Title: Statistical methods to identify mechanisms in studies of eco-evolutionary dynamics

Authors: Jelena H. Pantel[1,3], Lutz Becks[2]

[1]Ecological Modelling, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße 5, 45141 Essen, Germany

[2]University of Konstanz, Aquatic Ecology and Evolution, Limnological Institute University of Konstanz Mainaustraße 252 78464 Konstanz / Egg, Germany

[3]Corresponding author: Pantel, J.H. (jelena.pantel@uni-due.de). Laboratory website: www.jhpantel.com. Twitter: @jhpantel

Keywords:
- Hypothesis testing
- Biodiversity
- Bayesian statistics
- Eco-evolutionary dynamics
- Mechanistic models

Abstract:

While the reciprocal effects of ecological and evolutionary dynamics are increasingly recognized as an important driver for biodiversity, detection of such eco-evolutionary feedbacks, their underlying mechanisms, and their consequences remains challenging. Eco-evolutionary dynamics occur at different spatial and temporal scales and can leave signatures at different levels of organization (e.g., gene, protein, trait, community) that are often difficult to detect. Recent advances in statistical methods combined with alternative hypothesis tests provide a promising approach to identify potential eco-evolutionary drivers for observed data even in non-model systems that are not amenable to experimental manipulation. We discuss recent advances in eco-evolutionary modelling and statistical methods and discuss challenges for fitting mechanistic models to eco-evolutionary data.

Introduction

The field of eco-evolutionary dynamics centers on the paired observations that evolution can occur on overlapping timescales with ecological processes, and that ecological processes are influenced by traits that can both respond to and drive evolutionary change. Eco-evolutionary feedbacks occur when evolution of a trait or property impacts population or community dynamics (or vice-versa), which feeds back to drive further evolution (or ecological dynamics), in a continued cycle [1,2]. These eco-evolutionary dynamics need not be invoked if the timescales of ecological and evolutionary dynamics are sufficiently separate or if phenotypic variance is low enough that a mean phenotypic value captures ecological dynamics sufficiently [1]. However even with this clear definition, it remains that all biological systems experience evolution, and these evolutionary processes are occurring while ecological dynamics proceed. In the absence of intertwined feedback loops and especially in their presence, past or contemporary evolution can impact the dynamics of ecological processes. Studies have identified instances of eco-evolutionary interactions and feedbacks and demonstrated their strong impacts on populations, communities, and ecosystems [3,4,5, 6], but those impacts can be difficult to detect due to often complex relationship of ecological and evolutionary processes. Thus, determining potential eco-evolutionary drivers of population, community, and ecosystem dynamics remains a major challenge. This restricts our ability to evaluate the consequences of eco-evolutionary dynamics on contemporary and future biodiversity in natural settings and in non-model organisms.

Some studies have implemented methods (e.g. [7,8,9]) that ultimately take a target variable of interest - usually a phenotypic trait or vital rate shared across the species under study - constructs categories of contributing factors to variation in that target variable, and assigns total phenotypic variance to these fractions (often by holding one process constant and averaging across the change in the other effects; e.g. [10]). These fractions are linear and additive, but lack a mechanistic basis, and are calculated retrospectively, which means they are limited for predicting future changes. Analysis methods for the study of eco-evolutionary dynamics should instead consider the mechanistic basis of data structure, realistic models of variance and uncertainty, and the complexity of eco-evolutionary feedbacks and processes with distinct signatures at different temporal, spatial, and biodiversity scales (**Box 1**). Model-based hypothesis testing by comparing among alternative eco-evolutionary hypotheses has recently emerged as an alternative to existing correlative approaches and this framework can be used to address the unanswered questions in the field [11,12,13]. We summarize here the basic building blocks of using mechanistic models and statistical inference to pair eco-evolutionary data with a theoretical or simulation model. We provide guidance on how to construct alternative hypotheses, fit these to observed data, and generate posterior probabilities for these hypothesized processes. The goal of statistical inference for eco-evolutionary dynamics is not that all datasets must perfectly capture each step of historical dynamics, nor that all datasets must be fit to the same ANOVA-based linear model to categorize average effect sizes. We provide directions to make inference about the potentially complex processes that structure eco-evolutionary data, to identify whether eco-evolutionary feedbacks are operating in a specific system, and to determine the consequences of these dynamics for critical emergent properties such as resilience, stability, diversity, extinction, or ecosystem function.

A structured workflow for model-based hypothesis testing in eco-evolutionary dynamics

Eco-evolutionary research questions may center on whether only evolutionary or only ecological processes in isolation are sufficient to account for observed patterns of diversity or whether ecological and evolutionary processes interact in a feedback loop that results in novel dynamics and diversity patterns. Research questions can thus be formulated as a core of null and alternative hypotheses expressed in the form of alternative competing mechanistic models (**Figure 1**). Simulations from the models can then be matched to these alternative hypotheses, and rigorously compared to observed data to identify the variety of relevant processes (ecological, evolutionary, and eco-evolutionary feedback) that are consistent with this observed data. The simulations provide features (i.e. **summary statistics**; see Glossary) that are compared to corresponding features in the observed dataset, and many thousands of simulations should be run under each model to generate potential distributions of these features. Examples of these summary statistics in eco-evolutionary simulation models include spatial alpha diversity, frequencies of different range sizes, normalized lineage-through-time diversification curves [14], or summary statistics across three data axes (species abundances, population genetic variation, and trait values; [12]). We describe examples of such eco-evolutionary hypothesis formulations (**Box 1, Figure 1**), and how to compare observed data to these models (**Box 2**).

Choose a mechanistic model for eco-evolutionary hypothesis testing

The first step in eco-evolutionary hypothesis testing is to identify a mechanistic model to pair with observed data. This potential challenge is not unique to eco-evolutionary dynamics but is also a necessary step to make the field more predictive. Once the research question has been identified and the relevant biological data have been collected, there are a few options for selecting an appropriate model. The choice of model should be based on knowledge of the system (e.g., mechanisms and process that determine ecological dynamics, patterns, or outcomes), the relevant biodiversity (from genes to communities), the spatial and temporal scales relevant to the data of interest, as well as the summary statistics used to compare model simulations to observed data (**Box 1**). It is not always necessary to develop an entirely new model or design an experiment that estimates all parameters in a single theoretical model, as one can choose among existing models or tools to build models. These could be theoretical models, or simulation models that can reproduce the dynamics described in theoretical models. Numerous R software packages exist that allow simulations of commonly used ecological models such as logistic growth and multi-species interaction models (*gauseR*; [15]), consumer functional response models (*FRAIR*; [16]), infectious disease models (*EpiDynamics*; [17]), and island biogeography and metapopulation models (*EcoEvoApps*; [18]), and researchers can use open-source code to extend the model to consider evolution. The EcoEvo Mathematica package (https://github.com/cklausme/EcoEvo) allows simulations and equilibrium condition analyses of species- and trait-based eco-evolutionary models, using differential equations with intra- and inter-specific interactions and either quantitative genetic or adaptive dynamics trait models.

If models with known equations that capture the eco-evolutionary dynamics of interest do not exist or are too difficult to develop de novo, a researcher might instead use one from the emerging class of **general-use eco-evolutionary simulation models** (**Box 1**; **Supplement 1**). These generally use a fixed set of input parameters to model sequence, phenotypic, and/or phylogenetic evolution, where evolving traits impact how individuals grow, reproduce, die, move across a landscape, and interact with their environment and with other individuals, and where individuals inhabit a landscape with user-specified connectivity and movement patterns. These models can simultaneously consider both ecological and evolutionary diversity-generating processes (e.g., speciation, neutral and adaptive sequence and phenotypic evolution, spatial and environmental drivers of population and community structure), and generate known patterns that emerge from these processes (e.g., nucleotide diversity patterns, rank abundance curves, species-genetic diversity correlations). The models differ in some ways, such as the ease of implementing age- or stage-based population structure (*NEMO-AGE*, [19]), the inclusion of molecular evolution, (*NEMO-AGE*; *SLiM 4*, [20]), the built-in models for organismal movement (*RangeShifter*; [21]), and the types of built-in species interactions (**Box 1**).

Generate alternative hypotheses for processes that structure observed data

Research in eco-evolutionary dynamics has moved forward from asking whether evolution impacts ecological processes to identifying the mechanism of this effect [22,23,24]. Linking mechanistic modeling with observational data could allow investigation of these mechanisms even if they are not easily accessible through direct measurements (e.g., in field surveys, or for mechanisms that are typically accessible only through experimental manipulation). Some potential examples of hypotheses that could be studied by comparing data to mechanistic models include the precise form or mechanism of evolution (e.g., evolution of single or multiple traits, uncorrelated or correlated, the existence of trade-offs between evolution of life history traits and competition and their effects on community structure), or identifying feedbacks between ecological and evolutionary dynamics. Testing these types of hypotheses represents a step beyond showing only that ecology and evolution occur simultaneously.

To illustrate how eco-evolutionary hypothesis testing can work, we consider one example of eco-evolutionary dynamics on a micro-evolutionary timescale at a single spatial location - an epidemiological model with virulence evolution [25], where the number of individuals in susceptible (S) and infected (I) host populations are determined by a transmission-virulence tradeoff. Virulence (α) can evolve if the additive genetic variance for this trait $V > 0$, and the rate of change for this trait depends on the susceptible host population size ($d\alpha/dt = f(S)$). This link between the evolving trait and host population dynamics represents an eco-evolutionary feedback as $dS/dt = f(\alpha)$ and $d\alpha/dt = f(S)$ [1]. A null hypothesis for comparing empirical data to this model system could be that phenotypic evolution is not possible (the additive genetic variance for virulence $V = 0$), with an alternative hypothesis where evolution ($V > 0$) and a feedback between evolution and ecology is possible, and an additional alternative hypothesis where eco-evolutionary dynamics are removed from the system by decoupling the dependence of virulence evolution from host population dynamics ($d\alpha/dt \neq f(S)$; **Figure 1A**). Additional simulations can be used to explore the impacts of different trade-off functions (between virulence and transmission, or in the relationship between evolutionary change in virulence and population size of susceptible hosts).

For an example of eco-evolutionary dynamics on macroevolutionary timescales in multi-species communities at a large biogeographical scale, we used *gen3sis* ([14]; **Box 1**) to simulate phenotypic evolution, speciation, and community assembly in a hypothetical community across 40 million years and monitor the impact of various eco-evolutionary processes for emergent patterns of species diversity. In this example, we consider four alternative hypotheses for evolution (**Figure 1B-E**): with single vs. multiple traits, evolving at the same or different speeds, with traits either uncorrelated or correlated. Resulting patterns (e.g., of richness at the local α, regional γ, and among-site β levels, or of phylogenetic diversity) can be compared to observed data and model posterior probabilities can inform researchers of the likelihood that different eco-evolutionary processes have structured their data. A researcher could also test for the presence of eco-evolutionary feedbacks between evolution of dispersal ability and range expansion [26], the effects of uniform versus varying dispersal ability across species, or develop an alternative configuration where traits evolve as a tradeoff between local competitive ability and colonization ability [27,28]. Eco-evolutionary simulation models with explicit consideration of sequence evolution (e.g. *NEMO-AGE*, [17]; *sPEGG*, [24]; SLiM 4, [20]; **Box 1**) allow tests for the role that **genetic architecture** plays for structuring eco-evolutionary dynamics (e.g. species that are similar in niche axes but differ in their degree of standing genetic variation, mating systems, or other properties that impact adaptive capacity, i.e. [29]). Eco-evolutionary simulation models can test the impacts of these processes for emergent community properties such as coexistence, diversity, or resilience to perturbation.

Compare simulated data under each alternative hypothesis to observed data

The next step is to compare observed data to data simulated under null and alternative eco-evolutionary hypotheses and their associated models. A decision-making or classifying algorithm can accept or reject simulations based on their similarity to summary statistics of observed data (**Box 1, 2**). As observed and simulated data are compared based on the match in information-rich summary statistics, there is no a priori reason that absolute time series of sequences, trait, or abundance data are necessary to successfully identify the processes most likely to have generated an observed dataset. Diverse summary statistics across different scales and biological levels may be critical to address the problem of equifinality, where different combinations of processes or model parameters can produce similar patterns of data at a single biological level [11,30]. Before the step of comparing observed to simulated summary statistics, a researcher can combine re-classification of simulated datasets (e.g., leave-one-out **cross validation**) with a feature selection algorithm (e.g. a **boruta** algorithm; [31]) to both identify summary statistics with the most predictive power for a given target dataset and to help researchers consider what data properties are most critical to collect for hypothesis testing. To address data that is at an unknown point along a trajectory towards model equilibrium, the *MESS* model includes a parameter Λ that measures the 'fraction' of equilibrium that a system has obtained (the point at which the starting conditions of the simulation model are no longer detectable in the state of the system), which can be set to obtain simulations away from equilibrium and can be estimated from an observed empirical dataset [12,32].

It is not required to render guesses for or attempt measurements of all model parameters to compare observed and simulated data. To generate the thousands of simulations needed under each alternative hypothesis, researchers can instead sample candidate parameter values from a proposed **prior distribution**. These prior distributions should be based on a researcher's understanding of the model system when possible (e.g., an uninformative uniform distribution between 0-1 for a heritability parameter), or informative based on literature measures (e.g. an increased probability of weak species interactions, [33]; see [34,35]). After candidate simulations are run and summary statistics (or features) are extracted from each simulation run, a decision algorithm is used to accept or reject these simulated data features based on some distance between observed and simulated values. We describe how to use **Approximate Bayesian Computation** (ABC, [36]) for such an algorithmic approach in **Box 2**, but additional options are available (e.g., **Markov chain Monte Carlo** (MCMC), [37,38, 39]; **machine learning** (ML), [40]; see [41,42] for an overview of machine learning applied to model fitting in ecology and evolution and **Supplement 1** for an ML alternative to the analysis in **Box 2**). ABC is a Bayesian statistical computational approach to estimating **posterior distributions** of model parameters via random draws from prior distributions and comparing to observed values via a criterion for acceptance or rejection [36,43]. Researchers can generate their own models, alternative hypotheses, and associated simulated summary statistics, and use an existing *abc* R package [44] for model comparison. One application of ABC to eco-evolutionary hypothesis testing is the study of Baselga et al. [45] to estimate the relative importance of dispersal limitation and niche width in contemporary surveys of genetic and community structure (see also [46]). Machine learning could also be used to train a classifier algorithm on simulated datasets from known alternative model conditions, then classify the observed data and generate predicted model class probabilities. This approach was used by Overcast et al. [12] to classify whether datasets from contemporary surveys of diverse organismal groups were most consistent with neutral, environmental filtering, or competition eco-evolutionary models.

Model validation and generating future predictions

Once the decision algorithm (e.g. ABC) is used to produce posterior probabilities for each candidate model, researchers should consider posterior predictive checks to evaluate model fit [47,48], evaluate the importance of summary statistics for predictive ability [40,49,50], and also consider whether to use the model with the highest posterior probability support or to average over the models with weights based on their relative posterior support [51,52]. After these steps, researchers can gain valuable knowledge by considering the posterior distributions of model parameters that result from the simulations retained by the acceptance-rejection algorithm (**Box 2**). These posterior estimates of model parameters, some of which might have associated observations for validation and others of which are impossible or resource-intensive to estimate, are quite valuable for process inference, and can also serve as guidance for future data collection to confirm these estimates. They can also be used to make predictions for future dynamics of the system, with estimates of uncertainty, and with inclusion of alternative scenarios for future systemic change [53,54].

For studies of eco-evolutionary dynamics, predictive simulations can be used to determine the importance of eco-evolutionary dynamics and feedbacks for critical emergent properties such as stability, extinction risk, or biodiversity maintenance. For example in the host-disease model (**Figure 1A**), the impacts of including versus excluding the eco-evolutionary feedback loop for population stability could be evaluated by running simulations with parameters drawn from posterior distributions under each alternative model, and then generating a posterior distribution for an emergent property of the simulations such the **coefficient of variation** (CV, a metric often used to quantify stability; [55]). In one example, the study of Luiselli et al. [56] determined the relative impact of the combination of speciation rate and mechanism of competition for population genetic and community structure in a variety of taxonomic systems. They used the MESS simulation model to decompose the effects of speciation rates and competition type for emergent summary statistics of overall species richness, and community, genetic, and functional trait diversity. Their study found strong differences in the change in genetic diversity over time depending on whether speciation was included, whether species interacted neutrally, and on the mechanism of competition considered. Their results confirm that many of the properties we observe at the level of populations and communities are generated by eco-evolutionary processes. One recent study has established a mechanistic basis for the impact of eco-evolutionary dynamics for a critical, emergent system property of food web collapse. Barbour et al. [23] identified that the presence of a particular allele (AOP2$^+$ or AOP2$^-$) in host *Arabidopsis thaliana* plants could result in a 29% difference in extinction rates for associated aphids and parasitoids in an experimental food web.

Concluding remarks and future perspectives

It has previously been said that confirmation of eco-evolutionary predictions requires monitoring genetic properties of populations and subsequent community interactions over time intervals in which selection regimes are likely to have caused changes in ecologically relevant traits [8,57,58]. Studies of eco-evolutionary interactions and feedbacks are thus often limited to tractable taxonomic systems [59,60], with time-intensive measurements of phenotypes in common environments. The complex eco-evolutionary processes that provide structure in data are often collapsed into additive categories with effect sizes estimated using ANOVA-like variance partitioning without consideration for generalized probability distributions, fixed and random effects, or non-linear and interacting combinations of drivers of system responses (e.g. [8,9]). However, similar to the way that processes of evolution and the dynamics of natural selection leave distinct signatures on contemporary populations that can be analyzed to infer those historical structuring processes (e.g. [61,62,63]), past eco-evolutionary dynamics likely leave distinct signatures on contemporary populations and communities. Researchers must still scale the difficult challenge of identifying which features, emergent properties, and summary statistics from observed data capture the signatures of distinct eco-evolutionary processes (e.g., [3]). However, the path of using statistical or **process-based mechanistic models** to compare the likelihood of observed data to alternative models is used in other fields of research where historical processes cannot be observed (e.g. coalescent processes and demographic history, phylogenetic reconstruction and patterns, population genetic structure; [20,64,65,66]). The remaining challenges to uncover eco-evolutionary signatures on extant biodiversity are (i) generating data in contemporary populations and communities that captures the features necessary to infer eco-evolutionary structuring processes, (ii) accessing user-friendly versions of statistical models that can disentangle the numerous processes (drift, selection, gene flow, community assembly and metacommunity processes, multivariate environmental forcing; [67,68]) that combine to structure biodiversity from the gene to ecosystem level, and (iii) converging on mechanistic hypotheses and predictions for the impacts of eco-evolutionary dynamics in biological systems.

There is good reason to be optimistic on each front. First of all, simulations from mechanistic models under alternative eco-evolutionary hypotheses can be combined with exploration of diverse summary statistics and reclassification of subsets of simulated data to identify potential signatures of eco-evolutionary dynamics in natural systems (**Box 1**, using Step 2 and 3 to inform Step 1). As biodiversity science is entering an unprecedented era of technology-assisted, high-throughput data collection [69,70,71] and of open, reproducible data sharing [72,73], it is increasingly possible to test whether these signatures to identify eco-evolutionary dynamics are realized in diverse systems. Some recent examples of 'high-throughput' eco-evolutionary data collection exist, including fine time series of multi-species population size and trait values via fluid imaging technology [74,75], real-time observation of adaptive tracking to environmental change (e.g. *Drosophila melanogaster*, [24]; *Arabidopsis thaliana*, [76]; other non-model systems, [77,78,79]). Moving eco-evolutionary dynamics into the genomic era is a promising current direction of study [22,80]. The difficulty of detecting eco-evolutionary processes in complex systems could be alleviated by genomic data monitoring [81,82], especially in systems where candidate genes can be used for monitoring selection in real time and eDNA and high-throughput sequencing can be used for monitoring evolutionary dynamics and ranges of organisms across large spatial scales [83,84,85].

Second, we have discussed theoretical and simulation models that are moving towards modeling the full suite of interacting processes that structure biodiversity at multiple scales (spatial, temporal, diversity across levels of organization; **Box 1**; additional detail for models in **Supplement 1**), and some additional reviews addressed models that can be used to predict future biodiversity responses to environmental change [86,87]. There is still a need for statistical tools to detect effects and their magnitudes in experimental and field observational data, and the difficulty of distinguishing among the many signals that structure genes, traits, populations, and communities is compounded by the complexity of eco-evolutionary dynamics. Some new statistical models that address this need include an integrated reaction norm model linking genetic, phenotypic, and demographic processes [88] and a species distribution model with local adaptation and phenotypic plasticity (ΔSDMs; [89]). Modern statistical models for population and community ecology now consider critical processes such as observer, measurement, and process error, error propagation, or the existence of present but undetected species or life stages (e.g. [90,91,92,93,94,95]), and these features are likely to be important for eco-evolutionary models as well. For linking microevolution and models of eco-evolutionary dynamics, tools such as *sPEGG* [24], SLiM 4 [20], and the EcoEvo Mathematica package are currently available for generating features to compare with observed data. We also share the code used for all the simulations here in an open-source R package *ecoevoR* (https://github.com/jhpantel/ecoevoR).

There is an urgent need to move beyond establishing that evolution *can* be important for ecological processes or that eco-evolutionary feedback loops *might* exist in some systems. Are there particular environmental and spatial conditions [96] or features of community composition [97] or food web network structure [98] that make eco-evolutionary dynamics more or less likely to occur? The lack of information here highlights the reality that a background or null level of interacting eco-evolutionary processes is currently unknown (see **Outstanding Questions**). Published examples tend to focus on clear feedbacks between adaptive evolution and ecological dynamics in organisms that may have disproportionate adaptive capacity (e.g. *Daphnia*, [5,99,100]; *Arabidopsis*, [23]). However, it is currently unknown whether eco-evolutionary dynamics are driven by large effect alleles vs. allelic changes in numerous locations with cumulative small effects. Additionally, the role of adaptive evolution has been considered in many studies of eco-evolutionary dynamics, but fewer have established the role that sexual selection, genetic drift, or maladaptation might play (but see discussion of maladaptation in evolutionary rescue literature e.g. [101,102]), as these processes also structure genetic architecture, population demography, and species interactions [22,103,104,105]. Given that numerous evolutionary and ecological processes are continuously occurring, operating at a variety of time scales, answers to these research questions are needed for statistical models to be developed that implement accurate distributions for emergent properties that are often the targets of studies of the impacts of environmental change. Much of this review focuses on the methodology needed to detect signatures of eco-evolutionary dynamics in experimental and observed biodiversity data. We anticipate a research era in which these tools can be used to address urgent, compelling questions such as what the role of eco-evolutionary dynamics is for large-scale properties such as diversity, stability, and ecosystem functioning.

Acknowledgements:

J.H. Pantel acknowledges support from Professorinnenprogramm funded by the German Federal Ministry for Research and Education and the federal states.

References:

1. Lion, S. (2018) Theoretical approaches in evolutionary ecology: environmental feedback as a unifying perspective. *Am. Nat.* 191, 21–44
2. Bassar, R.D. *et al.* (2021) Towards a more precise - and accurate - view of eco-evolution. *Ecol. Lett.* 24, 623–625
3. Yoshida, T. *et al.* (2003) Rapid evolution drives ecological dynamics in a predator–prey system. *Nature* 424, 303–306
4. Bassar, R.D. *et al.* (2012) Direct and indirect ecosystem effects of evolutionary adaptation in the Trinidadian guppy (Poecilia reticulata). *Am. Nat.* 180, 167–185
5. Pantel, J.H. *et al.* (2015) Rapid local adaptation mediates zooplankton community assembly in experimental mesocosms. *Ecol. Lett.* 18, 992–1000
6. Frickel, J. *et al.* (2017) Rapid evolution of hosts begets species diversity at the cost of intraspecific diversity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11193–11198
7. Hairston, N.G., Jr *et al.* (2005) Rapid evolution and the convergence of ecological and evolutionary time. *Ecol. Lett.* 8, 1114–1127
8. Ellner, S.P. *et al.* (2011) Does rapid evolution matter? Measuring the rate of contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.* 14, 603–614
9. Govaert, L. *et al.* (2016) Eco-evolutionary partitioning metrics: assessing the importance of ecological and evolutionary contributions to population and community change. *Ecol. Lett.* 19, 839–853
10. Hattich, G.S.I. *et al.* (2022) Experimentally decomposing phytoplankton community change into ecological and evolutionary contributions. *Funct. Ecol.* 36, 120–132
11. Cabral, J.S. *et al.* (2019) Interactions between ecological, evolutionary and environmental processes unveil complex dynamics of insular plant diversity. *J. Biogeogr.* 46, ii–iv
12. Overcast, I. *et al.* (2021) A unified model of species abundance, genetic diversity, and functional diversity reveals the mechanisms structuring ecological communities. *Mol. Ecol. Resour.* 21, 2782–2800
13. Hagen, O. (2022) Coupling eco-evolutionary mechanisms with deep-time environmental dynamics to understand biodiversity patterns. *Ecography* DOI: 10.1111/ecog.06132
14. Hagen, O. *et al.* (2021) gen3sis: A general engine for eco-evolutionary simulations of the processes that shape Earth's biodiversity. *PLoS Biol.* 19, e3001340
15. Clark, A. *et al.* (2021) gauseR: Lotka-Volterra models for Gause's 'Struggle for existence'. R package version 1.1, https://CRAN.R-project.org/package=gauseR
16. Pritchard, D.W. *et al.* (2017) FRAIR: an R package for fitting and comparing consumer functional responses. *Methods Ecol. Evol.* 8, 1528–1534

17. Santos Baquero and Silveira Marques. (2020) Epidynamics: dynamic models in epidemiology. R package version 0.3.1, https://CRAN.R-project.org/package=EpiDynamics
18. McGuire, R.M. *et al.* (2022) EcoEvoApps: Interactive apps for theoretical models in ecology and evolutionary biology. *Ecol. Evol.* 12, e9556
19. Cotto, O. *et al.* (2020) Nemo-age: Spatially explicit simulations of eco-evolutionary dynamics in stage-structured populations under changing environments. *Methods Ecol. Evol.* 11, 1227–1236
20. Haller, B.C. and Messer, P.W. SLiM 4: Multispecies eco-evolutionary modeling. *Am. Nat.* 0, null
21. Bocedi, G. *et al.* (2021) RangeShifter 2.0: an extended and enhanced platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Ecography* 44, 1453–1462
22. Retel, C. *et al.* (2019) The feedback between selection and demography shapes genomic diversity during coevolution. *Sci Adv* 5, eaax0530
23. Barbour, M.A. *et al.* (2022) A keystone gene underlies the persistence of an experimental food web. *Science* 376, 70–73
24. Rudman, S.M. *et al.* (2022) Direct observation of adaptive tracking on ecological time scales in Drosophila. *Science* 375, eabj7484
25. Luo, S. and Koelle, K. (2013) Navigating the devious course of evolution: the importance of mechanistic models for identifying eco-evolutionary dynamics in nature. *Am. Nat.* 181 Suppl 1, S58-75
26. Fronhofer, E.A. and Altermatt, F. (2015) Eco-evolutionary feedbacks during experimental range expansions. *Nat. Commun.* 6, 6844
27. Pillai, P. and Guichard, F. (2012) Competition-colonization trade-offs, competitive uncertainty, and the evolutionary assembly of species. *PLoS One* 7, e33566
28. Pettersen, A.K. *et al.* (2020) Metabolic rate, context-dependent selection, and the competition-colonization trade-off. *Evol. Lett.* 4, 333–344
29. De Meester, L. *et al.* (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecol.* 23, 121–135
30. Kinnison, M.T. *et al.* (2015) Cryptic eco-evolutionary dynamics. *Ann. N. Y. Acad. Sci.* 1360, 120–144
31. Kursa, M.B. and Rudnicki, W.R. (2010) Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13
32. Overcast, I. *et al.* (2019) An integrated model of population genetics and community ecology. *J. Biogeogr.* 46, 816–829
33. Wootton, J.T. and Emmerson, M. (2005) Measurement of Interaction Strength in Nature. *Annu. Rev. Ecol. Evol. Syst.* 36, 419–444
34. Ellison, A.M. (2004) Bayesian inference in ecology. *Ecol. Lett.* 7, 509–520
35. Hooten, M.B. and Hobbs, N.T. (2015) A guide to Bayesian model selection for ecologists. *Ecol. Monogr.* 85, 3–28
36. Beaumont, M.A. (2010) Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics* 41, 379–406
37. Marjoram, P. *et al.* (2003) Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. U. S. A.* 100, 15324–15328
38. Hartig, F. *et al.* (2011) Statistical inference for stochastic simulation models – theory and application. *Ecol. Lett.* 14, 816–827
39. Newman, K. *et al.* (2022) State-space models for ecological time-series data: Practical model-fitting. *Methods Ecol. Evol.* 14, 26– 42
40. Pichler, M. and Hartig, F. (2022) Machine Learning and Deep Learning -- A review for Ecologists. *arXiv preprint* arXiv:2204.05023
41. Lucas, T.C.D. (2020) A translucent box: interpretable machine learning in ecology. *Ecol. Monogr.* 90, e01422

42. Borowiec, M.L. *et al.* (2022) Deep learning as a tool for ecology and evolution. *Methods Ecol. Evol.* 13, 1640–1660
43. Sunnåker, M. *et al.* (2013) Approximate Bayesian computation. *PLoS Comput. Biol.* 9, e1002803
44. Csilléry, K. *et al.* (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol. Evol.* 3, 475–479
45. Baselga, A. *et al.* (2022) Joint analysis of species and genetic variation to quantify the role of dispersal and environmental constraints in community turnover. *Ecography* 2022, e05808
46. Pontarp, M. *et al.* (2019) Inferring community assembly processes from macroscopic patterns using dynamic eco-evolutionary models and Approximate Bayesian Computation (ABC). *Methods Ecol. Evol.* 10, 450–460
47. Gelman, A. *et al.* (1996) Posterior predictive assessment of model fitness via realized discrepancies. *Stat. Sin.* 6, 733–760
48. Conn, P.B. *et al.* (2018) A guide to Bayesian model checking for ecologists. *Ecol. Monogr.* 88, 526–542
49. Jabot, F. and Chave, J. (2009) Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecol. Lett.* 12, 239–248
50. Blum, M.G.B. *et al.* (2013) A Comparative Review of Dimension Reduction Methods in Approximate Bayesian Computation. *SSO Schweiz. Monatsschr. Zahnheilkd.* 28, 189–208
51. Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution. *Trends Ecol. Evol.* 19, 101–108
52. Dormann, C.F. *et al.* (2018) Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs* 88, 485–504
53. Eyring, V. *et al.* (2019) Taking climate model evaluation to the next level. *Nat. Clim. Chang.* 9, 102–110
54. McIntire, E.J.B. *et al.* (2022) PERFICT: A Re-imagined foundation for predictive ecology. *Ecol. Lett.* 25, 1345–1351
55. Tilman, D. (1999) The ecological consequences of changes in biodiversity: A search for general principles101. *Ecology* 80, 1455–1474
56. Luiselli, J. *et al.* (2021) Detecting the ecological footprint of selection. *bioRxiv*, 2021.05.11.442553
57. G. F. Fussmann *et al.* (2007) Eco-Evolutionary Dynamics of Communities and Ecosystems. *Funct. Ecol.* 21, 465–477
58. Schoener, T.W. (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* 331, 426–429
59. De Meester, L. and Pantel, J. (2014) Eco-evolutionary dynamics in freshwater systems. *Journal of Limnology* 73 1 Suppl, 193-200
60. De Meester, L. *et al.* (2019) Analysing eco-evolutionary dynamics—The challenging complexity of the real world. *Funct. Ecol.* 33, 43–59
61. Sokal and Riska (1981) Geographic variation in Pemphigus populitransversus (Insecta: Aphididae). *Biol. J. Linn. Soc. Lond.* 15, 201–233
62. Hutchison, D.W. and Templeton, A.R. (1999) Correlation of pairwise genetic and geographic distance measures: inferring the relative influences of gene flow and drift on the distribution of genetic variability. *Evolution* 53, 1898–1914
63. Haller, B.C. and Messer, P.W. (2019) SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol. Biol. Evol.* 36, 632–637
64. Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214

65. Liu, L. (2008) BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* 24, 2542–2543
66. McPeek, M.A. (2008) The ecological dynamics of clade diversification and community assembly. *Am. Nat.* 172, E270-84
67. Vellend, M. (2010) Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–206
68. Vellend, M. (2016) The Theory of Ecological Communities (MPB-57). Princeton University Press.
69. Michener, W.K. and Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol. Evol.* 27, 85–93
70. Keitt, T.H. and Abelson, E.S. (2021) Ecology in the age of automation. *Science* 373, 858–859
71. Besson, M. *et al.* (2022) Towards the fully automated monitoring of ecological communities. *Ecol. Lett.* 25, 2753–2775
72. Lowndes, J.S.S. *et al.* (2017) Our path to better science in less time using open data science tools. *Nat Ecol Evol* 1, 160
73. Powers, S.M. and Hampton, S.E. (2019) Open science, reproducibility, and transparency in ecology. *Ecol. Appl.* 29, e01822
74. Wieczynski, D.J. *et al.* (2021) Linking species traits and demography to explain complex temperature responses across levels of organization. *Proc. Natl. Acad. Sci. U. S. A.* 118
75. Gibert, J.P. *et al.* (2022) Rapid eco-phenotypic feedbacks and the temperature response of biomass dynamics. *bioRxiv*, 2022.06.17.496633
76. Fulgione, A. *et al.* (2022) Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. *Nat. Commun.* 13, 1461
77. Gienapp, P. *et al.* (2017) Genomic Quantitative Genetics to Study Evolution in the Wild. *Trends Ecol. Evol.* 32, 897–908
78. Dudaniec, R.Y. *et al.* (2018) Signatures of local adaptation along environmental gradients in a range-expanding damselfly (Ischnura elegans). *Mol. Ecol.* 27, 2576–2593
79. Pfenninger, M. and Foucault, Q. (2022) Population genomic time series data of a natural population suggests adaptive tracking of fluctuating environmental changes. *Integr. Comp. Biol.* 62, 1812-1826
80. Rudman, S.M. *et al.* (2018) What genomic data can reveal about eco-evolutionary dynamics. *Nat. Ecol. Evol.* 2, 9–15
81. Davies, N. *et al.* (2014) The founding charter of the Genomic Observatories Network. *Gigascience* 3, 2
82. Emerson, B. *et al.* (2022) Collective and harmonised high throughput barcoding of insular arthropod biodiversity: toward a Genomic Observatories Network for islands. *Authorea* DOI: 10.22541/au.164840312.23663001
83. Shafer, A.B.A. *et al.* (2016) Forecasting Ecological Genomics: High-Tech Animal Instrumentation Meets High-Throughput Sequencing. *PLoS Biol.* 14, e1002350
84. Deiner, K. *et al.* (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895
85. Ellegaard, M. *et al.* (2020) Dead or alive: sediment DNA archives as tools for tracking aquatic evolution and adaptation. *Commun. Biol.* 3, 169
86. Waldvogel, A. *et al.* (2020) Evolutionary genomics can improve prediction of species' responses to climate change. *Evol. Lett.* 4, 4–18
87. Urban, M.C. *et al.* (2021) Coding for Life: Designing a Platform for Projecting and Protecting Global Biodiversity. *Bioscience* 72, 91–104
88. Lasky, J.R. *et al.* (2020) What processes must we understand to forecast regional-scale population dynamics? *Proc. Biol. Sci.* 287, 20202219
89. Benito Garzón, M. *et al.* (2019) ΔTraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. *New Phytol.* 222, 1757–1765

90. Ovaskainen, O. *et al.* (2017) How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecol. Lett.* 20, 561–576
91. Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the R-package Hmsc. *Methods Ecol. Evol.* 11, 442-447
92. Pollock, L.J. *et al.* (2014) Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5, 397–406
93. Pichler, M. and Hartig, F. (2021) A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods Ecol. Evol.* 12, 2159–2173
94. Devarajan, K. *et al.* (2020) Multi-species occupancy models: review, roadmap, and recommendations. *Ecography* 43, 1612–1624
95. Doser, J.W. *et al.* (2022) spOccupancy: An R package for single-species, multi-species, and integrated spatial occupancy models. *Methods Ecol. Evol.* 13, 1670-1678
96. Boussange, V. and Pellissier, L. (2022) Eco-evolutionary model on spatial graphs reveals how habitat structure affects phenotypic differentiation. *Commun. Biol.* 5, 668
97. Govaert, L. *et al.* (2021) Measuring the contribution of evolution to community trait structure in freshwater zooplankton. *Oikos* 130, 1773–1787
98. Pantel, J.H. *et al.* (2017) 14 questions for invasion in ecological networks. *Adv. Ecol. Res.* 56, 293-340
99. Colbourne, J.K. *et al.* (2011) The ecoresponsive genome of Daphnia pulex. *Science* 331, 555–561
100. Moody, E.K. *et al.* (2022) Eutrophication-driven eco-evolutionary dynamics indicated by differences in stoichiometric traits among populations of Daphnia pulicaria. *Freshw. Biol.* 67, 353–364
101. Gomulkiewicz, R. and Holt, R.D. (1995) When does evolution by natural selection prevent extinction? *Evolution* 49, 201–207
102. Hermann, R.J. and Becks, L. (2022) Change in prey genotype frequency rescues predator from extinction. *R. Soc. Open Sci.* 9, 220211
103. Farkas, T.E. *et al.* (2013) Evolution of camouflage drives rapid ecological change in an insect community. *Curr. Biol.* 23, 1835–1843
104. Rodríguez-Verdugo, A. *et al.* (2017) The genomic basis of eco-evolutionary dynamics. *Mol. Ecol.* 26, 1456–1464
105. Svensson, E.I. (2019) Eco-evolutionary dynamics of sexual selection and sexual conflict. *Funct. Ecol.* 33, 60–72
106. Connolly, S.R. *et al.* (2017) Process, Mechanism, and Modeling in Macroecology. *Trends Ecol. Evol.* 32, 835–844
107. Gallien, L. *et al.* (2017) The effects of intransitive competition on coexistence. *Ecol. Lett.* 20, 791–800

Glossary:

Approximate Bayesian Computation (ABC) - Approximate Bayesian Computation is a class of computational methods, based on a Bayesian statistical framework, to simulate posterior distributions of model parameters via random draws from those distributions and comparison to some true values of observations via some criterion for acceptance or rejection

boruta - A feature selection algorithm that uses a random forest classification algorithm to identify predictive features that are more informative than randomly generated features

coefficient of variation (CV) - The ratio of standard deviation to mean

cross validation - A model validation technique to assess how a statistical model or analysis will generalize to an independent data set, where a part of the data is withheld from training the model to assess its ability to classify the withheld data

genetic architecture –the underlying genetic basis that builds and controls a phenotypic trait of an individual or population and its variational properties, which can include the number of genes involved, the distribution of their effects, and their interactions (dominance, epistasis, pleiotropy)

general-use eco-evolutionary simulation model - A class of models to simulate eco-evolutionary processes, that include intrinsic population dynamics, species interactions with one another and with their environment, and genetic (molecular or quantitative trait) evolution. Models should include processes known to impact different levels of biodiversity (genes, traits, populations, and/or communities) and should accurately reproduce known patterns of emergent aggregate properties such as gene, species, or phylogenetic diversity

machine learning - A family of computational, algorithmic, or statistical approaches to build a model based on sample data (training data) to make predictions, decisions, or classifications without explicit programming of a data-generating model

Markov chain Monte Carlo (MCMC) - A class of algorithms to obtain a sample of a probability distribution, beginning from an arbitrarily chosen set of starting points and proceeding by moving at random from the starting point towards representative samples of the distribution based on some distance or rejection criterion

posterior distribution - Refers to a posterior probability distribution, or the updated probability distribution for a model parameter after taking into account information (new information in the form of data)

prior distribution - Refers to a prior probability distribution, or the expression of an uncertain distribution for the probability of different values of a model parameter, which captures prior expectations for the shape or structure of the parameter

process-based mechanistic model – A mechanistic model is a characterization of the state of a system as explicit functions of component parts and their associated actions and interactions, and a process-based model is a model that characterizes changes in a system's state as explicit functions of the events that drive those state changes [106]

summary statistic – A descriptive statistic calculated from data that gives compressed, summarized information, used to reduce data dimensionality (though information in the original data can be lost); may be refered to as 'features' in machine learning applications

Outstanding questions:

1. Does the occurrence of eco-evolutionary feedbacks depend on a few alleles with large effects or on many alleles with small effects in numerous directly and indirectly interacting species?

2. Under what conditions and how often do the time scales of ecological and evolutionary dynamics overlap, and does evolution lead to ecological dynamics that cannot be predicted by ecological processes alone and vice versa?

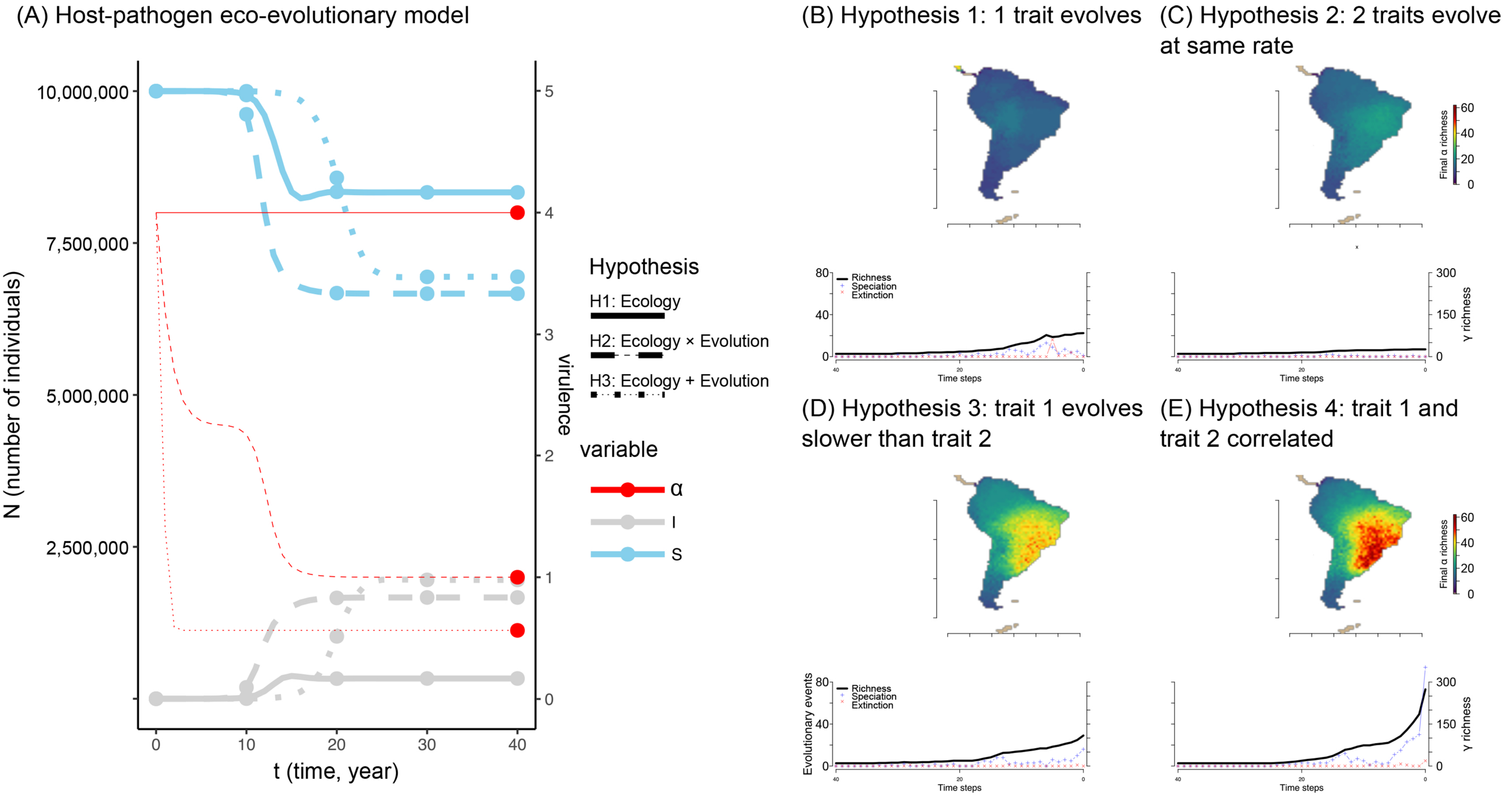
3. How do sexual selection, genetic drift, or mismatch structure genetic architecture, population demography, species interactions, and eco-evolutionary dynamics?

4. What is the role of eco-evolutionary dynamics for large-scale properties such as diversity, stability, and ecosystem functioning?

Tables & Figures:

Figure 1. Eco-evolutionary hypotheses for structure of observed disease and biodiversity dynamics. (a) Simulations of an eco-evolutionary epidemiological model with virulence evolution [25] are shown with many of the same model parameters (e.g. S_0 = X, $\alpha_0 = 4$, $\gamma = 2$; **Supplement 1** for additional parameter values), but under three alternative hypotheses that can be compared to observed data. These comparisons of observed to model-generated data under alternative hypotheses can help identify whether trait evolution and/or eco-evolutionary feedbacks structured the observed data. Population sizes of susceptible (S, light blue) and infected (I, light grey) individuals (left *y*-axis), and pathogen virulence (α, red, right *y*-axis), are given for: Hypothesis 1 (H1, solid line) in an ecological-only model where pathogen traits (α) cannot evolve, Hypothesis 2 (H2, dashed line) in an eco-evolutionary dynamics model where pathogen virulence can evolve and impacts host population size, and Hypothesis 3 (H3, dotted line) in an ecological and evolutionary model where virulence evolves independently from host population size. Points are shown for a coarser time series that could potentially be used for comparing observed to model-simulated data, including the final value for pathogen virulence (red points), as fine time series with records of all population and trait dynamics are difficult to obtain (see **Box 2** for eco-evolutionary hypothesis testing with such data). (b-e) Simulations of macroevolution, speciation, and community assembly using the gen3sis eco-evolutionary simulation engine [14], for a hypothetical community of 10 ancestral species, randomly placed as a single initial population on the map of South America (an 81x71 grid with cell-specific temperature values, these values impact population growth via the distance between a local population's trait value and the local optimum, and also impact local carrying capacity). Movement is via a cost function based on the distance between sites and a species-specific dispersal value, and evolution and speciation occur

across the dynamic landscape over 40 million years (**Supplement 1** for additional parameter values). We run simulations under 4 alternative eco-evolutionary hypotheses: (b) Hypothesis 1 (H1) with evolution in a single temperature-related trait that determines population growth, (c) Hypothesis 2 (H2) with evolution proceeding at the same rate in the temperature trait and in an additional dispersal-related trait, (d) Hypothesis 3 (H3) with evolution in these two traits that proceeds at different rates, and (e) Hypothesis 4 (H4) with evolution in the two traits at differing rates, where the two traits are correlated. For all four hypotheses, the final alpha diversity (local species richness) and time series of some summary statistics (number of speciation and extinction events, total regional richness) are shown. These summary statistics create features that could be used for comparison to observed data to distinguish among the alternative hypotheses that may have structured observed data (**Box 2**).


(A) Host-pathogen eco-evolutionary model
10,000,000
7,500,000
5,000,000
2,500,000
N (number of individuals)
0
10
20
30
40
t (time, year)
5
4
3
2
1
0
virulence
Hypothesis
H1: Ecology
H2: Ecology × Evolution
H3: Ecology + Evolution
variable
α
I
S
(B) Hypothesis 1: 1 trait evolves
(C) Hypothesis 2: 2 traits evolve at same rate
(D) Hypothesis 3: trait 1 evolves slower than trait 2
(E) Hypothesis 4: trait 1 and trait 2 correlated
Final α richness
60
40
20
0
Richness
Speciation
Extinction
80
40
0
Evolutionary events
Time steps
40
20
0
300
150
0
γ richness

Box 1. Workflow for model-based hypothesis testing in eco-evolutionary dynamics

1. Identify research question and data

a) Establish eco-evolutionary research questions, identify mechanisms thought to underly expected dynamics and patterns, and collect data (observational, experimental, field or lab-based, according to research needs)

b) Identify relevant summary statistics that are emergent properties of the data; e.g., nucleotide diversity, within- and among-population genetic variation, center and width of phenotypic distributions, phylogenetic structure, total population size, rank abundance curves, normalized lineage-through-time, diversification curves [14], alpha (local), beta (among-site turnover), and gamma (regional) diversity, frequencies of different range sizes, or summary statistics across three data axes (species abundances, population genetic variation, and trait values; [12]). Choosing the most informative summary statistics emerges from the research question and the data available (e.g., total abundance data at one time point will unlikely produce informative model comparisons). Ideally, the summary statistics are determined before data collection via steps 2-4, as simulations and the algorithms used for hypothesis testing can be used to inform what type of data might capture the signature of particular eco-evolutionary dynamics. Summary statistics can be derived from time series or individual points in time, and one should consider diverse summary statistics across different scales and biological levels to address the problem of equifinality [11].

2. Choose a mechanistic model for eco-evolutionary hypothesis testing

a) Identify relevant biodiversity (from genes to communities), spatial, and temporal scales, single or multiple species in a community, intra- and inter-specific interactions (e.g., consumer-resource, infectious disease dynamics, multi-species competition)

b) Find theoretical model or use general-use eco-evolutionary simulation model (**Box 1 Figure 1**) that best matches Step 1 and Step 2a and that can implement evolution (e.g., via molecular genetics, quantitative genetics, or adaptive dynamics model of traits that influence ecological model parameters), ecology (e.g., via a density dependent process or consideration of species interactions), and coupled eco-evolutionary processes

3. Generate alternative hypotheses for processes that structure observed data

a) Formulate a core of null and alternative hypotheses expressed in the form of alternative competing models differing in ecological, evolutionary, and coupled eco-evolutionary processes. Model alternatives may include varying forms of ecological (e.g., neutral or niche-based species interactions), evolutionary (e.g., correlated traits, neutral vs. adaptive evolution), or eco-evolutionary feedback (e.g., density-dependent trait evolution, dispersal-driven maladaptation impacting local fitness) relationships (**Figure 1**)

b) Run simulations under each model with sampling of candidate parameter values from a proposed prior distribution based on the biology of the system (e.g., a uniform distribution between 0-1 for a heritability parameter), literature search, or system knowledge

4. Compare simulated data under each alternative hypothesis to observed data

a) Identify algorithm to compare simulated and observed data

- Least squares
- Markov Chain Monte Carlo (MCMC)
- Machine learning (ML)
- Aproximate Bayesian Computation (ABC)

b) Identify summary statistics in simulation data that are compared to observed data

c) Evaluate summary statistics in simulated data and determine whether they are sufficient to distinguish among alternative hypotheses and their associated distinct eco-evolutionary mechanisms (i.e. test classification ability using training and test data, revisit summary statistics to maximize posterior probabilities from data simulated under known conditions) – note potential summary statistics that are emergent signatures of underlying eco-evolutionary dynamics (e.g. shifts in phase of predator-prey cycles [3])

d) Compare summary statistics in simulated data and observed data

e) Exclude simulations based on broad criteria (e.g., do not consider simulation runs where simulated population becomes extinct but observed population persists, where simulated species do not coexist but observed do, or otherwise where simulation does not resemble main features of observed data)

f) Compare simulated summary statistics to observed, reject simulations that are above a threshold level of similarity to observed data, and generate posterior probabilities for data under each alternative candidate model

5. Model validation and generating future predictions

a) Consider whether to use the model with the highest posterior probability support or to average over the models with weights based on their relative posterior support [51,52]

b) Use posterior predictive checks to evaluate model fit [47,48]

c) Use predictive simulations to determine the importance of eco-evolutionary dynamics and feedbacks for critical emergent properties (e.g., stability, extinction risk, or biodiversity maintenance)

Box 1, Figure 1. Summary of simulation models of eco-evolutionary dynamics and their relevant scales, which can be used to simulate the processes that structure observed biodiversity data. Models are given along with the spatial, temporal, and biodiversity scales considered by the model. We indicate if a level of biodiversity is modeled with filled symbols, and open symbols indicate some special considerations must be given to apply the model to that scale. We use +, - , and 0 to indicate positive, negative, or neutral interactions between species considered in a model (and a filled circle for species interactions means that all types are considered).

1) The MESS model considers multiple sites, but is spatially implicit, with a regional metacommunity immigration rate *m* (probability that an individual replacing a dead individual at a site is a migrant from the metacommunity).

2) Sequence evolution is not considered during forward-time simulations of community assembly, but parameters needed to simulate backward-time coalescent models for genes are estimated, and these coalescent simulations produce data consistent with user-specified sample size (number of individuals per species) and length of sequence.

3) Evolution for a single trait is considered, and no intraspecific variation is modeled.

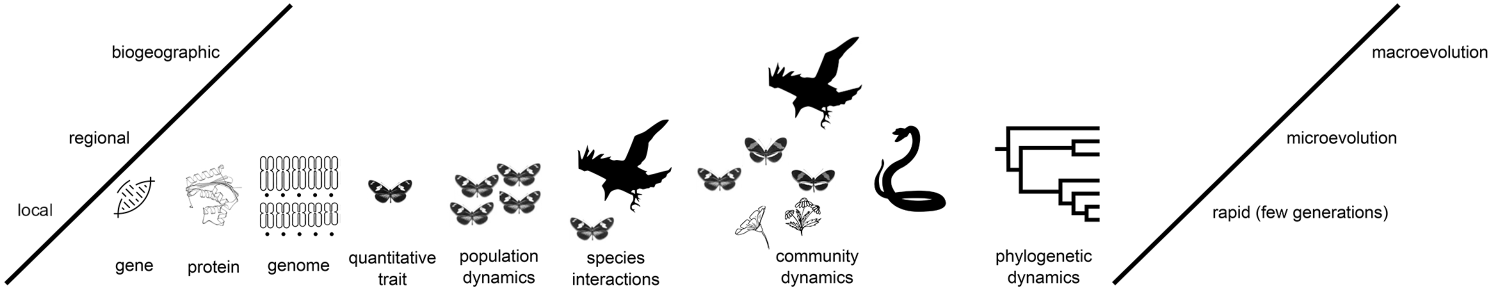
4) In the gen3sis model, the current implementation examples include competition. The species ecology is customizable and can be programmed to consider additional species interactions. However, divergence, speciation, and examples of competition are currently dependent on a single shared trait, which is traditionally used for competition models.

5) In the BioGEEM model, space is modeled to resemble an island, with 1 km^2-sized grids, with each cell assigned to an island side, elevation, and associated temperature.

6) Evolving traits determine a population’s suitability and growth on a grid type, dispersal properties, and stage-specific body masses. The model considers plant evolution and life stages and life form types unique to plants.

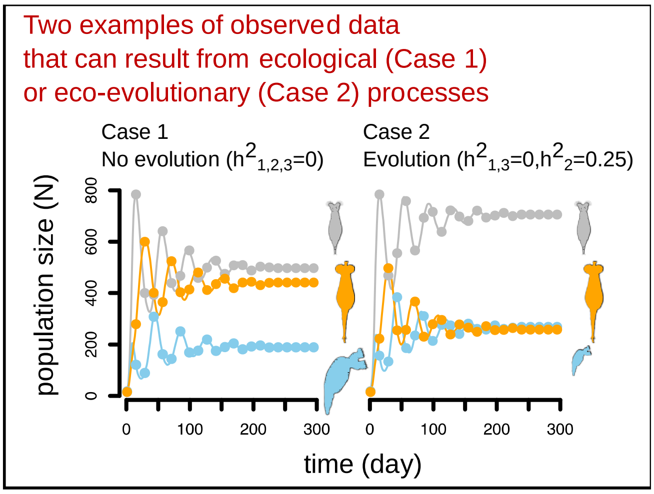
7) Competition is implemented via space limitation, determined by cell area.

8) In the RangeShifter model, all evolving traits are related to dispersal.



model	spatial scale	gene	protein	genome	quantitative trait	population dynamics	species interactions	community dynamics	phylogenetic dynamics	temporal scale	source language	Code URL
MESS	regional, biogeographic[1]	○[2]			○[3]	●	-/-,0/0	●	●	micro & macroevolution	Python	https://github.com/messDiv/MESS
gen3sis					●	●	-/-,0/0[4]		●	macroevolution	R	https://cran.r-project.org/package=gen3sis
BioGEEM	regional, biogeographic[5]				○[6]	●	-/-,0/0[7]	●	●	macroevolution	C++	https://github.com/julianoscabral/BioGEEM
sPEGG	local, regional	●		●	●	●	●	●		microevolution	C++	https://github.com/kewok/spegg
NEMO	local, regional	●		●		●	●			microevolution	C++	http://nemo2.sourceforge.net/
RangeShifter	regional, biogeographic	●		●	○[8]	●		●		microevolution	C++ : Windows GUI, Linux batch-mode, R package	https://github.com/rangeshifter
SLiM	local, regional	●		●	●	●	●			microevolution	C++, C, Eidos --> multi-platform GUI	http://messerlab.org/slim

Box 2. Eco-evolutionary hypothesis testing with ABC



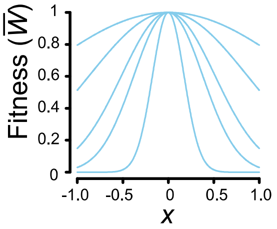
We demonstrate a five-step process to compare observed data to data produced by a theoretical model, to identify the most likely set of processes that produced the observed data. (a) **Candidate model development:** Our example of potential observed data is a time series (with values taken every two weeks, represented by points) of population size in a community with three species that coexist, compete for a shared resource, and have a key trait *x* (represented by the length of the organisms to the right of the plots) that impacts their mean fitness. The two time series were produced by a model with identical growth and competition parameters (values for interaction coefficients α_{ii} and α_{ij} were chosen to represent intransitive competition; [101]; see **Supplement 1** for additional parameter values of the simulation model used here), but in Case 1 the trait values are fixed and in Case 2, the trait values can evolve (via a model of evolutionary rescue; [102]). We focus on the data in Case 2 to use alternative hypothesis testing to identify the processes (ecological or eco-evolutionary) that produced the observed data. (b) **Simulations:** The first step is to come up with a model that could realistically produce the observed data under each of two alternative hypotheses. Here, we use a Beverton-Holt model with evolution in a heritable trait *x* that impacts population growth. We will compare Hypothesis 1 (H1) where evolution is not possible ($h^2 = 0$ for all species) and Hypothesis 2 (H2) where evolution is possible ($h^2 \geq 0$ for all species) and determine which is a better match to our observed data (the data in panel a, Case 2). We ran 100,000 simulations under the two alternative models. For all unknown or unmeasured parameters, a given simulation uses a random draw from a **prior distribution**, chosen to be uninformative (i.e., for model H2, heritability is drawn from a random uniform distribution between 0-1), informed by prior expert knowledge about the system (i.e. interaction coefficients are drawn from a beta distribution that places an increased likelihood on weak interactions ($\alpha \sim$ beta(0.25,10); [29]), or chosen using some value generator (e.g. as in [20])). For the random draws from the prior

Candidate model for observed data (Case 2)

$$N_{i,t+1} = \frac{\hat{W} e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}} N_{i,t}}{1+\alpha_{ii}N_{i,t}+\sum_j \alpha_{ij}N_{j,t}}$$

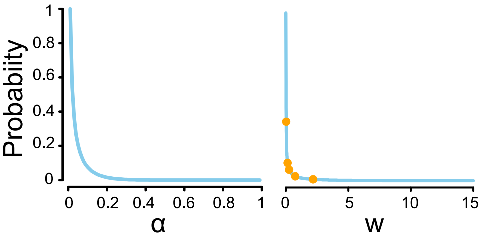
parameter	description
$N_{i,t}$	population size of species *i* at time *t*
$\hat{W}$	Growth rate of population with optimum trait value
P, *w*	width of phenotypic distribution, strength of Gaussian selection
$E\text{-}x_t$	Distance between environmental optimum trait value and population mean trait value
h^2	heritability
α_{ij}	Per-capita impact of species *j* on species *i*

Fitness distributions for values of *w*

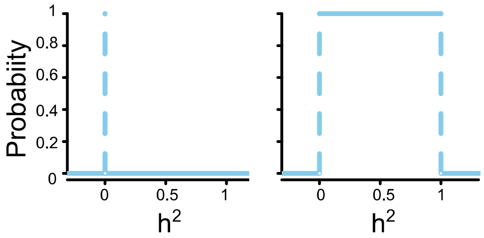


Prior parameter distributions under alternative ecological (H_1) and eco-evolutionary (H_2) hypotheses

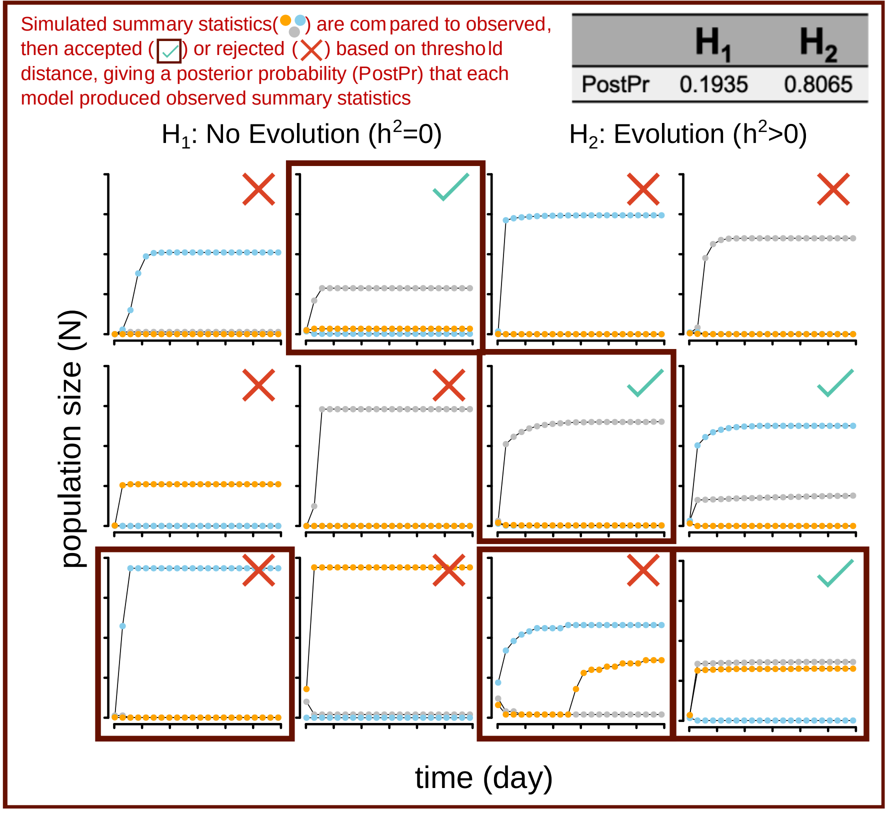
Prior shared under H_1 and H_2

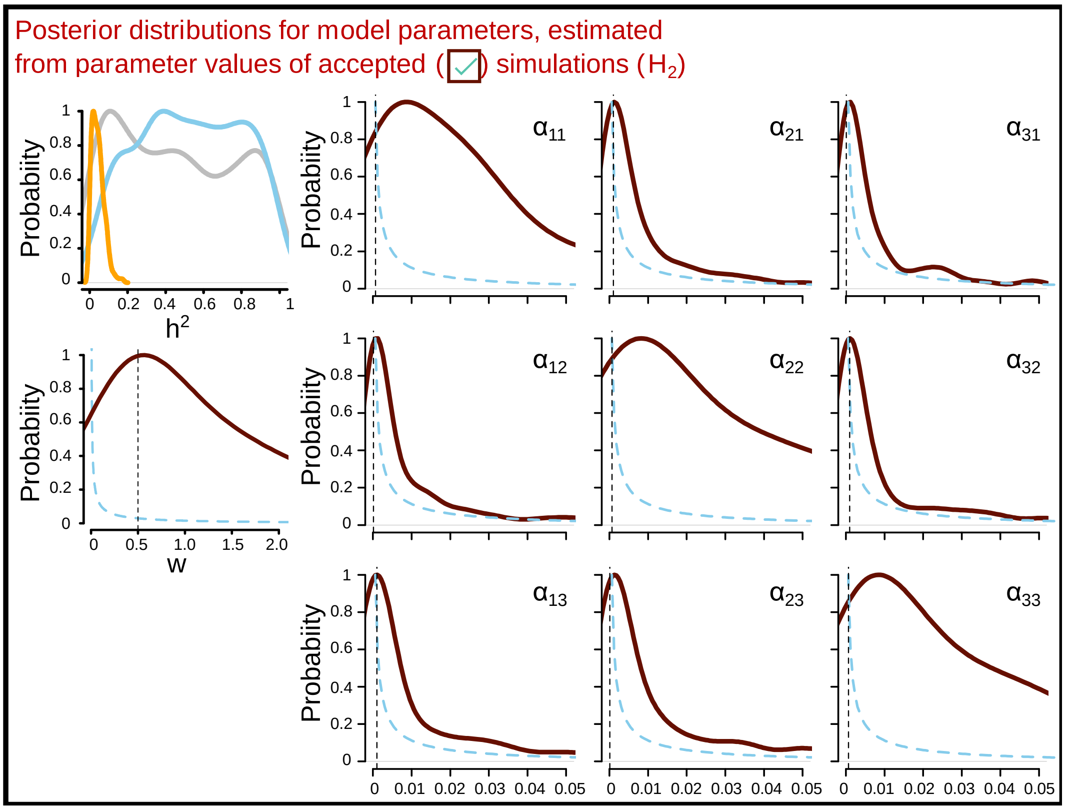


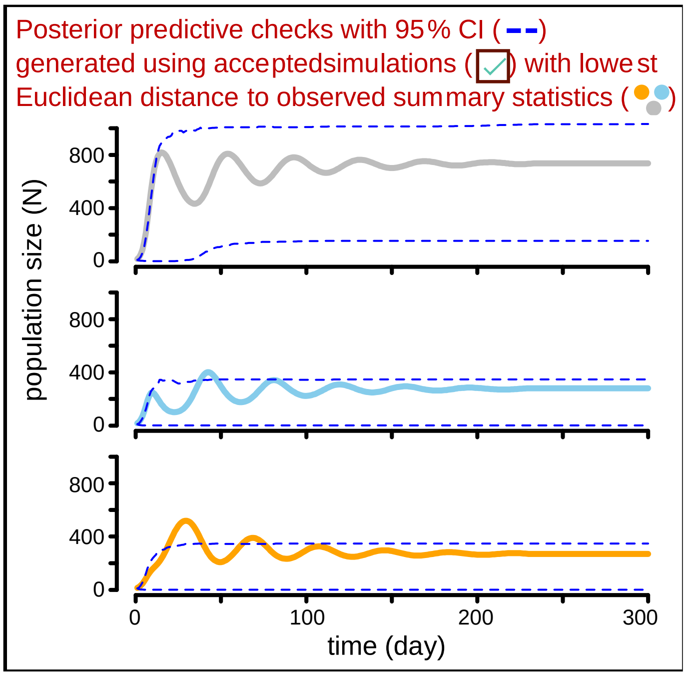
Prior differs under alternative ecological (H_1: $h^2 = 0$) and eco-evolutionary (H_2: $h^2 > 0$) hypotheses



distribution of *w* (the width of the Gaussian fitness function; example random draws are shown as orange points), the associated Gaussian fitness functions (the relationship between fitness on the *y*-axis and trait value on the *x*-axis) are shown. (c) **Model selection**: For each run of the simulation, the candidate parameter values drawn from the prior distributions produce a simulated dataset, which is then compared to the observed summary statistics (the points and $x_{t=300}$ trait values in panel a) using a distance function. We used a hidden-layer neural network to minimize the distance between observed and simulation summary statistics (using the R package *abc*; [36]). Some simulations are below this threshold distance and accepted (green checks) and some are above this threshold distance and rejected (red X). In this example, we limited the simulations to those where all 3 species had population size greater than 0, to better match the observed data (highlighted by dark red boxes). From this proportion of accepted models under each alternative hypothesis, a posterior model probability is generated (given in the table as PostPr). This indicates the probability the observed data was produced by each model. (d) **Parameter estimation:** Once a hypothesized model is identified (i.e. H2, the model with evolution in this instance, with a posterior probability of 80%), the randomly drawn parameter values that produced the accepted simulations can be used to estimate the **posterior**







distributions of model parameters that were not measured in the simulation [32]. In this example, the ABC process can successfully detect that Species 3 (colored orange) is unlikely to be the evolving species as the most probable values of h^2 are centered around 0. A higher weight is given to evolution ($h^2 > 0$) for Species 2 (colored light blue) compared to Species 1 (colored grey), but the ABC is unable to clearly resolve the true state of the system that $h^2_{1,3} = 0$ and $h^2_2=0.25$. We also show posterior distributions for the strength of selection *w* (prior: light blue, true value: black, posterior distribution: dark red) and intra- and interspecific interaction coefficients (α_{ii}) and α_{ij} respectively). (e) **Posterior predictions**: The quality of the selected model (H2, with evolution) is assessed by comparing the data (in panel a, Case 2) to predictions made under the accepted model and associated parameter posterior distributions. The 95% confidence intervals (blue dashed lines) for each species were obtained by running simulations using the parameter values from a subset of accepted simulations with lowest Euclidean distance to the observed values of summary statistics.