistence mechanisms. *Ecology Letters*, 16, 1294–1306.

Alto, B. W., Malicoate, J., Elliott, S. M., & Taylor, J. (2012). Demographic consequences of predators on prey: Trait and density mediated effects on mosquito larvae in containers. *PLoS ONE*, 7, 1–8.

Beaumont, M. A. (2010). Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(41), 379–406.

Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73, 1045–1055.

Brännström, A., Johansson, J., Loeuille, N., Kristensen, N., Troost, T. A., Lambers, R. H. R., & Dieckmann, U. (2012). Modelling the ecology and evolution of communities: A review of past achievements, current efforts, and future promises. *Evolutionary Ecology Research*, 14, 601–625.

Brännström, Å., Johansson, J., & von Festenberg, N. (2013). The hitchhiker's guide to adaptive dynamics. *Games*, 4, 304–328.

Cabral, J.S., Valente, L., & Hartig, F. (2017) Mechanistic simulation models in macroecology and biogeography: State-of-art and prospects. *Ecography*, 40, 267–280.

Cadotte, M. W., Davies, T. J., Regetz, J., Kembel, S. W., Cleland, E., & Oakley, T. H. (2010). Phylogenetic diversity metrics for ecological communities: Integrating species richness, abundance and evolutionary history. *Ecology Letters*, 13, 96–105.

Case, T. J. (2000). *An illustrated guide to theoretical ecology.* Oxford New York: Oxford University Press Inc.

Cavender-Bares, J., Kozak, K. H., Fine, P. V. A., & Kembel, S. W. (2009). The merging of community ecology and phylogenetic biology. *Ecology Letters*, 12, 693–715.

Cortez, M. H., & Ellner, S. P. (2010). Understanding rapid evolution in predator-prey interactions using the theory of fast-slow dynamical systems. *American Naturalist*, 176, E109–E127.

Csillery, K., Blum, M. G. B., Gaggiotti, O. E., & Francois, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*, 25, 410–418.

D'Amen, M., Rahbek, C., Zimmermann, N.E., & Guisan, A. (2015) Spatial predictions at the community level: From current approaches to future frameworks. *Biological Reviews*, 92, 169–187.

Dray, S., & Legendre, P. (2008). Testing the species traits-environment relationships: The fourth-corner problem revisited. *Ecology*, 89, 3400–3412.

Emerson, B. C., & Gillespie, R. G. (2008). Phylogenetic analysis of community assembly and structure over space and time. *Trends in Ecology & Evolution*, 23, 619–630.

Gao, C., Wang, H., Weng, E. S., Lakshmivarahan, S., Zhang, Y. F., & Luo, Y. Q. (2011). Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. *Ecological Applications*, 21, 1461–1473.

Geritz, S. A. H., Kisdi, E., Meszena, G., & Metz, J. A. J. (1998). Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology*, 12, 35–57.

Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., … Willig, M. R. (2009). Patterns and causes of species richness: A general simulation model for macroecology. *Ecology Letters*, 12, 873–886.

Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, 97, 3308–3314.

Jabot, F., & Bascompte, J. (2012). Bitrophic interactions shape biodiversity in space. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 4521–4526.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic engineering*, 82, 35–45.

Kraft, N. J. B., Valencia, R., & Ackerly, D. D. (2008). Functional traits and niche-based tree community assembly in an amazonian forest. *Science*, 322, 580–582.

Legendre, P., Galzin, R., & HarmelinVivien, M. L. (1997). Relating behavior to habitat: Solutions to the fourth-corner problem. *Ecology*, 78, 547–562.

Leibold, M. A., Economo, E. P., & Peres-Neto, P. (2010). Metacommunity phylogenetics: Separating the roles of environmental filters and historical biogeography. *Ecology Letters*, 13, 1290–1299.

Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P., & Stumpf, M. P. H. (2014). A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9, 439–456.

May, F., Giladi, I., Ristow, M., Ziv, Y., & Jeltsch, F. (2013). Metacommunity, mainland-island system or island communities? Assessing the regional dynamics of plant communities in a fragmented landscape. *Ecography*, 36, 842–853.

Mayfield, M. M., & Levine, J. M. (2010). Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecology Letters*, 13, 1085–1093.

Metz, J. A., Geritz, S. A., Meszena, G., Jacobs, F. J., & Van Heerwaarden, J. S. (1996). Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *Stochastic and Spatial Structures of Dynamical Systems*, 45, 183–231.

Mittelbach, G. G., & Schemske, D. W. (2015). Ecological and evolutionary perspectives on community assembly. *Trends in Ecology & Evolution*, 30, 241–247.

Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L. F., Chave, J., … Thuiller, W. (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87, 769–785.

Niu, S.L., Luo, Y.Q., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J.W., & Chapin, F.S. (2014) The role of data assimilation in predictive ecology. *Ecosphere*, 5, 1–16.

Pausas, J. G., & Verdu, M. (2010). The jungle of methods for evaluating phenotypic and phylogenetic structure of communities. *BioScience*, 60, 614–625.

Petchey, O. L. (2007). Effects of environmental variability on ecological communities: Testing the insurance hypothesis of biodiversity in

aquatic microcosms. *Impact of Environmental Variability on Ecological Systems*, 2, 179–196.

Petchey, O. L., & Gaston, K. J. (2006). Functional diversity: Back to basics and looking forward. *Ecology Letters*, 9, 741–758.

van der Plas, F., Janzen, T., Ordonez, A., Fokkema, W., Reinders, J., Etienne, R. S., & Olff, H. (2015). A new modeling approach estimates the relative importance of different community assembly processes. *Ecology*, 96, 1502–1515.

Pontarp, M., & Petchey, O. L. (2016). Community trait overdispersion due to trophic interactions: concerns for assembly process inference. *Proceedings of the Royal Society B Biological Sciences*, 283, 20161729. https://doi.org/10.1098/rspb.2016.1729

Pontarp, M., & Petchey, O. L. (2018). Ecological opportunity and predator–prey interactions: Linking eco-evolutionary processes and diversification in adaptive radiations. *Proceedings of the Royal Society B*, 285, 20172550. https://doi.org/10.1098/rspb.2017.2550

Pontarp, M., Ripa, J., & Lundberg, P. (2012). On the origin of phylogenetic structure in competitive metacommunities. *Evolutionary Ecology Research*, 14, 269–284.

Pontarp, M., Ripa, J., & Lundberg, P. (2015). The biogeography of adaptive radiations and the geographic overlap of sister species. *American Naturalist*, 186, 565–581.

Prangle, D., Fearnhead, P., Cox, M. P., Biggs, P. J., & French, N. P. (2014). Semi-automatic selection of summary statistics for ABC model choice. *Statistical Applications in Genetics and Molecular Biology*, 13, 67–82.

Ricklefs, R. E., & Travis, J. (1980). A morphological approach to the study of avian community organization. *Auk*, 97, 321–338.

Schluter, D. (1984). A variance test for detecting species associations, with some example applications. *Ecology*, 65, 998–1005.

Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 1760–1765.

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

Trisos, C. H., Petchey, O. L., & Tobias, J. A. (2014). Unraveling the interplay of community assembly processes acting on multiple niche axes across spatial scales. *American Naturalist*, 184, 593–608.

Urban, M.C., Bocedi, G., Hendry, A.P., Mihoub, J.B., Pe'er, G., Singer, A., … Travis, J.M.J. (2016) Improving the forecast for biodiversity under climate change. *Science*, 353, aad8466. https://doi.org/10.1126/science.aad8466

Urban, M. C., De Meester, L., Vellend, M., Stoks, R., & Vanoverbeke, J. (2012). A crucial step toward realism: Responses to climate change from an evolving metacommunity perspective. *Evolutionary Applications*, 5, 154–167.

Vamosi, S. M., Heard, S. B., Vamosi, J. C., & Webb, C. O. (2009). Emerging patterns in the comparative analysis of phylogenetic community structure. *Molecular Ecology*, 18, 572–592.

Vellend, M. (2010). Conceptual synthesis in community ecology. *Quarterly Review of Biology*, 85, 183–206.

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, 33, 475–505.

Wiens, J. J., Ackerly, D. D., Allen, A. P., Anacker, B. L., Buckley, L. B., Cornell, H. V., … Stephens, P. R. (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology Letters*, 13, 1310–1324.

Zurell, D., Berger, U., Cabral, J. S., Jeltsch, F., Meynard, C. N., Munkemuller, T., … Grimm, V. (2010). The virtual ecologist approach: Simulating data and observers. *Oikos*, 119, 622–635.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.