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 2, 45117 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 testing provides 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 modeling and statistical methods and discuss challenges for fitting mechanistic models to eco-evolutionary data.

Pairing ecological and evolutionary dynamics

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 feedback occurs 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 continuous 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, it remains that evolutionary processes are occurring while ecological dynamics proceed in all biological systems. 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 [3–6], but those impacts can be difficult to detect due to the often complex relationship between 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 (e.g., [7–9]) have implemented methods that take a target variable of interest (e.g., phenotypic trait, vital rate shared across the species under study), construct 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; [10]). These fractions are linear and additive, but lack a mechanistic basis; they are calculated retrospectively, and thus are limited for predicting future changes. Analysis methods for studying 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, Figure I**). 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 [11–13]. In this review we summarize 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 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, and 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, or their interaction in a feedback loop, account for observed 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, eco-evolutionary feedback) that are consistent with these observed data. The simulations provide features – referred to as **summary statistics** (see **Glossary**) – that are compared with 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 include spatial alpha diversity, range size variation, normalized lineage-through-time diversification curves [14], or diversity indices across three data axes (species abundances, population genetic variation, and trait values) [12]. We describe examples of such eco-evolutionary hypothesis formulations (**Box 1** and **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 it 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 some options for selecting an appropriate model. The choice of model should be based on knowledge of the system (e.g., mechanisms and processes 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 (see **Figure I** in **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 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 multispecies interaction models (*gauseR* [15]), consumer functional response models (*FRAIR* [16]), infectious disease models (*EpiDynamics* [17]), and island biogeography and metapopulation models (*EcoEvoApps* [18]). Researchers can use open- source code to extend the model to consider evolution. Alternatively, an existing EcoEvo Mathematica package (https://github.com/cklausme/EcoEvo) allows simulations and equilibrium analysis of species- and trait-based eco-evolutionary models, using differential equations with intra- and interspecific interactions, and 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 difficult to develop de novo, one can choose from the emerging class of **general-use eco-evolutionary simulation models** (**Box 1**, see online supplemental material S1F). 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 included features, such as implementing age- or stage-based population structure (NEMO-AGE; [19]), including molecular evolution (*NEMO-AGE*; *SLiM 4*; [20]), built-in models for organismal movement (*RangeShifter*; [21]), and the types of built-in species interactions (see **Figure I** in **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 [22–24]. 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 [25] and their effects on community structure), or identifying eco-evolutionary feedback dynamics. Testing these mechanistic hypotheses represents a step beyond showing only that ecology and evolution occur simultaneously.

To demonstrate eco-evolutionary hypothesis testing, we consider an example on a microevolutionary timescale at a single spatial location: an epidemiological model with virulence evolution [26], 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 feedbacks 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 multispecies communities on a large biogeographical scale, we used *gen3sis* [14] to simulate phenotypic evolution, speciation, and community assembly in a hypothetical community across 40 million years, and to monitor the impact of various eco-evolutionary processes for emergent patterns of species diversity. We consider four alternative hypotheses for evolution (**Figure 1B–E**): single versus 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. One could also test for the presence of eco-evolutionary feedbacks between evolution of dispersal ability and range expansion [27], 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 [28,29]. Eco-evolutionary simulation models with explicit consideration of sequence evolution (e.g., *NEMO-AGE* [17], *sPEGG* [24], *SLiM 4* [20]) (see **Figure I** in **Box 1**) allow tests for the role that **genetic architecture** plays in structuring eco-evolutionary dynamics (e.g., species with similar niche axes but different degrees of standing genetic variation, mating systems, or other properties that impact adaptive capacity [30]). 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 with data simulated under null and alternative eco-evolutionary hypotheses and their associated models, to estimate the posterior probability of each hypothesis. A decision-making or classifying algorithm can accept or reject simulations based on their similarity to summary statistics of observed data (**Boxes 1 and 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, traits, 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 [11,31]. Before the step of comparing observed to simulated summary statistics, one can combine reclassification of simulated datasets (e.g., leave-one-out **cross validation**) with a feature selection algorithm (e.g., a **boruta** algorithm [32]) to both identify summary statistics with the most predictive power for a given target dataset and to identify what data properties are most critical to collect for hypothesis testing. To address data at an unknown point along a trajectory towards system 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 system state), which can be set to obtain simulations away from equilibrium and can be estimated from an observed empirical dataset [12,33].

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 and 1 for a heritability parameter), or informative based on literature measures (e.g., an increased probability of weak species interactions) [34–36]. After simulations are run and summary statistics have been extracted, a decision algorithm can accept or reject a simulation based on its distance from observed summary statistics. We describe how to use **approximate Bayesian computation (ABC)** [37] for such an algorithmic approach in **Box 2**, but additional options are available: for example, **Markov**

chain Monte Carlo (MCMC) [38–40], **machine learning (ML)** [41] (see [42,43] for an overview of ML applied to model fitting in ecology and evolution, and supplemental material S1D for an ML alternative to the analysis in **Box 2**). ABC is a Bayesian statistical computational approach for estimating posterior distributions of model parameters via random draws from prior distributions and comparing with observed values via a criterion for acceptance or rejection [37,44]. Researchers can generate their own models, alternative hypotheses, and associated simulated summary statistics, and use an existing *abc* R package [45] for model comparison. One application of ABC to eco- evolutionary hypothesis testing is the study of Baselga *et al.* [46] to estimate the relative importance of dispersal limitation and niche width in contemporary surveys of genetic and community structure (see also [47]). ML 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. Overcast *et al.* [12] used this approach to classify whether datasets from contemporary surveys of diverse organismal groups were most consistent with neutral, environmental filtering, or competition eco-evolutionary models.

Parameter estimation, model validation, and generating future predictions

After model fitting is used to produce posterior probabilities for each candidate model, the posterior distributions of model parameters can also be estimated 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 useful for process inference, and can also serve as guidance for future data collection to confirm these estimates. Model checking and improvement is the next step: researchers should consider posterior predictive checks to evaluate model fit [48,49], evaluate the importance of summary statistics for predictive ability [41,50,51], 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 [52,53]. The best fit (or averaged) models and parameter estimates can 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 [54,55].

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 [56]. In one example, the study of Luiselli *et al.* [57] 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 depending on the mechanism of competition considered. Their results confirmed 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.

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 [8,58,59]. Studies of eco-evolutionary interactions and feedbacks are thus often limited to tractable taxonomic systems [60,61], 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, similarly to the way in which 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., [62–64]), 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 [20,65–67]). The remaining challenges to uncover eco-evolutionary signatures on extant biodiversity are generating data in contemporary populations and communities that capture the features necessary to infer eco-evolutionary structuring processes, 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 [68,69]) that combine to structure biodiversity from the gene to ecosystem level, and 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, biodiversity science is entering an unprecedented era of technology-assisted, high-throughput data collection [70–72] and of open, reproducible data sharing [73,74]. Some recent examples of ‘high-throughput’ eco-evolutionary data collection exist, including fine time series of multispecies population size and trait values via fluid imaging technology [75,76], real-time observation of adaptive tracking to environmental change (e.g., *Drosophila melanogaster* [24], *Arabidopsis thaliana* [77], other non-model systems [78–80]). Moving eco-evolutionary dynamics into the genomic era is a promising current direction of study [22,81]. The difficulty of detecting eco-evolutionary processes in complex systems could be alleviated by genomic data monitoring [82,83], 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 [84–86]. While this level of data coverage may not exist for all empirical studies, when available, such rich datasets can be used to evaluate summary statistics and data features that are potentially diagnostic of eco-evolutionary processes (**Box 1**, steps 1–4) and therefore useful as signatures of eco-evolutionary dynamics across diverse systems.

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 can be found in supplemental material S1E and S1F), and some additional reviews have addressed models that can be used to predict future biodiversity responses to environmental change [87,88]. 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 [89], and a species distribution model with local adaptation and phenotypic plasticity (ΔSDMs [90]). 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., [91–96]), 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 analyses here in an open-source R package *ecoevoR* (https://github.com/jhpantel/ecoevoR).

Future perspectives

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 [97] or features of community composition [98] or food web network structure [99] 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* [5,100,101], *Arabidopsis* [23]). However, it is currently unknown whether eco-evolutionary dynamics are driven by large-effect alleles versus 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., [102,103]), as these processes also structure genetic architecture, population demography, and species interactions [22,104–106]. 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. We acknowledge use of images from The Noun Project (DNA: Andrejs Kirma; genome: Emily Akins; flower 1: Maria Zamchy; flower 2: Tatyana; snake: TkBt; phylogeny: Leslie Coonrod), PhyloPic (butterflies: *Heliconius doris*, *H. melpomene*, *H. erato*; rotifer 1 and 3: *Philodinidae*; rotifer 2: *Rotaria macrura*, credit: Diego Fontaneto, Elisabeth A. Herniou, Chiara Boschetti, Manuela Caprioli, Giulio Melone, Claudia Ricci, and Timothy G. Barraclough (vectorized by T. Michael Keesey), license at https://creativecommons.org/licenses/by/3.0/ ; bird: *Corvus brachyrhynchos*), and the iCn3D protein viewer (using the ribonuclease H protein, PDB 2RN2, https://doi.org/10.2210/pdb2RN2/pdb ; Wang J, Youkharibache P, Marchler-Bauer A, Lanczycki C, Zhang D, Lu S, Madej T, Marchler GH, Cheng T, Chong LC, Zhao S, Yang K, Lin J, Cheng Z, Dunn R, Malkaram SA, Tai C-H, Enoma D, Busby B, Johnson NL, Tabaro F, Song G, Ge Y. iCn3D: From Web-Based 3D Viewer to Structural Analysis Tool in Batch Mode. Front. Mol. Biosci. 2022 9:831740.).

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tree.2023.03.011. The code used for all models and analyses is available in the R package *ecoevoR*: https://github.com/jhpantel/ecoevoR.

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. Hairon, 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. (2023) Coupling eco-evolutionary mechanisms with deep-time environmental dynamics to understand biodiversity patterns. Ecography, e06132 https://doi.org/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, O. and Silveira Marques, F. (2020) EpiDynamics: Dynamic Models in Epidemiology. R package version 0.3. 0. 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. (2022) SLiM 4: Multispecies eco-evolutionary modeling. Am. Nat. Published online December 1, 2022. https://doi.org/10.1086/723601

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. Sakarchi, J. and Germain, R.M. (2023) The evolution of competitive ability. Am. Nat. 201, 1–15

26. 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, S58–S75

27. Fronhofer, E.A. and Altermatt, F. (2015) Eco-evolutionary feedbacks during experimental range expansions. Nat. Commun. 6, 6844

28. Pillai, P. and Guichard, F. (2012) Competition-colonization trade-offs, competitive uncertainty, and the evolutionary assembly of species. PLoS One 7, e33566

29. Pettersen, A.K. et al. (2020) Metabolic rate, context-dependent selection, and the competition-colonization trade-off. Evol. Lett. 4, 333–344

30. De Meester, L. et al. (2002) The monopolization hypothesis and the dispersal–gene flow paradox in aquatic organisms. Acta Oecol. 23, 121–135

31. Kinnison, M.T. et al. (2015) Cryptic eco-evolutionary dynamics. Ann. N. Y. Acad. Sci. 1360, 120–144

32. Kursa, M.B. and Rudnicki, W.R. (2010) Feature selection with the Boruta package. J. Stat. Softw. 36, 1–13

33. Overcast, I. et al. (2019) An integrated model of population genetics and community ecology. J. Biogeogr. 46, 816–829

34. Wootton, J.T. and Emmerson, M. (2005) Measurement of interaction strength in nature. Annu. Rev. Ecol. Evol. Syst. 36, 419–444

35. Ellison, A.M. (2004) Bayesian inference in ecology. Ecol. Lett. 7, 509–520

36. Hooten, M.B. and Hobbs, N.T. (2015) A guide to Bayesian model selection for ecologists. Ecol. Monogr. 85, 3–28

37. Beaumont, M.A. (2010) Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41, 379–406

38. Marjoram, P. et al. (2003) Markov chain Monte Carlo withoutvlikelihoods. Proc. Natl. Acad. Sci. U. S. A. 100, 15324–15328

39. Hartig, F. et al. (2011) Statistical inference for stochastic simulation models – theory and application. Ecol. Lett. 14, 816–827

40. Newman, K. et al. (2022) State–space models for ecological time-series data: practical model-fitting. Methods Ecol. Evol. 14, 26–42

41. Pichler, M. and Hartig, F. (2023) Machine learning and deep learning – a review for ecologists. Methods Ecol. Evol. 14, 994–1016

42. Lucas, T.C.D. (2020) A translucent box: interpretable machine learning in ecology. Ecol. Monogr. 90, e01422

43. Borowiec, M.L. et al. (2022) Deep learning as a tool for ecology and evolution. Methods Ecol. Evol. 13, 1640–1660

44. Sunnåker, M. et al. (2013) Approximate Bayesian computation. PLoS Comput. Biol. 9, e1002803

45. Csilléry, K. et al. (2012) ABC: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3, 475–479

46. 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

47. 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

48. Gelman, A. et al. (1996) Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sin. 6, 733–760

49. Conn, P.B. et al. (2018) A guide to Bayesian model checking for ecologists. Ecol. Monogr. 88, 526–542

50. 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

51. 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

52. Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution. Trends Ecol. Evol. 19, 101–108

53. Dormann, C.F. et al. (2018) Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecol. Monogr. 88, 485–504

54. Eyring, V. et al. (2019) Taking climate model evaluation to the next level. Nat. Clim. Chang. 9, 102–110

55. McIntire, E.J.B. et al. (2022) PERFICT: a re-imagined foundation for predictive ecology. Ecol. Lett. 25, 1345–1351

56. Tilman, D. (1999) The ecological consequences of changes in biodiversity: a search for general principles. Ecology 80, 1455–1474

57. Luiselli, J. et al. (2022) Detecting the ecological footprint of selection. bioRxiv Published online December 10, 2022. https://doi.org/10.1101/2021.05.11.442553

58. Fussmann, G.F. et al. (2007) Eco-evolutionary dynamics of communities and ecosystems. Funct. Ecol. 21, 465–477

59. Schoener, T.W. (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. Science 331, 426–429

60. De Meester, L. and Pantel, J. (2014) Eco-evolutionary dynamics in freshwater systems. J. Limnol. 73, 193–200

61. De Meester, L. et al. (2019) Analysing eco-evolutionary dynamics – the challenging

complexity of the real world. Funct. Ecol. 33, 43–59

62. Sokal, R.R. and Riska, B. (1981) Geographic variation in Pemphigus populitransversus (Insecta: Aphididae). Biol. J. Linn. Soc. Lond. 15, 201–233

63. 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

64. Haller, B.C. and Messer, P.W. (2019) SLiM 3: forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol. 36, 632–637

65. Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214

66. Liu, L. (2008) BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24, 2542–2543

67. McPeek, M.A. (2008) The ecological dynamics of clade diversification and community assembly. Am. Nat. 172, E270–E284

68. Vellend, M. (2010) Conceptual synthesis in community ecology. Q. Rev. Biol. 85, 183–206

69. Vellend, M. (2016) The Theory of Ecological Communities, Monographs in Population Biology Vol. 57. Princeton University Press

70. Michener, W.K. and Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27, 85–93

71. Keitt, T.H. and Abelson, E.S. (2021) Ecology in the age of automation. Science 373, 858–859

72. Besson, M. et al. (2022) Towards the fully automated monitoring of ecological communities. Ecol. Lett. 25, 2753–2775

73. 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

74. Powers, S.M. and Hampton, S.E. (2019) Open science, reproducibility, and transparency in ecology. Ecol. Appl. 29, e01822

75. 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, e2104863118

76 Gibert, J.P. et al. (2023) Rapid eco-phenotypic feedback and the temperature response of biomass dynamics. Ecol. Evol. 13, e9685

77. Fulgione, A. et al. (2022) Parallel reduction in flowering time from de novo mutations enable evolutionary rescue in colonizing lineages. Nat. Commun. 13, 1461

78. Gienapp, P. et al. (2017) Genomic quantitative genetics to study evolution in the wild. Trends Ecol. Evol. 32, 897–908

79. 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

80. 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

81. Rudman, S.M. et al. (2018) What genomic data can reveal about eco-evolutionary dynamics. Nat. Ecol. Evol. 2, 9–15

82. Davies, N. et al. (2014) The founding charter of the Genomic Observatories Network. Gigascience 3, 2

83. Emerson, B.C. et al. (2022) Collective and harmonized high throughput barcoding of insular arthropod biodiversity: toward a Genomic Observatories Network for island. Mol. Ecol. Published online September 6, 2022, https://doi.org/10.22541/au.164840312.23663001

84. Shafer, A.B.A. et al. (2016) Forecasting ecological genomics: high-tech animal instrumentation meets high-throughput sequencing. PLoS Biol. 14, e1002350

85. Deiner, K. et al. (2017) Environmental DNA metabarcoding: transforming how we survey animal and plant communities. Mol. Ecol. 26, 5872–5895

86. Ellegaard, M. et al. (2020) Dead or alive: sediment DNA archives as tools for tracking aquatic evolution and adaptation. Commun. Biol. 3, 169

87. Waldvogel, A. et al. (2020) Evolutionary genomics can improve prediction of species' responses to climate change. Evol. Lett. 4, 4–18

88. Urban, M.C. et al. (2021) Coding for life: designing a platform for projecting and protecting global biodiversity. Bioscience 72, 91–104

89. Lasky, J.R. et al. (2020) What processes must we understand to forecast regional-scale population dynamics? Proc. Biol. Sci. 287, 20202219

90. Benito Garzón, M. et al. (2019) ΔTraitSDMs: species distribution models that account for local adaptation and phenotypic plasticity. New Phytol. 222, 1757–1765

91. 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

92. Tikhonov, G. et al. (2020) Joint species distribution modelling with the R-package Hmsc. Methods Ecol. Evol. 11, 442–447

93. 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

94. 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

95. Devarajan, K. et al. (2020) Multi-species occupancy models: review, roadmap, and recommendations. Ecography 43, 1612–1624

96. 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

97. Boussange, V. and Pellissier, L. (2022) Eco-evolutionary model on spatial graphs

reveals how habitat structure affects phenotypic differentiation. Commun. Biol. 5, 668

98. Govaert, L. et al. (2021) Measuring the contribution of evolution to community trait structure in freshwater zooplankton. Oikos 130, 1773–1787

99. Pantel, J.H. et al. (2017) 14 questions for invasion in ecological networks. Adv. Ecol. Res. 56, 293–340

100. Colbourne, J.K. et al. (2011) The ecoresponsive genome of Daphnia pulex. Science 331, 555–561

101. 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

102. Gomulkiewicz, R. and Holt, R.D. (1995) When does evolution by natural selection prevent extinction? Evolution 49, 201–207

103. Hermann, R.J. and Becks, L. (2022) Change in prey genotype frequency rescues predator from extinction. R. Soc. Open Sci. 9, 220211

104. Farkas, T.E. et al. (2013) Evolution of camouflage drives rapid ecological change in an insect community. Curr. Biol. 23, 1835–1843

105. Rodríguez-Verdugo, A. et al. (2017) The genomic basis of eco-evolutionary dynamics. Mol. Ecol. 26, 1456–1464

106. Svensson, E.I. (2019) Eco-evolutionary dynamics of sexual selection and sexual conflict. Funct. Ecol. 33, 60–72

107. Connolly, S.R. et al. (2017) Process, mechanism, and modeling in macroecology. Trends Ecol. Evol. 32, 835–844

108. 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

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

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)

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 referred 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?

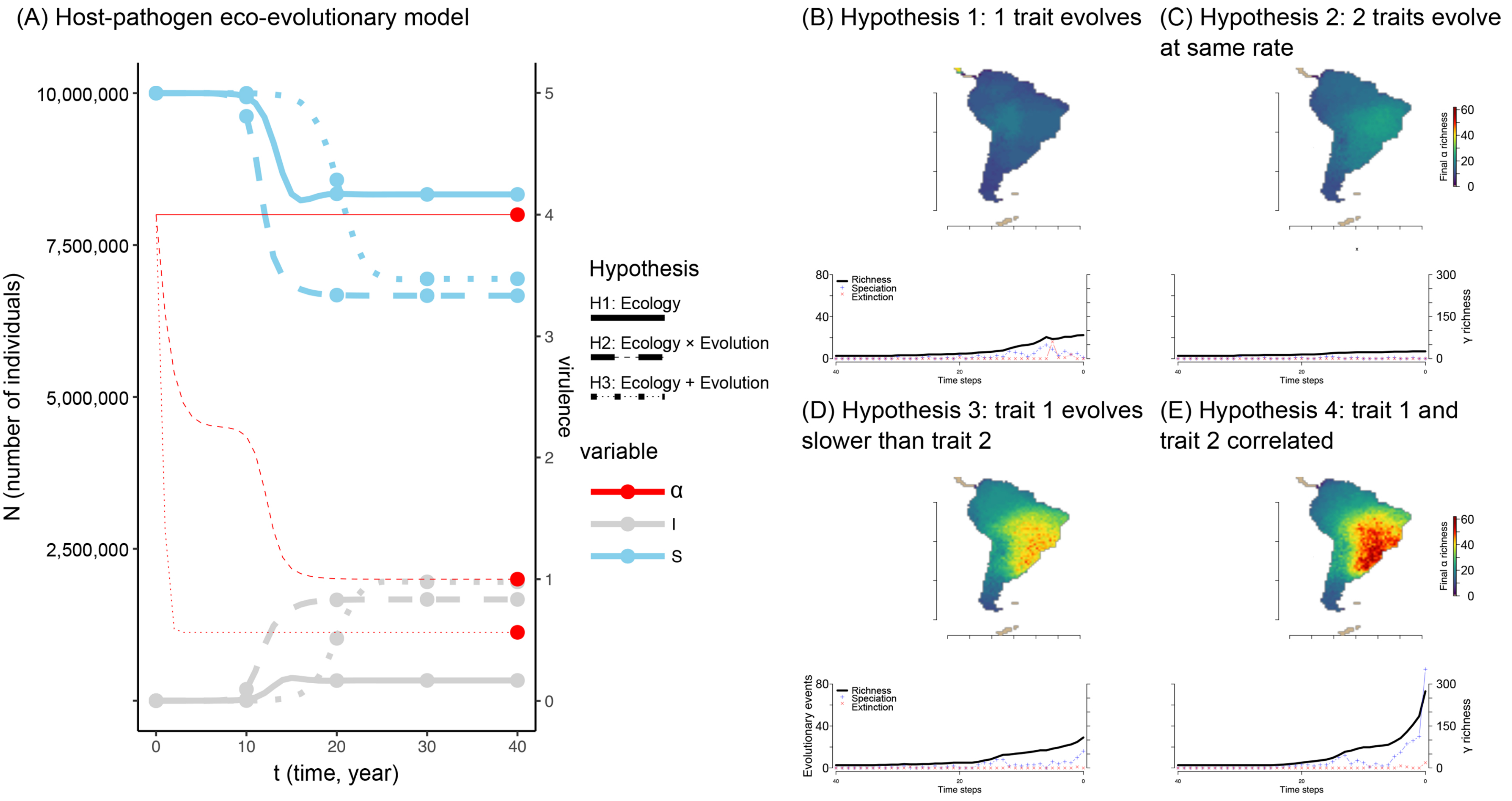
3. How do sexual selection, genetic drift, or evolutionary mismatch structure genetic architecture, population demography, species interactions, and eco-evolutionary dynamics?

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 the structure of observed disease and biodiversity dynamics. (A) Simulations of an eco-evolutionary epidemiological model with virulence evolution [26] are shown with many of the same model parameters (e.g., $S_0 = 10^7$, $\alpha_0 = 4$, $\gamma = 2$; see supplemental material S1A for additional parameter values), but under three alternative hypotheses that can be compared to observed data. Population sizes of susceptible (*S*, blue) and infected (*I*, gray) individuals (left *y*-axis), and pathogen virulence (α, red, right *y*-axis), are given for: hypothesis 1 (H1, unbroken line) in an ecological-only model where pathogen traits (α) cannot evolve, hypothesis 2 (H2, broken line) in an eco-evolutionary 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 show a coarser time series that could potentially be used for comparing observed to model-simulated data (red points are the final values for pathogen virulence), as fine time series with records of all population and trait dynamics are difficult to obtain. These comparisons of observed to model-generated data under alternative hypotheses (**Box 2**) can help identify whether trait evo- lution and/or eco-evolutionary feedbacks structured the observed data. (B–E) Simulations of speciation, trait evolution, and community assembly using the *gen3sis* eco-evolutionary simulation engine [14], for a hypothetical community of ten ancestral species, randomly placed as a single initial population on the map of South America (an 81 x 71 grid with cell-specific temperature values that impact carrying capacity and also population growth via the distance between a population's trait value and the local optimum). 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 (see supplemental material S1B for additional parameter values). We run simulations under four alternative eco-evolutionary hypotheses: (B) hypothesis 1 with evolution in a single temperature-related trait that determines population growth, (C) hypothesis 2 with evolution proceeding at the same rate in the temperature trait and in an additional dispersal-related trait that impacts the scale of the dispersal kernel, (D) hypothesis 3 with evolution in these two traits that proceeds at different rates, and (E) hypothesis 4 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, color-scale per grid cell in map) and time series of some summary statistics (number of speciation and extinction events, total regional richness) are shown. These summary statistics could be compared to observed values to distinguish among alternative hypotheses that may have structured observed data (**Box 2**).



(A) Host-pathogen eco-evolutionary model

(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

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

1. Identify research question and data

(i) Establish research questions, identify mechanisms thought to underlie expected dynamics and patterns, collect data (observational, experimental, field or lab- based, according to research needs).

(ii) Identify relevant summary statistics that are emergent properties of the data: for example, nucleotide diversity, within- and among-population genetic variation, species diversity (alpha, beta, gamma), center and width of phenotypic distributions, phylogenetic structure, total community abundance, rank abundance curves, normalized lineage-through-time diversification curves [14], variation in range sizes, summary statistics across three data axes (species abundances, population genetic variation, trait values [12]). Choosing the most informative summary statistics emerges from the research question and available data (e.g., total abundance data at one time point is unlikely to produce informative model comparisons). Summary statistics should be assessed before data collection (steps 2–4) and inform the type of data needed to capture the signature of particular eco-evolutionary dynamics. Summary statistics can come from time series or individual time points, 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

(i) Identify relevant biodiversity (from genes to communities), spatial, and temporal scales, single or multiple species, intraspecific and interspecific interactions (e.g., consumer resource, infectious disease dynamics, multispecies competition).

(ii) Use theoretical or **general-use eco-evolutionary simulation model** (**Figure I**) that best matches Step 1 and 2(i) and implements evolution (e.g., molecular genetics, quantitative genetics, or trait adaptive dynamics), ecology (e.g., density-dependent processes, consideration of species interactions), and coupled eco-evolutionary processes.

3. Generate alternative hypotheses for processes that structure observed data

(i) Formulate null and alternative hypotheses expressed as 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 versus adaptive evolution), or eco-evolutionary feedback (e.g., density-dependent trait evolution, dispersal-driven maladaptation impacting local fitness) relationships (**Figure I**).

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

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

(i) Identify algorithm to compare simulated and observed data:

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

(ii) Evaluate summary statistics in simulated data for sufficiency to discriminate 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]); see step 1(ii).

(iii) Compare summary statistics in simulated data and observed data.

(iv) Exclude simulations based on broad criteria (e.g., where a simulated population becomes extinct but an observed population persists) where simulation does not resemble main features of observed data.

(v) For ABC: compare simulated summary statistics to observed, reject simulations above a threshold similarity. For all: generate posterior probabilities for observed data under each alternative candidate model.

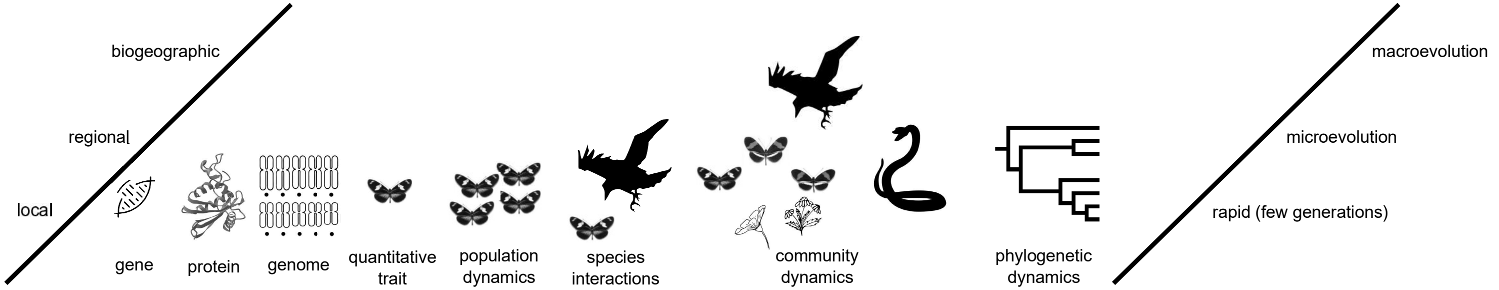
5. Parameter estimation, model validation, and generating future predictions

(i) 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 [52,53].

(ii) Use posterior predictive checks to evaluate model fit [48,49].

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

Box 1, Figure I. 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 modeled spatial, temporal, and biodiversity scales. Filled symbols indicate whether a biodiversity level is modeled, and open symbols indicate some special considerations needed to apply the model to that scale (corresponding numbers are explained in online supplemental material S1E). 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).



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	○[2]	[1]		○[3]	●	-/-,0/0	●	●	micro & macroevolution	Python	https://github.com/messDiv/MESS
gen3sis					●	●	●[4]	●	●	macroevolution	R, C++	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. Five-step process of eco-evolutionary hypothesis testing with ABC

Candidate model development

Our example observations are biweekly abundance data (**Figure IA**, points) of three coexisting species that compete for a shared resource, and have a key trait *x* (represented by the length of the pictured organisms) that impacts mean fitness. The two time series were produced by a model with identical growth and competition parameters (intransitive competition [108], additional parameter values in supplemental material S1C) but trait values are fixed in Case 1 and can evolve in Case 2 (evolutionary rescue) [102]. We focus on the data in Case 2 for alternative hypothesis testing to identify the processes (ecological or eco-evolutionary) that produced the observed data.

Simulations

The first step is to come up with a model that could realistically produce the observed data under alternative hypotheses (**Figure 1B**). We use a Beverton–Holt model with evolution in a heritable trait *x* that impacts population growth. We compare hypothesis 1 (H_1) without evolution ($h^2 = 0$ for all species) and hypothesis 2 (H_2) with evolution ($h^2 \geq 0$ for all species) to our observed data (**Figure IA**, 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 (**Figure IB**; i.e., for model H_2, heritability is drawn from a random uniform distribution between 0 and 1), informed by prior knowledge about the system (i.e., interaction coefficients are drawn from a beta distribution with increased likelihood of weak interactions ($\alpha \sim \text{beta}(0.25,10)$ [34]), or using some value generator [14]. For the random draws from the prior distribution of *w* (width of the Gaussian fitness function; example draws are shown as orange points), the associated fitness functions (the relationship between fitness, *y*-axis, and trait value, *x*-axis) are shown.

Model selection

For each simulation run, the candidate parameter values drawn from prior distributions produce a simulated dataset, which is then compared to the observed summary statistics (points and $x_{t=300}$ trait values in **Figure IA**) using a distance function (**Figure IC**) (a hidden-layer neural network that minimizes distance between observed and simulation summary statistics; R package *abc* [45]). Simulations below a threshold distance are accepted (green checks) and above this threshold are rejected (red X). Here, we limited simulations to those where all three species had a population size >0 (black boxes), to better match the observed data. From this proportion of accepted models (black box, green check) under each alternative hypothesis, a posterior model probability is generated (PostPr). This indicates the probability that the observed data were produced by each model.

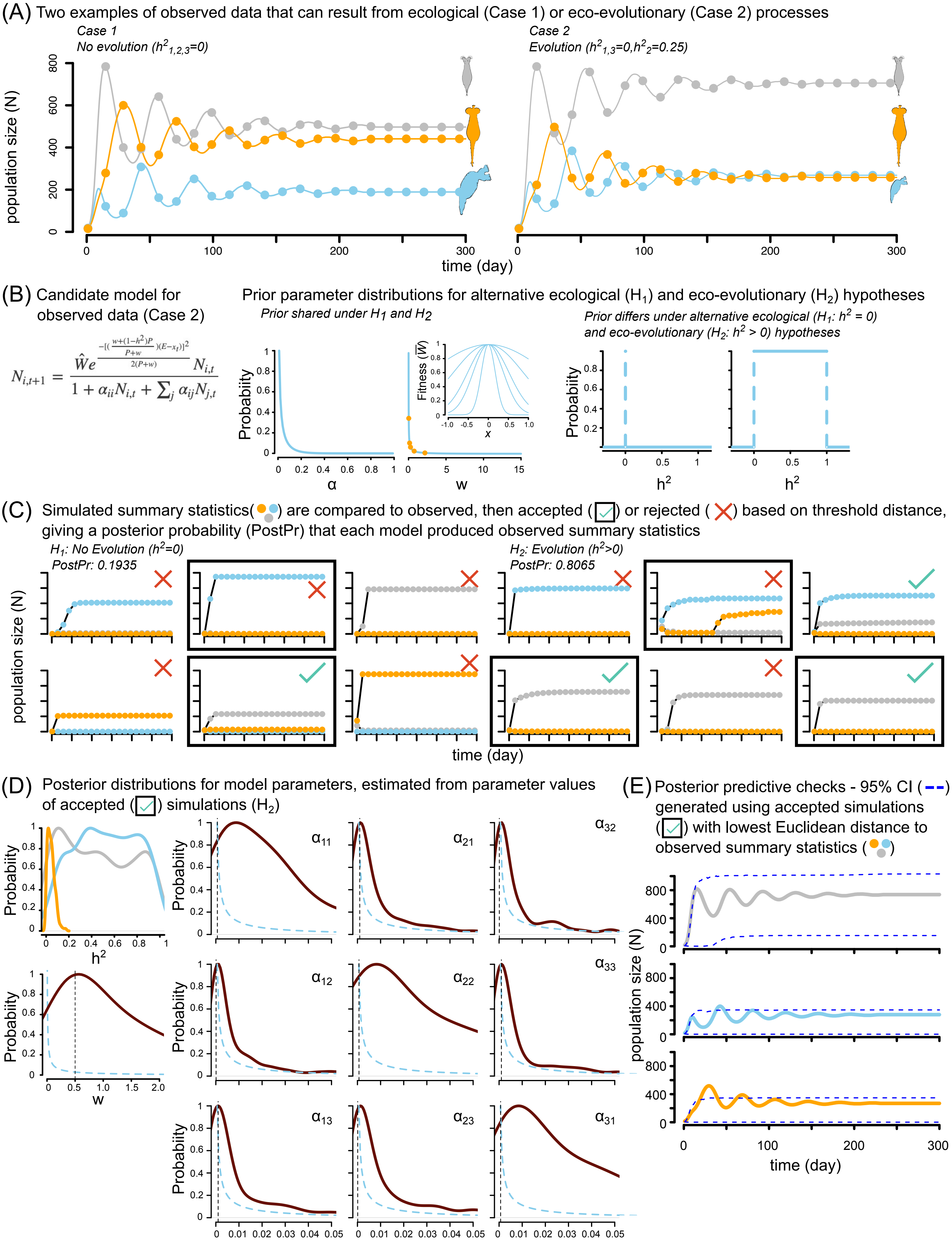
Parameter estimation

Once a hypothesized model is identified (H_2 with evolution, 81% posterior probability), the randomly drawn parameter values that produced accepted simulations can estimate the posterior distributions of model parameters [37] (**Figure ID**). Here, the ABC process can successfully detect that species 3 (orange) is unlikely to be evolving as the most probable values of h^2 are centered around 0, but it cannot clearly resolve the true system state that $h^2_{1,3} = 0$ and $h^2_2=0.25$ (species 2: blue, species 3: gray). We also show posterior distributions for the strength of selection *w* (prior: dashed blue line, true value: dashed black line, posterior distribution: solid black line) and interaction coefficients (α_{ii} and α_{ij}).

Posterior predictions

The quality of the selected model (H_2, with evolution) is assessed by comparing the data (**Figure IA**, Case 2) to predictions made under the accepted model and associated parameter posterior distributions. The 95% confidence intervals (**Figure IE**, blue broken lines) 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.

Box 2, Figure I. Illustrated five-step process of eco-evolutionary hypothesis testing with ABC (full methods are given in supplementary material S1C).


(A) Two examples of observed data that can result from ecological (Case 1) or eco-evolutionary (Case 2) processes
Case 1
No evolution ($h^2_{1,2,3}=0$)
Case 2
Evolution ($h^2_{1,3}=0, h^2_2=0.25$)
population size (N)
0
200
400
600
800
0
100
200
300
time (day)
(B) Candidate model for observed data (Case 2)
$N_{i,t+1} = \frac{\hat{W}e^{\frac{-[(\frac{w+(1-h^2)P}{P+w})(E-x_t)]^2}{2(P+w)}}N_{i,t}}{1+\alpha_{ii}N_{i,t}+\sum_j \alpha_{ij}N_{j,t}}$
Prior parameter distributions for 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
Probability
α
w
Fitness ($\overline{W}$)
x
h^2
(C) Simulated summary statistics(●●●) are compared to observed, then accepted (✓) or rejected (✗) based on threshold distance, giving a posterior probability (PostPr) that each model produced observed summary statistics
H_1: No Evolution ($h^2=0$)
PostPr: 0.1935
H_2: Evolution ($h^2>0$)
PostPr: 0.8065
population size (N)
time (day)
(D) Posterior distributions for model parameters, estimated from parameter values of accepted (✓) simulations (H_2)
Probability
h^2
w
α_{11}
α_{21}
α_{32}
α_{12}
α_{22}
α_{33}
α_{13}
α_{23}
α_{31}
0
0.01
0.02
0.03
0.04
0.05
(E) Posterior predictive checks - 95% CI (- -) generated using accepted simulations (✓) with lowest Euclidean distance to observed summary statistics (●●●)
population size (N)
time (day)
0
400
800
0
100
200
300