



HAL
open science

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

Jelena H Pantel, Lutz Becks

► **To cite this version:**

Jelena H Pantel, Lutz Becks. Statistical methods to identify mechanisms in studies of eco-evolutionary dynamics. *Trends in Ecology & Evolution*, 2023, 38 (8), pp.760 - 772. 10.1016/j.tree.2023.03.011 . hal-04735915

HAL Id: hal-04735915

<https://hal.science/hal-04735915v1>

Submitted on 22 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

2

3 **Authors:** Jelena H. Pantel^{1,3}, Lutz Becks²

4

5 ¹Ecological Modelling, Faculty of Biology, University of Duisburg-Essen, Universitätsstraße
6 2, 45117 Essen, Germany

7

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

10

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

13

14 **Keywords:**

15 - Hypothesis testing

16 - Biodiversity

17 - Bayesian statistics

18 - Eco-evolutionary dynamics

19 - Mechanistic models

20

21 **Abstract:**

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

32

33

34 *Pairing ecological and evolutionary dynamics*

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

54
55 Some studies (e.g., [7–9]) have implemented methods that take a target variable of interest
56 (e.g., phenotypic trait, vital rate shared across the species under study), construct categories
57 of contributing factors to variation in that target variable, and assigns total phenotypic
58 variance to these fractions (often by holding one process constant and averaging across the
59 change in the other effects; [10]). These fractions are linear and additive, but lack a
60 mechanistic basis; they are calculated retrospectively, and thus are limited for predicting
61 future changes. Analysis methods for studying eco-evolutionary dynamics should instead
62 consider the mechanistic basis of data structure, realistic models of variance and uncertainty,
63 and the complexity of eco-evolutionary feedbacks and processes with distinct signatures at
64 different temporal, spatial, and biodiversity scales (**Box 1, Figure I**). Model-based hypothesis
65 testing by comparing among alternative eco-evolutionary hypotheses has recently emerged
66 as an alternative to existing correlative approaches, and this framework can be used to
67 address the unanswered questions in the field [11–13]. In this review we summarize the basic
68 building blocks of using mechanistic models and statistical inference to pair eco-evolutionary
69 data with a theoretical or simulation model. We provide guidance on how to construct
70 alternative hypotheses, fit these to observed data, and generate posterior probabilities for
71 these hypothesized processes. The goal of statistical inference for eco-evolutionary
72 dynamics is not that all datasets must perfectly capture each step of historical dynamics, nor
73 that all datasets must fit to the same ANOVA-based linear model to categorize average effect
74 sizes. We provide directions to make inference about the potentially complex processes that
75 structure eco-evolutionary data, to identify whether eco-evolutionary feedbacks are operating
76 in a specific system, and to determine the consequences of these dynamics for critical
77 emergent properties such as resilience, stability, diversity, extinction, and ecosystem
78 function.

79

80 *A structured workflow for model-based hypothesis testing in eco-evolutionary dynamics*
81 Eco-evolutionary research questions may center on whether only evolutionary or only
82 ecological processes in isolation, or their interaction in a feedback loop, account for observed
83 diversity patterns. Research questions can thus be formulated as a core of null and alternative
84 hypotheses expressed in the form of alternative competing mechanistic models (**Figure 1**).
85 Simulations from the models can then be matched to these alternative hypotheses and
86 rigorously compared to observed data to identify the variety of relevant processes (ecological,
87 evolutionary, eco-evolutionary feedback) that are consistent with these observed data. The
88 simulations provide features – referred to as **summary statistics** (see **Glossary**) – that are
89 compared with corresponding features in the observed dataset, and many thousands of
90 simulations should be run under each model to generate potential distributions of these
91 features. Examples of these summary statistics include spatial alpha diversity, range size
92 variation, normalized lineage-through-time diversification curves [14], or diversity indices
93 across three data axes (species abundances, population genetic variation, and trait values)
94 [12]. We describe examples of such eco-evolutionary hypothesis formulations (**Box 1** and
95 **Figure 1**), and how to compare observed data to these models (**Box 2**).

96

97 **Choose a mechanistic model for eco-evolutionary hypothesis testing**

98 The first step in eco-evolutionary hypothesis testing is to identify a mechanistic model to pair
99 with observed data. This potential challenge is not unique to eco-evolutionary dynamics, but
100 it is also a necessary step to make the field more predictive. Once the research question has
101 been identified and the relevant biological data have been collected, there are some options
102 for selecting an appropriate model. The choice of model should be based on knowledge of
103 the system (e.g., mechanisms and processes that determine ecological dynamics, patterns,
104 or outcomes), the relevant biodiversity (from genes to communities), the spatial and temporal
105 scales relevant to the data of interest, as well as the summary statistics used to compare
106 model simulations to observed data (see **Figure I** in **Box 1**). It is not always necessary to
107 develop an entirely new model or design an experiment that estimates all parameters in a
108 single theoretical model, as one can choose among existing models or tools to build models.
109 These could be theoretical models, or simulation models that reproduce the dynamics
110 described in theoretical models. Numerous R software packages exist that allow simulations
111 of commonly used ecological models such as logistic growth and multispecies interaction
112 models (*gauseR* [15]), consumer functional response models (*FRAIR* [16]), infectious
113 disease models (*EpiDynamics* [17]), and island biogeography and metapopulation models
114 (*EcoEvoApps* [18]). Researchers can use open- source code to extend the model to consider
115 evolution. Alternatively, an existing EcoEvo Mathematica package
116 (<https://github.com/cklausme/EcoEvo>) allows simulations and equilibrium analysis of
117 species- and trait-based eco-evolutionary models, using differential equations with intra- and
118 interspecific interactions, and quantitative genetic or adaptive dynamics trait models.

119

120 If models with known equations that capture the eco-evolutionary dynamics of interest do not
121 exist or are difficult to develop de novo, one can choose from the emerging class of **general-**
122 **use eco-evolutionary simulation models** (**Box 1**, see online supplemental material S1F).
123 These generally use a fixed set of input parameters to model sequence, phenotypic, and/or
124 phylogenetic evolution, where evolving traits impact how individuals grow, reproduce, die,
125 move across a landscape, and interact with their environment and with other individuals, and
126 where individuals inhabit a landscape with user-specified connectivity and movement

127 patterns. These models can simultaneously consider both ecological and evolutionary
128 diversity-generating processes (e.g., speciation, neutral and adaptive sequence and
129 phenotypic evolution, spatial and environmental drivers of population and community
130 structure), and generate known patterns that emerge from these processes (e.g., nucleotide
131 diversity patterns, rank abundance curves, species–genetic diversity correlations). The
132 models differ in some included features, such as implementing age- or stage-based
133 population structure (*NEMO-AGE*; [19]), including molecular evolution (*NEMO-AGE*; *SLiM 4*;
134 [20]), built-in models for organismal movement (*RangeShifter*; [21]), and the types of built-in
135 species interactions (see **Figure I** in **Box 1**).

136

137 **Generate alternative hypotheses for processes that structure observed data**

138 Research in eco-evolutionary dynamics has moved forward from asking whether evolution
139 impacts ecological processes to identifying the mechanism of this effect [22–24]. Linking
140 mechanistic modeling with observational data could allow investigation of these mechanisms
141 even if they are not easily accessible through direct measurements (e.g., in field surveys, or
142 for mechanisms that are typically accessible only through experimental manipulation). Some
143 potential examples of hypotheses that could be studied by comparing data to mechanistic
144 models include the precise form or mechanism of evolution (e.g., evolution of single or
145 multiple traits, uncorrelated or correlated, the existence of trade-offs between evolution of life
146 history traits and competition [25] and their effects on community structure), or identifying
147 eco-evolutionary feedback dynamics. Testing these mechanistic hypotheses represents a
148 step beyond showing only that ecology and evolution occur simultaneously.

149

150 To demonstrate eco-evolutionary hypothesis testing, we consider an example on a
151 microevolutionary timescale at a single spatial location: an epidemiological model with
152 virulence evolution [26], where the number of individuals in susceptible (*S*) and infected (*I*)
153 host populations are determined by a transmission–virulence tradeoff. Virulence (α) can
154 evolve if the additive genetic variance for this trait $V > 0$, and the rate of change for this trait
155 depends on the susceptible host population size: $da/dt = f(S)$. This link between the evolving
156 trait and host population dynamics represents an eco-evolutionary feedback as $dS/dt = f(\alpha)$
157 and $da/dt = f(S)$ [1]. A null hypothesis for comparing empirical data to this model system could
158 be that phenotypic evolution is not possible (the additive genetic variance for virulence $V =$
159 0), with an alternative hypothesis where evolution ($V > 0$) and a feedback between evolution
160 and ecology is possible, and an additional alternative hypothesis where eco-evolutionary
161 feedbacks are removed from the system by decoupling the dependence of virulence evolution
162 from host population dynamics: $da/dt \neq f(S)$ (**Figure 1A**). Additional simulations can be used
163 to explore the impacts of different trade-off functions (between virulence and transmission,
164 or in the relationship between evolutionary change in virulence and population size of
165 susceptible hosts).

166

167 For an example of eco-evolutionary dynamics on macroevolutionary timescales in
168 multispecies communities on a large biogeographical scale, we used *gen3sis* [14] to simulate
169 phenotypic evolution, speciation, and community assembly in a hypothetical community
170 across 40 million years, and to monitor the impact of various eco-evolutionary processes for
171 emergent patterns of species diversity. We consider four alternative hypotheses for evolution
172 (**Figure 1B–E**): single versus multiple traits, evolving at the same or different speeds, with
173 traits either uncorrelated or correlated. Resulting patterns (e.g., of richness at the local α ,

174 regional γ , and among-site β levels, or of phylogenetic diversity) can be compared to
175 observed data, and model posterior probabilities can inform researchers of the likelihood that
176 different eco-evolutionary processes have structured their data. One could also test for the
177 presence of eco-evolutionary feedbacks between evolution of dispersal ability and range
178 expansion [27], the effects of uniform versus varying dispersal ability across species, or
179 develop an alternative configuration where traits evolve as a tradeoff between local
180 competitive ability and colonization ability [28,29]. Eco-evolutionary simulation models with
181 explicit consideration of sequence evolution (e.g., *NEMO-AGE* [17], *sPEGG* [24], *SLiM 4* [20])
182 (see **Figure I** in **Box 1**) allow tests for the role that **genetic architecture** plays in structuring
183 eco-evolutionary dynamics (e.g., species with similar niche axes but different degrees of
184 standing genetic variation, mating systems, or other properties that impact adaptive capacity
185 [30]). Eco-evolutionary simulation models can test the impacts of these processes for
186 emergent community properties such as coexistence, diversity, or resilience to perturbation.
187

188 **Compare simulated data under each alternative hypothesis to observed data**

189 The next step is to compare observed data with data simulated under null and alternative
190 eco-evolutionary hypotheses and their associated models, to estimate the posterior
191 probability of each hypothesis. A decision-making or classifying algorithm can accept or reject
192 simulations based on their similarity to summary statistics of observed data (**Boxes 1 and 2**).
193 As observed and simulated data are compared based on the match in information-rich
194 summary statistics, there is no a priori reason that absolute time series of sequences, traits,
195 or abundance data are necessary to successfully identify the processes most likely to have
196 generated an observed dataset. Diverse summary statistics across different scales and
197 biological levels may be critical to address the problem of equifinality, where different
198 combinations of processes or model parameters can produce similar patterns of data at a
199 single biological level [11,31]. Before the step of comparing observed to simulated summary
200 statistics, one can combine reclassification of simulated datasets (e.g., leave-one-out **cross**
201 **validation**) with a feature selection algorithm (e.g., a **boruta** algorithm [32]) to both identify
202 summary statistics with the most predictive power for a given target dataset and to identify
203 what data properties are most critical to collect for hypothesis testing. To address data at an
204 unknown point along a trajectory towards system equilibrium, the *MESS* model includes a
205 parameter Λ that measures the ‘fraction’ of equilibrium that a system has obtained (the point
206 at which the starting conditions of the simulation model are no longer detectable in the system
207 state), which can be set to obtain simulations away from equilibrium and can be estimated
208 from an observed empirical dataset [12,33].
209

210 It is not required to render guesses for or attempt measurements of all model parameters to
211 compare observed and simulated data. To generate the thousands of simulations needed
212 under each alternative hypothesis, researchers can instead sample candidate parameter
213 values from a proposed **prior distribution**. These prior distributions should be based on a
214 researcher’s understanding of the model system when possible (e.g., an uninformative
215 uniform distribution between 0 and 1 for a heritability parameter), or informative based on
216 literature measures (e.g., an increased probability of weak species interactions) [34–36]. After
217 simulations are run and summary statistics have been extracted, a decision algorithm can
218 accept or reject a simulation based on its distance from observed summary statistics. We
219 describe how to use **approximate Bayesian computation (ABC)** [37] for such an
220 algorithmic approach in **Box 2**, but additional options are available: for example, **Markov**

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

236

237 **Parameter estimation, model validation, and generating future predictions**

238 After model fitting is used to produce posterior probabilities for each candidate model, the
239 posterior distributions of model parameters can also be estimated from the simulations
240 retained by the acceptance–rejection algorithm (**Box 2**). These posterior estimates of model
241 parameters – some of which might have associated observations for validation, and others
242 of which are impossible or resource-intensive to estimate – are useful for process inference,
243 and can also serve as guidance for future data collection to confirm these estimates. Model
244 checking and improvement is the next step: researchers should consider posterior predictive
245 checks to evaluate model fit [48,49], evaluate the importance of summary statistics for
246 predictive ability [41,50,51], and also consider whether to use the model with the highest
247 posterior probability support or to average over the models with weights based on their
248 relative posterior support [52,53]. The best fit (or averaged) models and parameter estimates
249 can be used to make predictions for future dynamics of the system, with estimates of
250 uncertainty, and with inclusion of alternative scenarios for future systemic change [54,55].

251

252 For studies of eco-evolutionary dynamics, predictive simulations can be used to determine
253 the importance of eco-evolutionary dynamics and feedbacks for critical emergent properties
254 such as stability, extinction risk, or biodiversity maintenance. For example, in the host–
255 disease model (**Figure 1A**), the impacts of including versus excluding the eco-evolutionary
256 feedback loop for population stability could be evaluated by running simulations with
257 parameters drawn from posterior distributions under each alternative model, and then
258 generating a posterior distribution for an emergent property of the simulations such the
259 **coefficient of variation (CV)**, a metric often used to quantify stability [56]. In one example,
260 the study of Luiselli *et al.* [57] determined the relative impact of the combination of speciation
261 rate and mechanism of competition for population genetic and community structure in a
262 variety of taxonomic systems. They used the *MESS* simulation model to decompose the
263 effects of speciation rates and competition type for emergent summary statistics of overall
264 species richness, and community, genetic, and functional trait diversity. Their study found
265 strong differences in the change in genetic diversity over time depending on whether
266 speciation was included, whether species interacted neutrally, and depending on the
267 mechanism of competition considered. Their results confirmed that many of the properties

268 we observe at the level of populations and communities are generated by eco-evolutionary
269 processes. One recent study has established a mechanistic basis for the impact of eco-
270 evolutionary dynamics for a critical, emergent system property of food web collapse. Barbour
271 *et al.* [23] identified that the presence of a particular allele (AOP2⁺ or AOP2⁻) in host
272 *Arabidopsis thaliana* plants could result in a 29% difference in extinction rates for associated
273 aphids and parasitoids in an experimental food web.

274

275 *Future perspectives*

276 It has previously been said that confirmation of eco-evolutionary predictions requires
277 monitoring genetic properties of populations and subsequent community interactions over
278 time intervals in which selection regimes are likely to have caused changes in ecologically
279 relevant traits [8,58,59]. Studies of eco-evolutionary interactions and feedbacks are thus often
280 limited to tractable taxonomic systems [60,61], with time-intensive measurements of
281 phenotypes in common environments. The complex eco-evolutionary processes that provide
282 structure in data are often collapsed into additive categories with effect sizes estimated using
283 ANOVA-like variance partitioning without consideration for generalized probability
284 distributions, fixed and random effects, or non-linear and interacting combinations of drivers
285 of system responses (e.g., [8,9]). However, similarly to the way in which processes of
286 evolution and the dynamics of natural selection leave distinct signatures on contemporary
287 populations that can be analyzed to infer those historical structuring processes (e.g., [62–
288 64]), past eco-evolutionary dynamics likely leave distinct signatures on contemporary
289 populations and communities. Researchers must still scale the difficult challenge of
290 identifying which features, emergent properties, and summary statistics from observed data
291 capture the signatures of distinct eco-evolutionary processes (e.g., [3]). However, the path of
292 using statistical or **process-based mechanistic models** to compare the likelihood of
293 observed data to alternative models is used in other fields of research where historical
294 processes cannot be observed (e.g., coalescent processes and demographic history,
295 phylogenetic reconstruction and patterns, population genetic structure [20,65–67]). The
296 remaining challenges to uncover eco-evolutionary signatures on extant biodiversity are
297 generating data in contemporary populations and communities that capture the features
298 necessary to infer eco-evolutionary structuring processes, accessing user-friendly versions
299 of statistical models that can disentangle the numerous processes (drift, selection, gene flow,
300 community assembly and metacommunity processes, multivariate environmental forcing
301 [68,69]) that combine to structure biodiversity from the gene to ecosystem level, and
302 converging on mechanistic hypotheses and predictions for the impacts of eco-evolutionary
303 dynamics in biological systems.

304

305 There is good reason to be optimistic on each front. First, biodiversity science is entering an
306 unprecedented era of technology-assisted, high-throughput data collection [70–72] and of
307 open, reproducible data sharing [73,74]. Some recent examples of ‘high-throughput’ eco-
308 evolutionary data collection exist, including fine time series of multispecies population size
309 and trait values via fluid imaging technology [75,76], real-time observation of adaptive
310 tracking to environmental change (e.g., *Drosophila melanogaster* [24], *Arabidopsis thaliana*
311 [77], other non-model systems [78–80]). Moving eco-evolutionary dynamics into the genomic
312 era is a promising current direction of study [22,81]. The difficulty of detecting eco-
313 evolutionary processes in complex systems could be alleviated by genomic data monitoring
314 [82,83], especially in systems where candidate genes can be used for monitoring selection

315 in real time and eDNA and high-throughput sequencing can be used for monitoring
316 evolutionary dynamics and ranges of organisms across large spatial scales [84–86]. While
317 this level of data coverage may not exist for all empirical studies, when available, such rich
318 datasets can be used to evaluate summary statistics and data features that are potentially
319 diagnostic of eco-evolutionary processes (**Box 1**, steps 1–4) and therefore useful as
320 signatures of eco-evolutionary dynamics across diverse systems.

321
322 Second, we have discussed theoretical and simulation models that are moving towards
323 modeling the full suite of interacting processes that structure biodiversity at multiple scales
324 (spatial, temporal, diversity across levels of organization [**Box 1**]) (additional detail for models
325 can be found in supplemental material S1E and S1F), and some additional reviews have
326 addressed models that can be used to predict future biodiversity responses to environmental
327 change [87,88]. There is still a need for statistical tools to detect effects and their magnitudes
328 in experimental and field observational data, and the difficulty of distinguishing among the
329 many signals that structure genes, traits, populations, and communities is compounded by
330 the complexity of eco-evolutionary dynamics. Some new statistical models that address this
331 need include an integrated reaction norm model linking genetic, phenotypic, and
332 demographic processes [89], and a species distribution model with local adaptation and
333 phenotypic plasticity (Δ SDMs [90]). Modern statistical models for population and community
334 ecology now consider critical processes such as observer, measurement, and process error,
335 error propagation, or the existence of present but undetected species or life stages (e.g., [91–
336 96]), and these features are likely to be important for eco-evolutionary models as well. For
337 linking microevolution and models of eco-evolutionary dynamics, tools such as *sPEGG* [24],
338 *SLiM 4* [20], and the EcoEvo Mathematica package are currently available for generating
339 features to compare with observed data. We also share the code used for all the analyses
340 here in an open-source R package *ecoevoR* (<https://github.com/jhpantel/ecoevoR>).

341 342 *Future perspectives*

343 There is an urgent need to move beyond establishing that evolution can be important for
344 ecological processes or that eco-evolutionary feedback loops might exist in some systems.
345 Are there particular environmental and spatial conditions [97] or features of community
346 composition [98] or food web network structure [99] that make eco-evolutionary dynamics
347 more or less likely to occur? The lack of information here highlights the reality that a
348 background or null level of interacting eco-evolutionary processes is currently unknown (see
349 **Outstanding questions**). Published examples tend to focus on clear feedbacks between
350 adaptive evolution and ecological dynamics in organisms that may have disproportionate
351 adaptive capacity (e.g., *Daphnia* [5,100,101], *Arabidopsis* [23]). However, it is currently
352 unknown whether eco-evolutionary dynamics are driven by large-effect alleles versus allelic
353 changes in numerous locations with cumulative small effects. Additionally, the role of
354 adaptive evolution has been considered in many studies of eco-evolutionary dynamics, but
355 fewer have established the role that sexual selection, genetic drift, or maladaptation might
356 play (but see discussion of maladaptation in evolutionary rescue literature: e.g., [102,103]),
357 as these processes also structure genetic architecture, population demography, and species
358 interactions [22,104–106]. Given that numerous evolutionary and ecological processes are
359 continuously occurring, operating at a variety of time scales, answers to these research
360 questions are needed for statistical models to be developed that implement accurate
361 distributions for emergent properties that are often the targets of studies of the impacts of

362 environmental change. Much of this review focuses on the methodology needed to detect
363 signatures of eco-evolutionary dynamics in experimental and observed biodiversity data. We
364 anticipate a research era in which these tools can be used to address urgent, compelling
365 questions such as what the role of eco-evolutionary dynamics is for large-scale properties
366 such as diversity, stability, and ecosystem functioning.

367

368 **Acknowledgements:**

369 J.H. Pantel acknowledges support from Professorinnenprogramm funded by the German
370 Federal Ministry for Research and Education and the federal states. We acknowledge use of
371 images from The Noun Project (DNA: Andrejs Kirma; genome: Emily Akins; flower 1: Maria
372 Zamchy; flower 2: Tatyana; snake: TkBt; phylogeny: Leslie Coonrod), PhyloPic (butterflies:
373 *Heliconius doris*, *H. melpomene*, *H. erato*; rotifer 1 and 3: *Philodinidae*; rotifer 2: *Rotaria*
374 *macrura*, credit: Diego Fontaneto, Elisabeth A. Herniou, Chiara Boschetti, Manuela Caprioli,
375 Giulio Melone, Claudia Ricci, and Timothy G. Barraclough (vectorized by T. Michael Keeseey),
376 license at <https://creativecommons.org/licenses/by/3.0/> ; bird: *Corvus brachyrhynchos*), and
377 the iCn3D protein viewer (using the ribonuclease H protein, PDB 2RN2,
378 <https://doi.org/10.2210/pdb2RN2/pdb> ; Wang J, Youkharibache P, Marchler-Bauer A,
379 Lanczycki C, Zhang D, Lu S, Madej T, Marchler GH, Cheng T, Chong LC, Zhao S, Yang K,
380 Lin J, Cheng Z, Dunn R, Malkaram SA, Tai C-H, Enoma D, Busby B, Johnson NL, Tabaro F,
381 Song G, Ge Y. iCn3D: From Web-Based 3D Viewer to Structural Analysis Tool in Batch
382 Mode. *Front. Mol. Biosci.* 2022 9:831740.).

383

384 **Supplemental information**

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

388

389 **References:**

- 390 1. Lion, S. (2018) Theoretical approaches in evolutionary ecology: environmental feedback
391 as a unifying perspective. *Am. Nat.* 191, 21–44
- 392 2. Bassar, R.D. et al. (2021) Towards a more precise – and accurate – view of eco-
393 evolution. *Ecol. Lett.* 24, 623–625
- 394 3. Yoshida, T. et al. (2003) Rapid evolution drives ecological dynamics in a predator–prey
395 system. *Nature* 424, 303–306
- 396 4. Bassar, R.D. et al. (2012) Direct and indirect ecosystem effects of evolutionary
397 adaptation in the Trinidadian guppy (*Poecilia reticulata*). *Am. Nat.* 180, 167–185
- 398 5. Pantel, J.H. et al. (2015) Rapid local adaptation mediates zooplankton community
399 assembly in experimental mesocosms. *Ecol. Lett.* 18, 992–1000
- 400 6. Frickel, J. et al. (2017) Rapid evolution of hosts begets species diversity at the cost of
401 intraspecific diversity. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11193–11198
- 402 7. Hairon, N.G., Jr et al. (2005) Rapid evolution and the convergence of ecological and
403 evolutionary time. *Ecol. Lett.* 8, 1114–1127
- 404 8. Ellner, S.P. et al. (2011) Does rapid evolution matter? Measuring the rate of
405 contemporary evolution and its impacts on ecological dynamics. *Ecol. Lett.* 14, 603–

- 407 9. Govaert, L. et al. (2016) Eco-evolutionary partitioning metrics: assessing the importance
408 of ecological and evolutionary contributions to population and community change. *Ecol.*
409 *Lett.* 19, 839–853
- 410 10. Hattich, G.S.I. et al. (2022) Experimentally decomposing phytoplankton community
411 change into ecological and evolutionary contributions. *Funct. Ecol.* 36, 120–132
- 412 11. Cabral, J.S. et al. (2019) Interactions between ecological, evolutionary and
413 environmental processes unveil complex dynamics of insular plant diversity. *J.*
414 *Biogeogr.* 46, ii–iv
- 415 12. Overcast, I. et al. (2021) A unified model of species abundance, genetic diversity, and
416 functional diversity reveals the mechanisms structuring ecological communities. *Mol.*
417 *Ecol. Resour.* 21, 2782–2800
- 418 13. Hagen, O. (2023) Coupling eco-evolutionary mechanisms with deep-time environmental
419 dynamics to understand biodiversity patterns. *Ecography*, e06132
420 <https://doi.org/10.1111/ecog.06132>
- 421 14. Hagen, O. et al. (2021) gen3sis: a general engine for eco-evolutionary simulations of
422 the processes that shape Earth’s biodiversity. *PLoS Biol.* 19, e3001340
- 423 15. Clark, A. et al. (2021) *gauseR*: Lotka–Volterra models for Gause’s ‘Struggle for
424 existence’ R package version 1.1. <https://CRAN.R-project.org/package=gauseR>
- 425 16. Pritchard, D.W. et al. (2017) *FRAIR*: an R package for fitting and comparing consumer
426 functional responses. *Methods Ecol. Evol.* 8, 1528–1534
- 427 17. Santos Baquero, O. and Silveira Marques, F. (2020) *EpiDynamics*: Dynamic Models in
428 Epidemiology. R package version 0.3. 0. [https://CRAN.R-project.org/package=](https://CRAN.R-project.org/package=EpiDynamics)
429 *EpiDynamics*
- 430 18. McGuire, R.M. et al. (2022) *EcoEvoApps*: interactive apps for theoretical models in
431 ecology and evolutionary biology. *Ecol. Evol.* 12, e9556
- 432 19. Cotto, O. et al. (2020) *Nemo-age*: spatially explicit simulations of eco-evolutionary
433 dynamics in stage-structured populations under changing environments. *Methods Ecol.*
434 *Evol.* 11, 1227–1236
- 435 20. Haller, B.C. and Messer, P.W. (2022) *SLiM 4*: Multispecies eco-evolutionary modeling.
436 *Am. Nat.* Published online December 1, 2022. <https://doi.org/10.1086/723601>
- 437 21. Bocedi, G. et al. (2021) *RangeShifter 2.0*: an extended and enhanced platform for
438 modelling spatial eco-evolutionary dynamics and species’ responses to environmental
439 changes. *Ecography* 44, 1453–1462
- 440 22. Retel, C. et al. (2019) The feedback between selection and demography shapes
441 genomic diversity during coevolution. *Sci. Adv.* 5, eaax0530
- 442 23. Barbour, M.A. et al. (2022) A keystone gene underlies the persistence of an
443 experimental food web. *Science* 376, 70–73
- 444 24. Rudman, S.M. et al. (2022) Direct observation of adaptive tracking on ecological time
445 scales in *Drosophila*. *Science* 375, eabj7484

- 446 25. Sakarchi, J. and Germain, R.M. (2023) The evolution of competitive ability. *Am. Nat.*
447 201, 1–15
- 448 26. Luo, S. and Koelle, K. (2013) Navigating the devious course of evolution: the
449 importance of mechanistic models for identifying eco-evolutionary dynamics in nature.
450 *Am. Nat.* 181, S58–S75
- 451 27. Fronhofer, E.A. and Altermatt, F. (2015) Eco-evolutionary feedbacks during
452 experimental range expansions. *Nat. Commun.* 6, 6844
- 453 28. Pillai, P. and Guichard, F. (2012) Competition-colonization trade-offs, competitive
454 uncertainty, and the evolutionary assembly of species. *PLoS One* 7, e33566
- 455 29. Pettersen, A.K. et al. (2020) Metabolic rate, context-dependent selection, and the
456 competition-colonization trade-off. *Evol. Lett.* 4, 333–344
- 457 30. De Meester, L. et al. (2002) The monopolization hypothesis and the dispersal–gene
458 flow paradox in aquatic organisms. *Acta Oecol.* 23, 121–135
- 459 31. Kinnison, M.T. et al. (2015) Cryptic eco-evolutionary dynamics. *Ann. N. Y. Acad. Sci.*
460 1360, 120–144
- 461 32. Kursa, M.B. and Rudnicki, W.R. (2010) Feature selection with the Boruta package. *J.*
462 *Stat. Softw.* 36, 1–13
- 463 33. Overcast, I. et al. (2019) An integrated model of population genetics and community
464 ecology. *J. Biogeogr.* 46, 816–829
- 465 34. Wootton, J.T. and Emmerson, M. (2005) Measurement of interaction strength in nature.
466 *Annu. Rev. Ecol. Evol. Syst.* 36, 419–444
- 467 35. Ellison, A.M. (2004) Bayesian inference in ecology. *Ecol. Lett.* 7, 509–520
- 468 36. Hooten, M.B. and Hobbs, N.T. (2015) A guide to Bayesian model selection for
469 ecologists. *Ecol. Monogr.* 85, 3–28
- 470 37. Beaumont, M.A. (2010) Approximate Bayesian computation in evolution and ecology.
471 *Annu. Rev. Ecol. Evol. Syst.* 41, 379–406
- 472 38. Marjoram, P. et al. (2003) Markov chain Monte Carlo without likelihoods. *Proc. Natl.*
473 *Acad. Sci. U. S. A.* 100, 15324–15328
- 474 39. Hartig, F. et al. (2011) Statistical inference for stochastic simulation models – theory
475 and application. *Ecol. Lett.* 14, 816–827
- 476 40. Newman, K. et al. (2022) State–space models for ecological time-series data: practical
477 model-fitting. *Methods Ecol. Evol.* 14, 26–42
- 478 41. Pichler, M. and Hartig, F. (2023) Machine learning and deep learning – a review for
479 ecologists. *Methods Ecol. Evol.* 14, 994–1016
- 480 42. Lucas, T.C.D. (2020) A translucent box: interpretable machine learning in ecology. *Ecol.*
481 *Monogr.* 90, e01422
- 482 43. Borowiec, M.L. et al. (2022) Deep learning as a tool for ecology and evolution. *Methods*
483 *Ecol. Evol.* 13, 1640–1660

- 484 44. Sunnåker, M. et al. (2013) Approximate Bayesian computation. *PLoS Comput. Biol.* 9,
485 e1002803
- 486 45. Csilléry, K. et al. (2012) ABC: an R package for approximate Bayesian computation
487 (ABC). *Methods Ecol. Evol.* 3, 475–479
- 488 46. Baselga, A. et al. (2022) Joint analysis of species and genetic variation to quantify the
489 role of dispersal and environmental constraints in community turnover. *Ecography*
490 2022, e05808
- 491 47. Pontarp, M. et al. (2019) Inferring community assembly processes from macroscopic
492 patterns using dynamic eco-evolutionary models and Approximate Bayesian
493 Computation (ABC). *Methods Ecol. Evol.* 10, 450–460
- 494 48. Gelman, A. et al. (1996) Posterior predictive assessment of model fitness via realized
495 discrepancies. *Stat. Sin.* 6, 733–760
- 496 49. Conn, P.B. et al. (2018) A guide to Bayesian model checking for ecologists. *Ecol.*
497 *Monogr.* 88, 526–542
- 498 50. Jabot, F. and Chave, J. (2009) Inferring the parameters of the neutral theory of
499 biodiversity using phylogenetic information and implications for tropical forests. *Ecol.*
500 *Lett.* 12, 239–248
- 501 51. Blum, M.G.B. et al. (2013) A comparative review of dimension reduction methods in
502 approximate Bayesian computation. *SSO Schweiz. Monatsschr. Zahnheilkd.* 28, 189–
503 208
- 504 52. Johnson, J.B. and Omland, K.S. (2004) Model selection in ecology and evolution.
505 *Trends Ecol. Evol.* 19, 101–108
- 506 53. Dormann, C.F. et al. (2018) Model averaging in ecology: a review of Bayesian,
507 information-theoretic, and tactical approaches for predictive inference. *Ecol. Monogr.*
508 88, 485–504
- 509 54. Eyring, V. et al. (2019) Taking climate model evaluation to the next level. *Nat. Clim.*
510 *Chang.* 9, 102–110
- 511 55. McIntire, E.J.B. et al. (2022) PERFICT: a re-imagined foundation for predictive ecology.
512 *Ecol. Lett.* 25, 1345–1351
- 513 56. Tilman, D. (1999) The ecological consequences of changes in biodiversity: a search for
514 general principles. *Ecology* 80, 1455–1474
- 515 57. Luiselli, J. et al. (2022) Detecting the ecological footprint of selection. *bioRxiv* Published
516 online December 10, 2022. <https://doi.org/10.1101/2021.05.11.442553>
- 517 58. Fussmann, G.F. et al. (2007) Eco-evolutionary dynamics of communities and
518 ecosystems. *Funct. Ecol.* 21, 465–477
- 519 59. Schoener, T.W. (2011) The newest synthesis: understanding the interplay of
520 evolutionary and ecological dynamics. *Science* 331, 426–429
- 521 60. De Meester, L. and Pantel, J. (2014) Eco-evolutionary dynamics in freshwater systems.
522 *J. Limnol.* 73, 193–200
- 523 61. De Meester, L. et al. (2019) Analysing eco-evolutionary dynamics – the challenging

- 524 complexity of the real world. *Funct. Ecol.* 33, 43–59
- 525 62. Sokal, R.R. and Riska, B. (1981) Geographic variation in *Pemphigus populitransversus*
526 (Insecta: Aphididae). *Biol. J. Linn. Soc. Lond.* 15, 201–233
- 527 63. Hutchison, D.W. and Templeton, A.R. (1999) Correlation of pairwise genetic and
528 geographic distance measures: inferring the relative influences of gene flow and drift on
529 the distribution of genetic variability. *Evolution* 53, 1898–1914
- 530 64. Haller, B.C. and Messer, P.W. (2019) SLiM 3: forward genetic simulations beyond the
531 Wright–Fisher model. *Mol. Biol. Evol.* 36, 632–637
- 532 65. Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by
533 sampling trees. *BMC Evol. Biol.* 7, 214
- 534 66. Liu, L. (2008) BEST: Bayesian estimation of species trees under the coalescent model.
535 *Bioinformatics* 24, 2542–2543
- 536 67. McPeck, M.A. (2008) The ecological dynamics of clade diversification and community
537 assembly. *Am. Nat.* 172, E270–E284
- 538 68. Vellend, M. (2010) Conceptual synthesis in community ecology. *Q. Rev. Biol.* 85, 183–
539 206
- 540 69. Vellend, M. (2016) *The Theory of Ecological Communities*, Monographs in Population
541 Biology Vol. 57. Princeton University Press
- 542 70. Michener, W.K. and Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-
543 intensive science. *Trends Ecol. Evol.* 27, 85–93
- 544 71. Keitt, T.H. and Abelson, E.S. (2021) Ecology in the age of automation. *Science* 373,
545 858–859
- 546 72. Besson, M. et al. (2022) Towards the fully automated monitoring of ecological
547 communities. *Ecol. Lett.* 25, 2753–2775
- 548 73. Lowndes, J.S.S. et al. (2017) Our path to better science in less time using open data
549 science tools. *Nat. Ecol. Evol.* 1, 160
- 550 74. Powers, S.M. and Hampton, S.E. (2019) Open science, reproducibility, and
551 transparency in ecology. *Ecol. Appl.* 29, e01822
- 552 75. Wieczynski, D.J. et al. (2021) Linking species traits and demography to explain complex
553 temperature responses across levels of organization. *Proc. Natl. Acad. Sci. U. S. A.*
554 118, e2104863118
- 555 76. Gibert, J.P. et al. (2023) Rapid eco-phenotypic feedback and the temperature response
556 of biomass dynamics. *Ecol. Evol.* 13, e9685
- 557 77. Fulgione, A. et al. (2022) Parallel reduction in flowering time from de novo mutations
558 enable evolutionary rescue in colonizing lineages. *Nat. Commun.* 13, 1461
- 559 78. Gienapp, P. et al. (2017) Genomic quantitative genetics to study evolution in the wild.
560 *Trends Ecol. Evol.* 32, 897–908
- 561 79. Dudaniec, R.Y. et al. (2018) Signatures of local adaptation along environmental
562 gradients in a range-expanding damselfly (*Ischnura elegans*). *Mol. Ecol.* 27, 2576–2593

- 563 80. Pfenninger, M. and Foucault, Q. (2022) Population genomic time series data of a
564 natural population suggests adaptive tracking of fluctuating environmental changes.
565 *Integr. Comp. Biol.* 62, 1812–1826
- 566 81. Rudman, S.M. et al. (2018) What genomic data can reveal about eco-evolutionary
567 dynamics. *Nat. Ecol. Evol.* 2, 9–15
- 568 82. Davies, N. et al. (2014) The founding charter of the Genomic Observatories Network.
569 *Gigascience* 3, 2
- 570 83. Emerson, B.C. et al. (2022) Collective and harmonized high throughput barcoding of
571 insular arthropod biodiversity: toward a Genomic Observatories Network for island. *Mol.*
572 *Ecol.* Published online September 6, 2022, [https://doi.org/10.](https://doi.org/10.22541/au.164840312.23663001)
573 [22541/au.164840312.23663001](https://doi.org/10.22541/au.164840312.23663001)
- 574 84. Shafer, A.B.A. et al. (2016) Forecasting ecological genomics: high-tech animal
575 instrumentation meets high-throughput sequencing. *PLoS Biol.* 14, e1002350
- 576 85. Deiner, K. et al. (2017) Environmental DNA metabarcoding: transforming how we
577 survey animal and plant communities. *Mol. Ecol.* 26, 5872–5895
- 578 86. Ellegaard, M. et al. (2020) Dead or alive: sediment DNA archives as tools for tracking
579 aquatic evolution and adaptation. *Commun. Biol.* 3, 169
- 580 87. Waldvogel, A. et al. (2020) Evolutionary genomics can improve prediction of species'
581 responses to climate change. *Evol. Lett.* 4, 4–18
- 582 88. Urban, M.C. et al. (2021) Coding for life: designing a platform for projecting and
583 protecting global biodiversity. *Bioscience* 72, 91–104
- 584 89. Lasky, J.R. et al. (2020) What processes must we understand to forecast regional-scale
585 population dynamics? *Proc. Biol. Sci.* 287, 20202219
- 586 90. Benito Garzón, M. et al. (2019) Δ TraitSDMs: species distribution models that account
587 for local adaptation and phenotypic plasticity. *New Phytol.* 222, 1757–1765
- 588 91. Ovaskainen, O. et al. (2017) How to make more out of community data? A conceptual
589 framework and its implementation as models and software. *Ecol. Lett.* 20, 561–576
- 590 92. Tikhonov, G. et al. (2020) Joint species distribution modelling with the R-package
591 *Hmsc*. *Methods Ecol. Evol.* 11, 442–447
- 592 93. Pollock, L.J. et al. (2014) Understanding co-occurrence by modelling species
593 simultaneously with a Joint Species Distribution Model (JSDM). *Methods Ecol. Evol.* 5,
594 397–406
- 595 94. Pichler, M. and Hartig, F. (2021) A new joint species distribution model for faster and
596 more accurate inference of species associations from big community data. *Methods*
597 *Ecol. Evol.* 12, 2159–2173
- 598 95. Devarajan, K. et al. (2020) Multi-species occupancy models: review, roadmap, and
599 recommendations. *Ecography* 43, 1612–1624
- 600 96. Doser, J.W. et al. (2022) *spOccupancy*: an R package for single-species, multi-species,
601 and integrated spatial occupancy models. *Methods Ecol. Evol.* 13, 1670–1678
- 602 97. Boussange, V. and Pellissier, L. (2022) Eco-evolutionary model on spatial graphs

603 reveals how habitat structure affects phenotypic differentiation. *Commun. Biol.* 5, 668
604 98. Govaert, L. et al. (2021) Measuring the contribution of evolution to community trait
605 structure in freshwater zooplankton. *Oikos* 130, 1773–1787
606 99. Pantel, J.H. et al. (2017) 14 questions for invasion in ecological networks. *Adv. Ecol.*
607 *Res.* 56, 293–340
608 100. Colbourne, J.K. et al. (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*
609 331, 555–561
610 101. Moody, E.K. et al. (2022) Eutrophication-driven eco- evolutionary dynamics indicated
611 by differences in stoichiometric traits among populations of *Daphnia pulicaria*. *Freshw.*
612 *Biol.* 67, 353–364
613 102. Gomulkiewicz, R. and Holt, R.D. (1995) When does evolution by natural selection
614 prevent extinction? *Evolution* 49, 201–207
615 103. Hermann, R.J. and Becks, L. (2022) Change in prey genotype frequency rescues
616 predator from extinction. *R. Soc. Open Sci.* 9, 220211
617 104. Farkas, T.E. et al. (2013) Evolution of camouflage drives rapid ecological change in an
618 insect community. *Curr. Biol.* 23, 1835–1843
619 105. Rodríguez-Verdugo, A. et al. (2017) The genomic basis of eco-evolutionary dynamics.
620 *Mol. Ecol.* 26, 1456–1464
621 106. Svensson, E.I. (2019) Eco-evolutionary dynamics of sexual selection and sexual
622 conflict. *Funct. Ecol.* 33, 60–72
623 107. Connolly, S.R. et al. (2017) Process, mechanism, and modeling in macroecology.
624 *Trends Ecol. Evol.* 32, 835–844
625 108. Gallien, L. et al. (2017) The effects of intransitive competition on coexistence. *Ecol.*
626 *Lett.* 20, 791–800
627
628
629
630
631
632
633
634
635
636
637

638 **Glossary:**

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

643

644 **Boruta:** A feature selection algorithm that uses a random forest classification algorithm to
645 identify predictive features that are more informative than randomly generated features

646

647 **Coefficient of variation (CV):** The ratio of standard deviation to mean

648

649 **Cross validation:** A model validation technique to assess how a statistical model or analysis
650 will generalize to an independent data set, where a part of the data is withheld from training
651 the model to assess its ability to classify the withheld data

652

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

659

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

664

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

668

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

673

674 **Posterior distribution:** Refers to a posterior probability distribution, or the updated
675 probability distribution for a model parameter after taking into account information (new
676 information in the form of data)

677

678 **Prior distribution:** Refers to a prior probability distribution, or the expression of an uncertain
679 distribution for the probability of different values of a model parameter, which captures prior
680 expectations for the shape or structure of the parameter

681

682 **Process-based mechanistic model:** A mechanistic model is a characterization of the state
683 of a system as explicit functions of component parts and their associated actions and

684 interactions, and a process-based model is a model that characterizes changes in a system's
685 state as explicit functions of the events that drive those state changes [106]

686

687 **Summary statistic:** A descriptive statistic calculated from data that gives compressed,
688 summarized information, used to reduce data dimensionality (though information in the
689 original data can be lost); may be referred to as 'features' in machine learning applications

690

691

692 **Outstanding questions:**

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

696

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

700

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

703

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

706

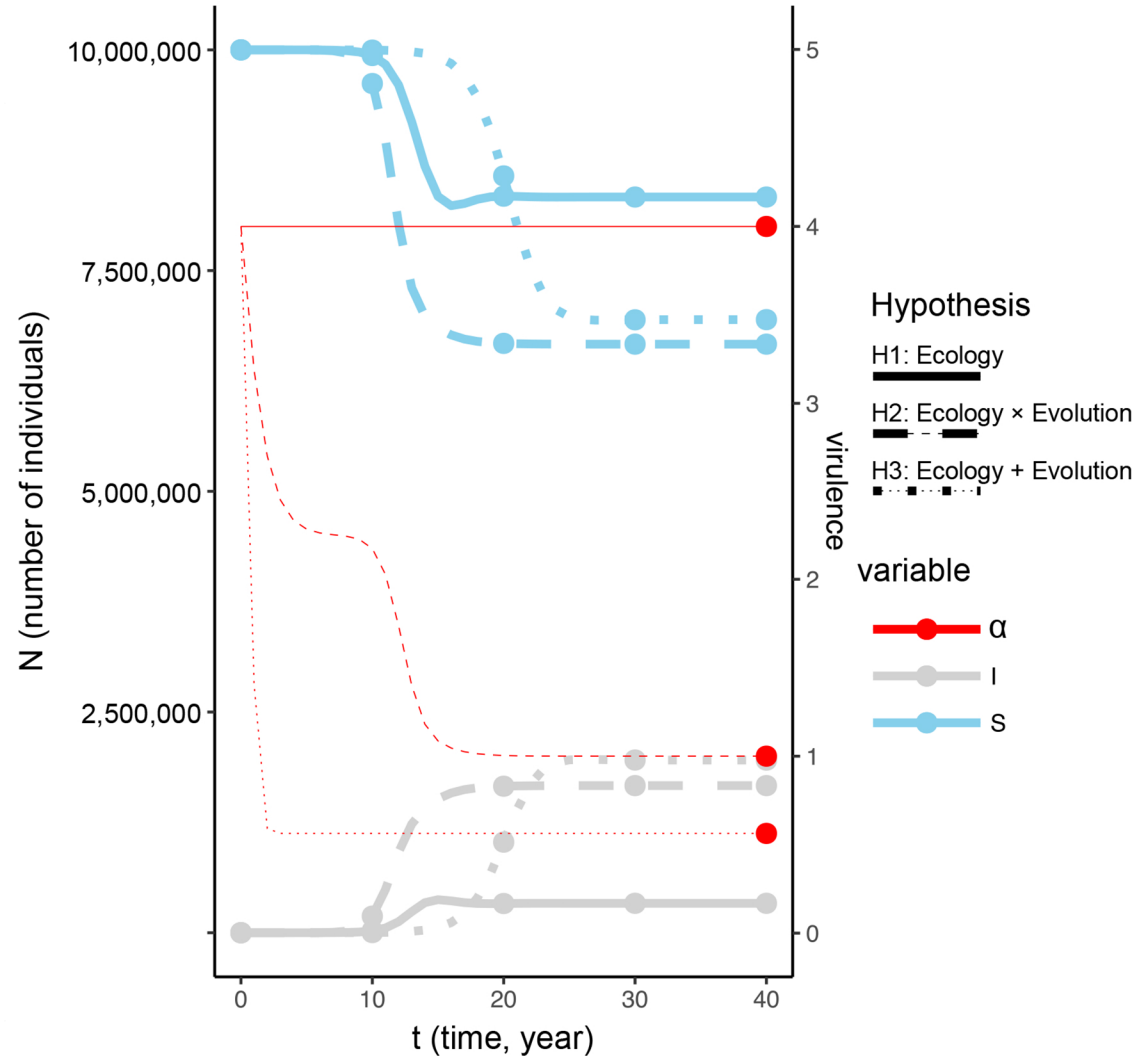
707 **Tables & Figures:**

708 **Figure 1.** Eco-evolutionary hypotheses for the structure of observed disease and biodiversity
709 dynamics. (A) Simulations of an eco-evolutionary epidemiological model with virulence
710 evolution [26] are shown with many of the same model parameters (e.g., $S_0 = 10^7$, $\alpha_0 = 4$, γ
711 $= 2$; see supplemental material S1A for additional parameter values), but under three
712 alternative hypotheses that can be compared to observed data. Population sizes of
713 susceptible (S , blue) and infected (I , gray) individuals (left y-axis), and pathogen virulence (α ,
714 red, right y-axis), are given for: hypothesis 1 (H1, unbroken line) in an ecological-only model
715 where pathogen traits (α) cannot evolve, hypothesis 2 (H2, broken line) in an eco-evolutionary
716 model where pathogen virulence can evolve and impacts host population size, and
717 hypothesis 3 (H3, dotted line) in an ecological and evolutionary model where virulence
718 evolves independently from host population size. Points show a coarser time series that could
719 potentially be used for comparing observed to model-simulated data (red points are the final
720 values for pathogen virulence), as fine time series with records of all population and trait
721 dynamics are difficult to obtain. These comparisons of observed to model-generated data
722 under alternative hypotheses (**Box 2**) can help identify whether trait evolution and/or eco-
723 evolutionary feedbacks structured the observed data. (B–E) Simulations of speciation, trait
724 evolution, and community assembly using the *gen3sis* eco-evolutionary simulation engine
725 [14], for a hypothetical community of ten ancestral species, randomly placed as a single initial
726 population on the map of South America (an 81 x 71 grid with cell-specific temperature values
727 that impact carrying capacity and also population growth via the distance between a
728 population's trait value and the local optimum). Movement is via a cost function based on the
729 distance between sites and a species-specific dispersal value, and evolution and speciation

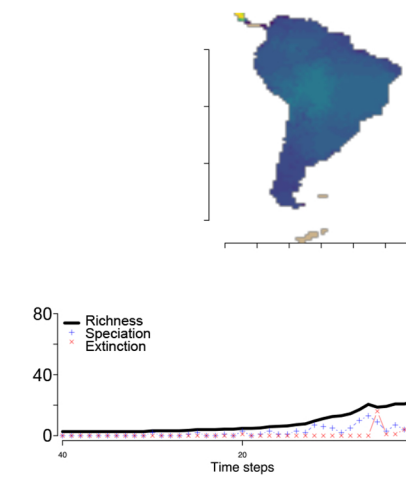
730 occur across the dynamic landscape over 40 million years (see supplemental material S1B
731 for additional parameter values). We run simulations under four alternative eco-evolutionary
732 hypotheses: (B) hypothesis 1 with evolution in a single temperature-related trait that
733 determines population growth, (C) hypothesis 2 with evolution proceeding at the same rate
734 in the temperature trait and in an additional dispersal-related trait that impacts the scale of
735 the dispersal kernel, (D) hypothesis 3 with evolution in these two traits that proceeds at
736 different rates, and (E) hypothesis 4 with evolution in the two traits at differing rates, where
737 the two traits are correlated. For all four hypotheses, the final alpha diversity (local species
738 richness, color-scale per grid cell in map) and time series of some summary statistics (number
739 of speciation and extinction events, total regional richness) are shown. These summary
740 statistics could be compared to observed values to distinguish among alternative hypotheses
741 that may have structured observed data (**Box 2**).

742

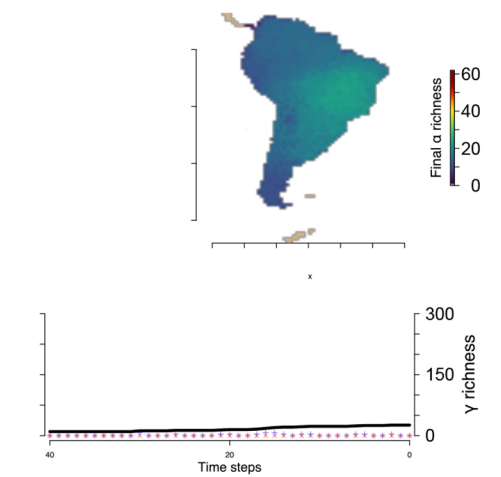
(A) Host-pathogen eco-evolutionary model



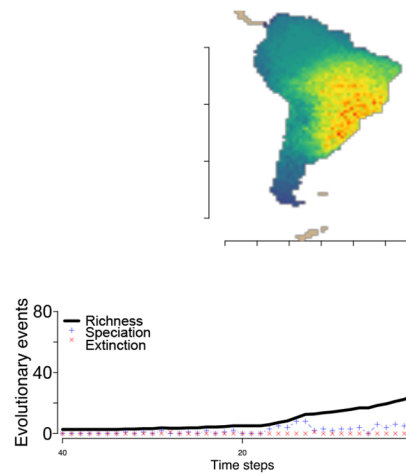
(B) Hypothesis 1: 1 trait evolves



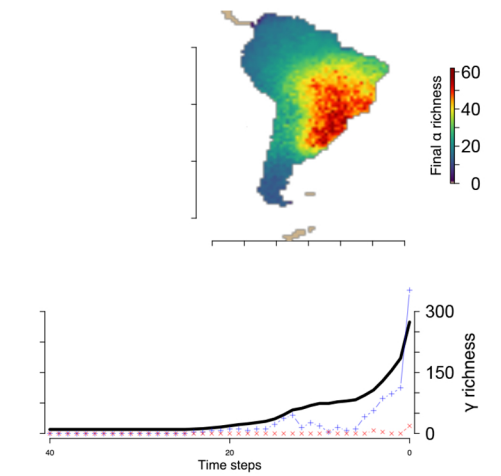
(C) Hypothesis 2: 2 traits evolve at same rate



(D) Hypothesis 3: trait 1 evolves slower than trait 2



(E) Hypothesis 4: trait 1 and trait 2 correlated



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

744 **1. Identify research question and data**

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

748 (ii) Identify relevant summary statistics that are emergent properties of the data: for example,
749 nucleotide diversity, within- and among-population genetic variation, species diversity (alpha,
750 beta, gamma), center and width of phenotypic distributions, phylogenetic structure, total
751 community abundance, rank abundance curves, normalized lineage-through-time
752 diversification curves [14], variation in range sizes, summary statistics across three data axes
753 (species abundances, population genetic variation, trait values [12]). Choosing the most
754 informative summary statistics emerges from the research question and available data (e.g.,
755 total abundance data at one time point is unlikely to produce informative model comparisons).
756 Summary statistics should be assessed before data collection (steps 2–4) and inform the
757 type of data needed to capture the signature of particular eco-evolutionary dynamics.
758 Summary statistics can come from time series or individual time points, and one should
759 consider diverse summary statistics across different scales and biological levels to address
760 the problem of equifinality [11].

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

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

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

769 **3. Generate alternative hypotheses for processes that structure observed data**

770 (i) Formulate null and alternative hypotheses expressed as competing models differing in
771 ecological, evolutionary, and coupled eco-evolutionary processes. Model alternatives may
772 include varying forms of ecological (e.g., neutral or niche-based species interactions),
773 evolutionary (e.g., correlated traits, neutral versus adaptive evolution), or eco-evolutionary
774 feedback (e.g., density-dependent trait evolution, dispersal-driven maladaptation impacting
775 local fitness) relationships (**Figure I**).

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

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

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

- 781 • Least squares
- 782 • Markov Chain Monte Carlo (MCMC)
- 783 • Machine learning (ML)
- 784 • Approximate Bayesian computation (ABC)

785 (ii) Evaluate summary statistics in simulated data for sufficiency to discriminate among
786 alternative hypotheses and their associated distinct eco-evolutionary mechanisms (i.e., test
787 classification ability using training and test data, revisit summary statistics to maximize
788 posterior probabilities from data simulated under known conditions); note potential summary
789 statistics that are emergent signatures of underlying eco-evolutionary dynamics (e.g., shifts
790 in phase of predator–prey cycles [3]); see step 1(ii).

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

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

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

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

799 (i) Consider whether to use the model with the highest posterior probability support or to
800 average over the models with weights based on their relative posterior support [52,53].

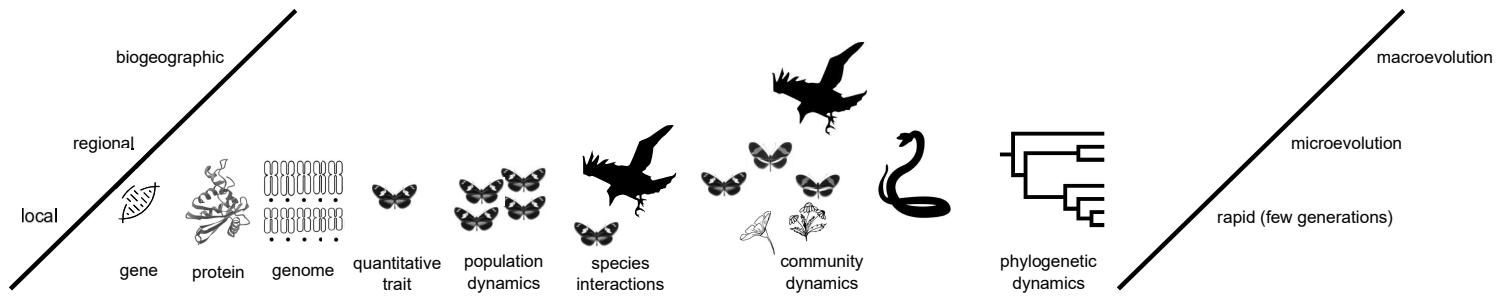
801 (ii) Use posterior predictive checks to evaluate model fit [48,49].

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

805

806 **Box 1, Figure I.** Summary of simulation models of eco-evolutionary dynamics and their
807 relevant scales, which can be used to simulate the processes that structure observed
808 biodiversity data. Models are given along with the modeled spatial, temporal, and biodiversity
809 scales. Filled symbols indicate whether a biodiversity level is modeled, and open symbols
810 indicate some special considerations needed to apply the model to that scale (corresponding
811 numbers are explained in online supplemental material S1E). We use +, −, and 0 to indicate
812 positive, negative, or neutral interactions between species considered in a model (and a filled
813 circle for species interactions means that all types are considered).

814



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://messengerlab.org/slim

815 **Box 2.** Five-step process of eco-evolutionary hypothesis testing with ABC

816 **Candidate model development**

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

825

826 **Simulations**

827 The first step is to come up with a model that could realistically produce the observed data
828 under alternative hypotheses (**Figure 1B**). We use a Beverton–Holt model with evolution in
829 a heritable trait x that impacts population growth. We compare hypothesis 1 (H_1) without
830 evolution ($h^2 = 0$ for all species) and hypothesis 2 (H_2) with evolution ($h^2 \geq 0$ for all species)
831 to our observed data (**Figure IA**, Case 2). We ran 100,000 simulations under the two
832 alternative models. For all unknown or unmeasured parameters, a given simulation uses a
833 random draw from a **prior distribution**, chosen to be uninformative (**Figure IB**; i.e., for model
834 H_2 , heritability is drawn from a random uniform distribution between 0 and 1), informed by
835 prior knowledge about the system (i.e., interaction coefficients are drawn from a beta
836 distribution with increased likelihood of weak interactions ($\alpha \sim \text{beta}(0.25, 10)$ [34]), or using
837 some value generator [14]. For the random draws from the prior distribution of w (width of the
838 Gaussian fitness function; example draws are shown as orange points), the associated
839 fitness functions (the relationship between fitness, y -axis, and trait value, x -axis) are shown.

840

841 **Model selection**

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

852

853 **Parameter estimation**

854 Once a hypothesized model is identified (H_2 with evolution, 81% posterior probability), the
855 randomly drawn parameter values that produced accepted simulations can estimate the
856 posterior distributions of model parameters [37] (**Figure ID**). Here, the ABC process can
857 successfully detect that species 3 (orange) is unlikely to be evolving as the most probable
858 values of h^2 are centered around 0, but it cannot clearly resolve the true system state that
859 $h^2_{1,3} = 0$ and $h^2_2 = 0.25$ (species 2: blue, species 3: gray). We also show posterior distributions
860 for the strength of selection w (prior: dashed blue line, true value: dashed black line, posterior
861 distribution: solid black line) and interaction coefficients (α_{ij} and α_{ji}).

862

863 **Posterior predictions**

864 The quality of the selected model (H_2 , with evolution) is assessed by comparing the data
865 (**Figure IA**, Case 2) to predictions made under the accepted model and associated parameter
866 posterior distributions. The 95% confidence intervals (**Figure IE**, blue broken lines) were
867 obtained by running simulations using the parameter values from a subset of accepted
868 simulations with lowest Euclidean distance to the observed values of summary statistics.

869

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

