

## Uncovering Cryptic Coevolution\*

Scott L. Nuismer,<sup>1,†</sup> Bob Week,<sup>2</sup> and Luke J. Harmon<sup>1</sup>

1. Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844; 2. Department of Integrative Biology, Michigan State University, East Lansing, Michigan 48824

Submitted March 8, 2021; Accepted August 12, 2021; Electronically published April 19, 2022

**ABSTRACT:** Studies of coevolution in the wild have largely focused on reciprocally specialized species pairs with striking and exaggerated phenotypes. Textbook examples include interactions between toxic newts and their garter snake predators, long-tongued flies and the flowers they pollinate, and weevils with elongated rostra used to bore through the defensive pericarp of their host plants. Although these studies have laid a foundation for understanding coevolution in the wild, they have also contributed to the widespread impression that coevolution is a rare and quirky sideshow to the day-to-day grind of ecology and evolution. In this perspective, we argue that the focus of coevolution has been biased toward the obvious and ignored the cryptic. We have focused on the obvious—studies of reciprocally specialized species pairs with exaggerated phenotypes—mainly because we have lacked the statistical tools required to study coevolution in more generalized and phenotypically mundane systems. Building from well-established coevolutionary theory, we illustrate how model-based approaches can be used to remove this barrier and begin estimating the strength of coevolutionary selection indirectly using routinely collected data, thus uncovering cryptic coevolution in more typical communities. By allowing the distribution of coevolutionary selection to be estimated across genomes, phylogenies, and communities and over deep timescales, these novel approaches have the potential to revolutionize the way we study coevolution. As we develop a road map to these next-generation approaches, we highlight recent studies making notable progress in this direction.

**Keywords:** species interactions, comparative methods, ecological networks, community structure.

### Introduction

Reciprocal evolutionary change between interacting species—coevolution—has long been a guiding principle for understanding how traits of interacting species evolve and

how species interactions themselves diversify (Ehrlich and Raven 1964; Janzen 1980; Berenbaum et al. 1986; Thompson 1994; Agrawal et al. 2012). The past several decades have seen an explosion of coevolutionary theory, largely driven by mathematical and computational advances that make studying coevolution in complex genetic, phenotypic, and ecological systems feasible (Dieckmann and Law 1996; Otto and Nuismer 2004; Dercole et al. 2006; Gandon and Otto 2007; M'Gonigle et al. 2009; Nuismer et al. 2013; Debarre et al. 2014; Nuismer 2017; Week et al. 2021). The resulting models have shed new light on long-standing questions in ecology and evolutionary biology and identified myriad ways coevolution influences the structure, function, and diversification of biological systems. For instance, analysis of multilocus population genetic models has refined our understanding of the conditions under which coevolution favors sexual reproduction or increased mutation rates (Otto and Nuismer 2004; M'Gonigle et al. 2009). Quantitative genetic and adaptive dynamics methods have clarified when coevolution favors increased phenotypic and taxonomic diversification (Doebeli and Dieckmann 2000; Yoder and Nuismer 2010). Furthermore, new moment-based approaches and network-based methods have illuminated how coevolution influences the structure, function, and stability of ecological networks and communities (Nuismer et al. 2013, 2018; Guimaraes et al. 2017). Together, these new methods have revolutionized coevolutionary modeling by allowing us to move well beyond localized pairwise interactions mediated by a single genetic locus or quantitative trait.

As the complexity and scope of mathematical and computational models of coevolution have raced ahead, the development of statistical approaches for estimating the strength and form of coevolutionary selection in natural populations has fallen woefully behind. Coevolutionary selection is a product of genotype  $\times$  genotype interactions for fitness and is the central process responsible for coevolution and reciprocal adaptation. As a consequence, empiricists have few tools

\* This contribution is part of a group of Focused Topic articles that was organized by Paulo R. Guimaraes Jr. and Ana Paula A. Assis.

† Corresponding author; email: snuismer@uidaho.edu.

ORCIDs: Nuismer, <https://orcid.org/0000-0001-9817-0056>; Week, <https://orcid.org/0000-0001-7687-4757>.

at their disposal if they wish to quantify the importance of coevolutionary interactions among more than a single, isolated pair of interacting species. Even for a single, isolated species pair, rigorously demonstrating coevolutionary selection is a formidable task, requiring knowledge of individual genotypes or phenotypes and the fitness consequences of encounters between individuals (Brodie and Ridenhour 2003; Ridenhour 2005; Heath and Nuismer 2014; MacPherson et al. 2018). Given this, it is little surprise that the best-studied cases of coevolution involve rather unusual systems with highly exaggerated traits or extremely tight reciprocal specialization (Zangerl and Berenbaum 2003; Toju and Sota 2006; Hanifin et al. 2008; Pauw et al. 2009; King et al. 2011). The unfortunate consequence of focusing on these unusual or highly specialized biological systems, however, is the perception that coevolution is a rare and quirky sideshow to the main evolutionary acts shaping the diversity of life on a day-to-day basis.

Reconciling the diverse consequences of coevolution demonstrated by emerging theory with its perceived rarity requires the development of new statistical approaches that allow the strength of coevolutionary selection to be estimated more broadly. By coupling coevolutionary models with data that can be collected rapidly and efficiently from diverse systems, these new approaches will allow us to estimate the strength of cryptic coevolutionary selection in systems where exaggerated traits and reciprocal specialization are not pronounced. The ultimate goal of these new methods should be to move from estimating the strength of coevolutionary selection between a pair of genes, traits, or species to estimating the distribution of coevolutionary selection within entire genomes, phenomes, and communities. Here, we outline recent progress that has been made in this direction by moving away from the direct estimation of coevolutionary selection using conventional statistical approaches and toward model-based statistical approaches that estimate coevolutionary selection even when cryptic.

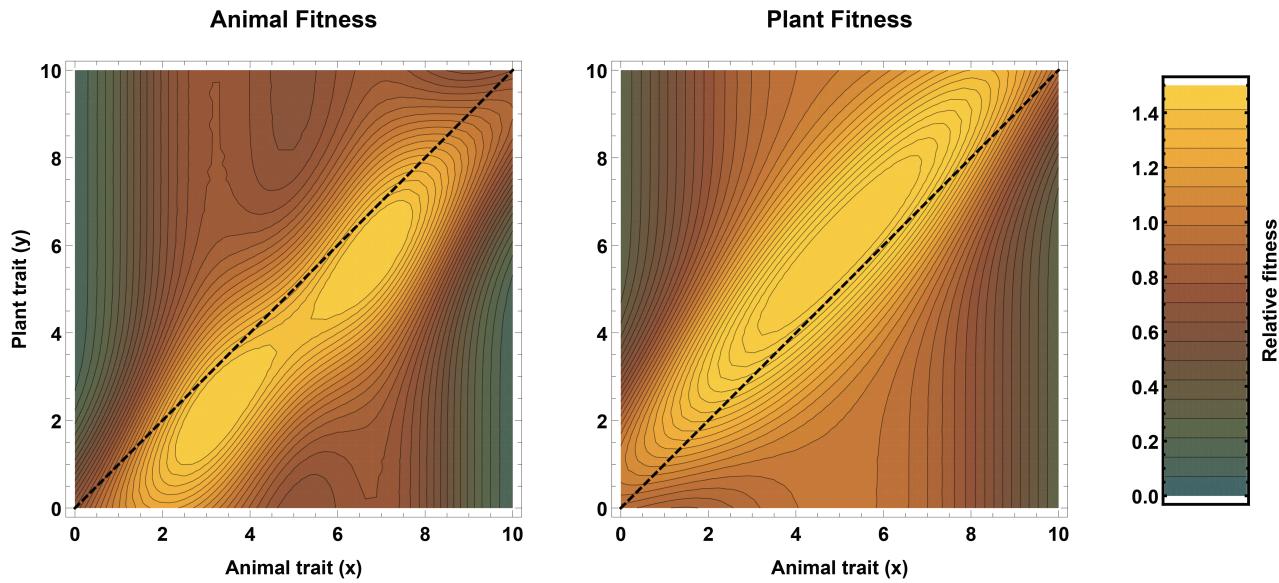
#### **Direct Measurement of Pairwise Coevolution Is Hard**

The distinguishing feature of coevolution is reciprocity, and it is that reciprocity that makes it hard to measure in natural systems. Two species A and B coevolve when evolutionary change in species A alters the pattern of selection experienced by species B, which in turn leads to an evolutionary change in species B that modifies selection acting on species A (Janzen 1980). Establishing the existence of this reciprocal evolutionary feedback between species, and understanding how it influences ecological and evolutionary change, is the essential challenge of coevolutionary biology. However, the importance of coevolution cannot generally be evaluated using conventional methods, such as estimat-

ing selection gradients, comparing divergence of quantitative traits and neutral molecular markers ( $Q_{ST}$ - $F_{ST}$ ), or comparing rates of synonymous and nonsynonymous substitution, all of which deal with the effects of selection on single species. Instead, demonstrating coevolution requires methods that explicitly estimate the strength of the genotype  $\times$  genotype or phenotype  $\times$  phenotype interactions for fitness that underpin reciprocal selection (fig. 1).

Although a few methods now exist to formally estimate the strength of reciprocal selection from phenotypic data (Ridenhour 2005) or to scan genomes for genotype  $\times$  genotype interactions (MacPherson et al. 2018), these methods have not yet been broadly applied. The reason these methods have not been widely adopted is largely their stringent data requirements: highly replicated measurement of phenotypes and/or genotypes for interacting pairs of individuals and the consequences of the interaction for each individual's lifetime fitness.

A consequence of the challenges associated with measuring coevolution directly in the wild then is an overreliance on the obvious. Much of what we know comes from well-studied systems with unique natural history that allows genotype  $\times$  genotype or phenotype  $\times$  phenotype interactions for fitness to be demonstrated using clever and labor-intensive experimental manipulations (e.g., Zangerl and Berenbaum 2003; Toju and Sota 2006; Pauw et al. 2009). For example, coevolution has been established as an important mechanism driving the escalation of offensive and defensive traits in the interaction between the Japanese camellia, *Camellia japonica*, and its seed predatory weevil, *Curculio camelliae* (Toju and Sota 2006). Demonstrating the potential for coevolution in this system is facilitated by the existence of wildly exaggerated traits with a clear functional link to the fitness consequences of interaction (rostrum and pericarp) and the feasibility of executing pairwise performance trials in the laboratory that can be used to establish phenotype  $\times$  phenotype interactions for fitness. Similarly, coevolution has been implicated in the interactions between the long-tongued fly, *Moegistorhynchus longirostris*, and the plant it pollinates, *Lapeirousia anceps*, using elegant experimental manipulations that allow phenotype  $\times$  phenotype interactions for fitness to be established (Pauw et al. 2009). Here too, establishing the potential for coevolution is facilitated by the existence of exaggerated traits with clear functional links to fitness. Even in these wonderfully amenable systems, however, direct estimation of coevolutionary selection in the field has remained elusive. The reason is that estimating reciprocal (coevolutionary) selection directly requires measuring the traits of naturally interacting pairs of individuals and at least a surrogate of their lifetime fitness. This same scenario plays out over and over again, subtly and implicitly shifting our perception of coevolution through ascertainment bias. We now collectively, even if subconsciously, expect coevolution



**Figure 1:** Hypothetical phenotype  $\times$  phenotype interactions for fitness in a pollinating animal (*left*) and the plant it pollinates (*right*). Animal fitness is maximized when the animal has a slightly larger trait value than the plant it pollinates, as might be the case when nectar extraction is maximized by having a beak slightly longer than the depth of the plant's corolla. Plant fitness is also maximized by having a slightly larger trait value than the animal pollinating it, as might be the case if pollen transfer is maximized when a longer corolla requires more sustained pollinator effort to reach the nectar reward. The result is a phenotype  $\times$  phenotype interaction where the fitness of each individual depends not only on its own trait value but also on the trait value of individuals with which it interacts.

to be most important in that unusual subset of interactions mediated by flashy and exaggerated traits. At the same time, we may be ignoring countless examples of cryptic coevolution where the signature of the coevolutionary process is less pronounced (Thompson 1994, 2005a).

#### Indirect Methods Expand the Scope of Coevolutionary Studies

Advances in mathematical and computational statistics are beginning to open the door to new methods that allow the strength of coevolutionary selection to be estimated by coupling mechanistic models of coevolution to more readily collected types of data (table 1). One advantage shared by many of these methods is that they allow coevolutionary selection to be quantified indirectly without requiring direct measurements of the effect of one species on the fitness of another. These methods promise to broaden our perception of coevolution's generality and taxonomic distribution and greatly increase the scale of coevolutionary studies. By avoiding the need to directly estimate genotype  $\times$  genotype or phenotype  $\times$  phenotype interactions for fitness in natural populations, these new methods can be scaled up to study the distribution of coevolution across multiple populations, entire genomes, or clades of interacting species (fig. 2).

In the sections that follow, we provide a brief overview of some important progress in this direction, highlighting new

methods for estimating the strength of coevolutionary selection indirectly using spatially replicated genomic data, phenotypic data, data from metacommunities of interacting species, and data over deep timescales from phylogenetic trees.

#### *Identifying the Intensity and Distribution of Coevolutionary Selection across Genomes*

Advances in sequencing technology are making it increasingly feasible to get high-resolution marker data from multiple individuals across multiple populations. These new technologies have brought to life the possibility of identifying loci involved in coevolution by scanning for genotype  $\times$  genotype interactions for fitness using coevolutionary genome-wide association studies (co-GWAS; MacPherson et al. 2018; Wang et al. 2018; Ebert and Fields 2020). Co-GWAS extend classical genome-wide association studies to pairs of interacting species. In brief, they scan the genomes of the interacting species for genetic markers in one species that interact statistically with genetic markers in the other species to influence the outcome of the interaction (e.g., infection or resistance). Unfortunately, as with other direct approaches for investigating coevolution, this requires information on the outcome or fitness consequences of the interaction for both members of an interacting species pair. Thus, even though it is now feasible to generate genomic sequence data for a large number of replicate pairs of interacting individuals, co-GWAS

**Table 1:** Indirect, model-based approaches for inferring coevolution

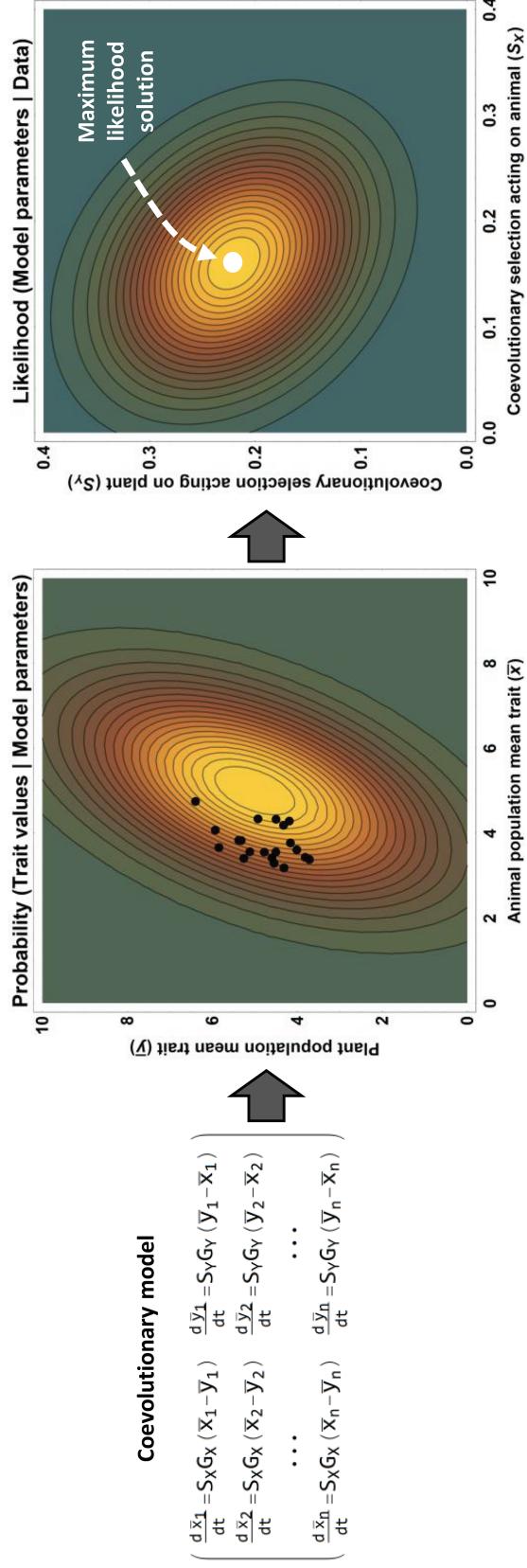
Modeling approach	Scale	Statistical method	Data required	Limitation(s)	Reference
Population genetic simulations	Metapopulation	Approximate Bayesian computation	Genetic marker frequencies from a pair of interacting species across multiple populations	Assumes populations are independent	Maerkle and Tellier 2020
Mathematical quantitative genetics	Metapopulation	Maximum likelihood	Population mean phenotypes for a pair of interacting species across multiple populations	Assumes weak coevolution; assumes populations are independent; assumes fixed genetic variance	Week and Nuismer 2019
Quantitative genetic simulations	Metapopulation	Approximate Bayesian computation	Population mean phenotypes for a pair of interacting species across multiple populations	Assumes fixed genetic variance	Nuismer and Week 2019
Quantitative genetic simulations	Clade	Maximum likelihood	Phenotypes and phylogenetic relationships for a community of interacting species	Assumes weak coevolution; assumes fixed genetic variance	Drury et al. 2016

are likely to be restricted to laboratory experiments for the foreseeable future because measuring the outcomes and reciprocal fitness consequences of individual interactions in the wild remains intractable in most systems. Indirect, model-based approaches circumvent the challenge of measuring reciprocal fitness consequences directly and thus expand the range of systems where the genomic signature of cryptic coevolution can be investigated.

For example, one way to identify loci experiencing coevolutionary selection is to scan the genome for genetic markers that are correlated across species (Nuismer et al. 2017). This approach builds on techniques developed for studying local adaptation to the abiotic environment that scan the genome for markers that correlate with spatially structured environmental variables, such as water stress, soil nutrient and mineral composition, and temperature profiles (Coop et al. 2010). The coevolutionary extension, then, is to scan the genomes of interacting species in search of markers that covary across space or time (Nuismer et al. 2017). Strong theoretical support exists for this approach, demonstrating that local adaptation in species interactions results from spatial covariation between genes of the interacting species (Nuismer and Gandon 2008; Gandon and Nuismer 2009). Using this method, candidate coevolving loci are identified as interspecific marker pairs that correlate more strongly than expected under a neutral model. In cases where strong local adaptation has been previously demonstrated and isolation by distance is weak, this method can successfully identify those loci currently experiencing coevolutionary selection (Nuismer et al. 2017). More recently, an approach to distinguishing loci responding to coevolutionary selection from those evolving

neutrally has also been developed for cases where genomic data are available from replicate natural populations (Maerkle and Tellier 2020). This approach uses a combination of coalescent simulations and approximate Bayesian computation (ABC) to evaluate support for neutral and coevolutionary models given that candidate loci have been preidentified. The computational feasibility of using this method as a genomic scan and its statistical performance when used in this way remain unexplored however.

These initial efforts to develop indirect methods for identifying loci responding to coevolutionary selection show promise but lag behind similar approaches developed for identifying loci within single species that contribute to local adaptation (Hoban et al. 2016). Specifically, existing coevolutionary approaches do not yet integrate effective methods for distinguishing the genomic signature of coevolution from the background noise created by the interaction of drift, mutation, and gene flow in spatially structured populations (Lotterhos and Whitlock 2015). In particular, as spatial population genetic structure becomes more pronounced, the coevolutionary approach described above suffers from a high rate of false positives (Nuismer et al. 2017). An additional challenge is distinguishing between interspecific correlations attributable to coevolutionary selection and those attributable to shared adaptation to a common environment or unidirectional adaptation of one partner to another. Similar problems—and solutions—in population genetics point the way forward (Hoban et al. 2016). Overall, adapting methods that remove cryptic stratification by including genetic relatedness as a random effect offer a promising path to more robust indirect methods for identifying coevolving loci.



**Figure 2:** Schematic of an indirect method for estimating the strength of coevolutionary selection using phenotypic data. A mathematical model is developed that predicts how population mean phenotypes of interacting species coevolve (left). The model is used to predict the probability that a population is characterized by a particular combination of trait means given a set of model parameters (center; color shading quantifies probability of trait combinations given model parameters). The black dots in the center panel represent population mean phenotypes of plant and animal pairs estimated from different populations. This data can then be used to generate a likelihood surface (right; brighter yellow indicates higher likelihood). This likelihood surface defines the probability of observing the data as a function of model parameters defining the strength of coevolutionary selection (right). The maximum likelihood estimate for the strength of coevolutionary selection is then defined by a parameter combination corresponding to the peak of the likelihood surface indicated by the white dot.

### *Estimating the Intensity and Distribution of Coevolutionary Selection across Phenotypes*

Indirect approaches inferring coevolutionary selection across phenotypes have a long history, in part spurred by Thompson's (2005b) geographic mosaic theory of coevolution. This theory led to a proliferation of studies indirectly investigating species interactions across multiple populations using species' phenotypes. In many cases these studies measure quantitative traits of interacting species across replicate populations and focus on spatial patterns of trait covariation (Brodie and Ridenhour 2003; Zangerl and Berenbaum 2003; Toju and Sota 2006; Pauw et al. 2009). Often, positive correlations have been interpreted as providing evidence for coevolution even while recognizing that multiple noncoevolutionary mechanisms can also lead to similar patterns and correlations (Gomulkiewicz et al. 2007; Nuismer et al. 2010). In some systems with amenable natural history, coevolutionary and noncoevolutionary hypotheses can be distinguished by employing the types of experimental manipulation described in "Direct Measurement of Pairwise Coevolution Is Hard" (Toju and Sota 2006; Pauw et al. 2009). When these experimental manipulations are not possible, patterns of trait covariation are ambiguous with respect to inferring coevolution (Nuismer et al. 2010).

Newer approaches move beyond trait correlations and can be applied to a much greater range of systems. These approaches use statistical methods and spatially replicated data to parameterize coevolutionary models and compare their relative support with models that do not include coevolutionary selection. For example, using previously developed mathematical models predicting the bivariate distribution of population mean trait values for interacting species across a metapopulation, Week and Nuismer (2019) derived maximum likelihood solutions for the model parameters quantifying the strength of coevolutionary selection. The likelihood of the parameterized coevolutionary model can then be compared with the likelihood of the simpler model lacking coevolution using a likelihood ratio test. Application to the well-studied interactions between the Japanese camellia, *Camellia japonica*, and its seed predatory weevil, *Curculio camelliae* (Toju and Sota 2006), and between the long-tongued fly, *Moegistorhynchus longirostris*, and the plant it pollinates, *Lapeirousia anceps* (Pauw et al. 2009), supported the hypothesis of ongoing coevolutionary selection in both systems. Moving forward, these approaches can be easily applied to systems lacking obviously exaggerated traits but where cryptic coevolution between interacting species is suspected.

In addition to identifying the signature of coevolution, model-based approaches allow parameters defining the strength of coevolutionary selection to be estimated, creating the possibility of predicting the future dynamics and

outcome of the coevolutionary process. Nuismer and Week (2019) developed a Bayesian extension of their method that allows for strong selection and gene flow among populations while maintaining many of the other simplifying assumptions of quantitative genetic models (e.g., Gaussian phenotype distributions). Parameter estimates from this method are reliable as long as a large number of populations have been sampled and reasonable estimates are available for important background parameters, such as effective population sizes and additive genetic variances.

These emerging statistical approaches remove one of the key impediments to studying coevolution and quantifying coevolutionary selection: the need to conduct experiments demonstrating phenotype  $\times$  phenotype interactions for fitness. Unfortunately, as currently implemented these approaches are univariate and focus on a single pair of pre-defined phenotypes with an obvious mechanistic connection to the outcome of the interaction. To expand the reach of these methods and realize their full potential to remove biases in the types of systems exposed to coevolutionary study, they need to be generalized to screen multivariate suites of candidate traits, not just those exaggerated traits that so often define the interactions chosen for coevolutionary studies. Because coevolutionary models that integrate multidimensional phenotype space already exist for single populations (Gilman et al. 2012; Debarre et al. 2014) and simple communities (Nuismer and Doeblei 2004; Nuismer 2017; Assis et al. 2020), a clear path exists to realizing this next generation of multidimensional inference tools. A greater constraint to moving this line of research forward is the lack of spatially replicated multivariate trait data for most interacting species pairs and the increasingly formidable challenge of estimating background parameters, such as the additive genetic variance-covariance matrix for each species. An additional challenge that arises with multivariate trait data is the potential for spurious inference driven by selection acting on correlated traits. There is every reason to think, however, that with population mean phenotypes estimated for a suite of candidate traits in each species across a large number of replicate populations, it will be possible to identify the subset of traits experiencing coevolutionary selection, including those cryptically coevolving traits that show no overt signs of exaggeration. As this methodology is extended to larger numbers of traits and systems, we will also begin to understand the phenomic distribution of coevolutionary selection and its overall influence on integrated organismal phenotypes.

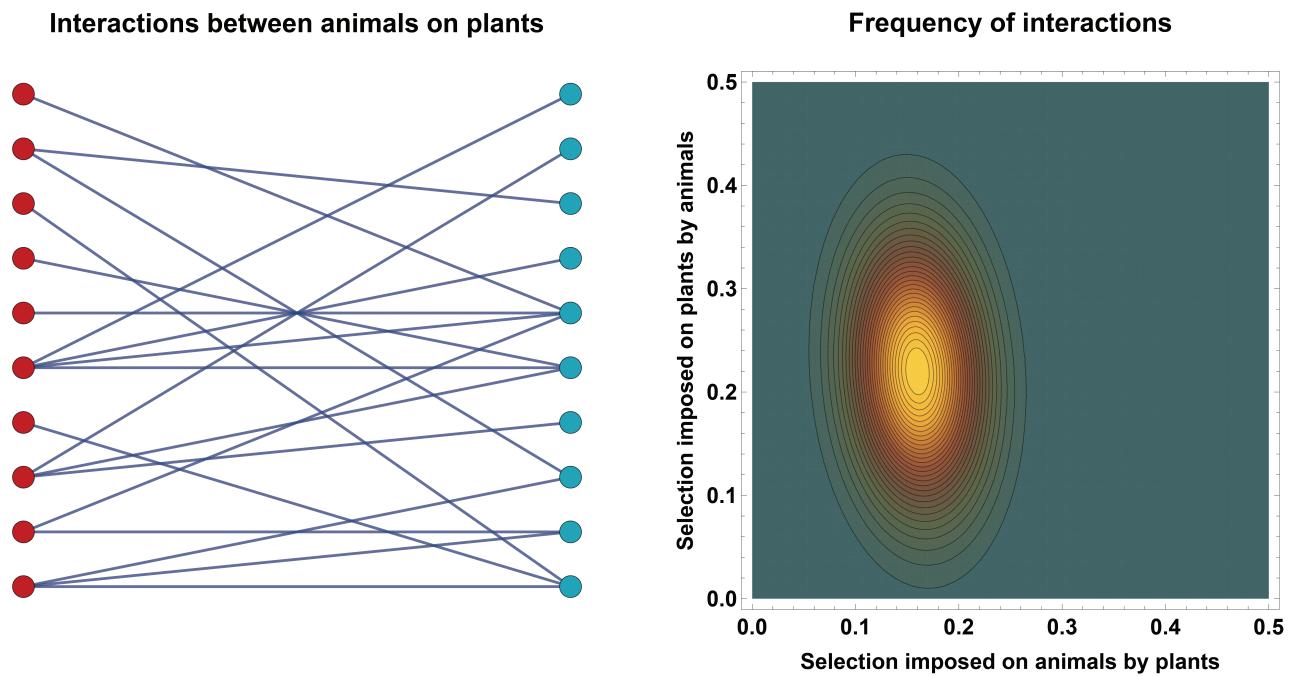
### *Estimating the Intensity and Distribution of Coevolution within Communities*

Work over the past several decades has made enormous strides in increasing our understanding of the distribution

and network structure of ecological interactions within complex biological communities (Gómez et al. 2009; Thebaud and Fontaine 2010; Poisot et al. 2012; Guimaraes et al. 2017; de Andreazzi et al. 2020; Segar et al. 2020). This understanding helps to predict how communities might respond to extirpation of individual species or other anthropogenic disturbances (May 1973; Montoya and Sole 2003; Memmott et al. 2004; Rezende et al. 2007; Nuismer et al. 2018). What we still know almost nothing about, however, is the distribution of coevolutionary selection within ecological communities. For instance, is coevolutionary selection generally focused on a small subset of reciprocally specialized species pairs, or is it more broadly distributed throughout the community? Do stereotypical patterns exist in the network structure of coevolutionary selection? Is coevolutionary selection sufficiently strong and widespread for it to influence the response of communities to disturbance? Answering these fundamental questions about the nature of coevolution at the scale of entire communities requires overcoming a profound technical challenge: devising methods that allow the strength of coevolutionary selection to be estimated for the  $n(n - 1)/2$  species pairs that make up a community of size  $n$ . Unfortunately, directly estimating the strength of coevolutionary selection experienced by all possible pairs of species is, for all intents and purposes, an insurmountable challenge. Instead, progress will require the development of indirect model-

based approaches that estimate the distribution of coevolutionary selection from data that can be collected at the scale of entire communities (fig. 3).

One existing set of methods that come close to revealing insights about coevolution focuses on community assembly (Kraft et al. 2007). The field of phylogenetic community ecology uses phylogenetic data in the form of trees with branch lengths that show the timing of divergence among present-day species (Webb et al. 2002). These trees are usually accompanied by other information about these species, such as how they co-occur in communities and the traits that those species have (Perronne et al. 2017). By comparing patterns in the trees and trait data with what one would expect under a neutral, noninteractive model, one can detect the interactions that species have with one another and with their environment (Kraft et al. 2007). For example, a pattern where species in local communities are overdispersed on phylogenetic trees is often interpreted as the imprint of competition among species with similar traits (Webb et al. 2002). Similar approaches can also be carried out with fossil data, where models support inferences of clade interaction and replacement over long timescales (Lidgard et al. 1993). Unfortunately, most current approaches to macroevolutionary community assembly relate to coevolutionary theory only tangentially and do not yield quantitative estimates for the strength of coevolutionary selection.



**Figure 3:** Development of indirect methods will allow us to move away from describing connectance among community members (*left*) and toward a richer description of how the strength of selection imposed by interactions is distributed (*right*).

The limitations of phylogenetic community ecology could be overcome by building on a growing collection of mechanistic coevolutionary models predicting how communities of interacting species evolve and coevolve (Dieckmann and Law 1996; Loeuille 2010; Nuismer et al. 2013; Cortez 2016; Andreazzi et al. 2017). These models generally focus on communities where coevolutionary interactions between species are mediated by a single quantitative trait, such as phenology in plants or body size in animals, and can be broken down into two broad types. In the first type, feedbacks between ecology and evolution are assumed to be negligible, reducing model complexity sufficiently for analytical solutions to be derived in some cases (e.g., Nuismer et al. 2013). The second type of model is more general and directly integrates the dynamic interplay between coevolution and population dynamics (e.g., Dieckmann and Law 1996). Although these models rarely admit simple analytical solutions, they can be easily simulated to generate predictions for coevolution's influence on the ecological structure and function of communities. Both types of models can be used to develop indirect methods for estimating the distribution of coevolutionary selection within communities, but the methods used in each case differ.

A plausible path to inferring the contemporary strength of coevolutionary selection among species pairs within a complex community could capitalize on existing coevolutionary models to develop maximum likelihood and Bayesian approaches. For those models that focus on only evolutionary dynamics, previous work has derived analytical predictions for the equilibrium distribution of species' trait means as a function of the strength of coevolutionary selection (Nuismer et al. 2013). Specifically, this work studied mutualistic coevolution between plants and animals and demonstrated that the phenotypes of plant species and animal species follow distributions with the mean and variance of plant and animal groups defined by the average strength of coevolutionary selection and a suite of background parameters. Consequently, it is possible to calculate the probability of observing any set of plant and animal phenotypes as a function of the strength of coevolutionary selection as long as independent estimates of key background parameters are available. Although implementing a maximum likelihood approach in this vein is straightforward from a mathematical and computational standpoint, its utility is likely to be restricted by the sparse availability of estimates for important background parameters, such as the strength of abiotic stabilizing selection and additive genetic variation. For models that explicitly track joint evolutionary and ecological dynamics, the challenge is greater because of an increase in model complexity and parameter richness but so are the opportunities for robust inference. As a first pass, the assumption that an ecological and evolutionary equilibrium has been reached could be used to simulate the joint distribution of mean pheno-

types and population densities at steady state. Integrating these steady-state simulations into an ABC framework would allow the strength of coevolution to be estimated if prior distributions for other model parameters could be sufficiently well-defined. An important assumption of these approaches, as they exist now, is the absence of phylogenetic constraints on phenotypic evolution. Thus, a rich area of future development will be developing methods that merge community-level coevolution with phylogenetic history.

#### *Inferring the Historical Intensity and Distribution of Coevolution over Deep Time*

A suite of methods allow investigation of the effects of coevolution over macroevolutionary timescales (Weber et al. 2017; Harmon et al. 2019; Hembry and Weber 2020). These methods analyze either fossil occurrence data or phylogenetic trees, comparing patterns in the data to what one would expect if species were not interacting (Nuismer and Harmon 2015; Manceau et al. 2017). Macroevolutionary approaches to coevolution are appealing in their potential ability to uncover patterns that emerge over very long timescales. The confluence of new methods and better data has resulted in a suite of studies showing convincingly that species interactions influence macroevolution (Drury et al. 2016; Harmon et al. 2019). However, the models that can be fitted to comparative data are often simplistic and/or phenomenological. This limitation means that current methods are good at detecting the imprint of interactions on long-term data but can be only weakly connected to the coevolutionary models of species interactions discussed elsewhere in this article. Most macroevolutionary methods do not even allow definitive inference in favor of coevolution as defined in this article. For example, patterns of phylogenetic signal or cophylogeny can result from many distinct processes, some of which are not the result of coevolution. In this section, we give a brief overview of current methods and suggest ways to better integrate them with microcoevolutionary models.

A major focus in the macroevolutionary study of species interactions has been on the evolutionary dynamics of species' traits. Here, the focus is on detecting the imprint of coevolution on the temporal dynamics of trait evolution, either using fossil time-series data or fitting models to trait evolution along the branches of a phylogenetic tree. In both of these fields, macroevolutionary models for species trait evolution originally focused on a suite of random walk models, including Brownian motion, Ornstein-Uhlenbeck, and various modifications (Felsenstein 1985; Martins and Hansen 1997). For reasons more related to statistical inference than biology, these models long assumed that species evolve independently of one another, precluding any direct inferences of species interactions and coevolution. This limitation has been overcome by newer models in both comparative methods

and paleontology that allow interactions to affect trait evolution (e.g., Nuismer and Harmon 2015; Drury et al. 2016, 2018; Manceau et al. 2017; Weber et al. 2017; Adams and Nason 2018). In turn, empirical analyses have shown that models with species interactions tend to fit data better than models without (Drury et al. 2016). A good recent example is Lomascolo et al. (2019), which combines interaction networks, models of trait evolution, and ancestral state reconstruction to infer coevolution between plants and their pollinators.

However, much more work needs to be done. For one thing, most current approaches are univariate and do not allow investigation of the interactions of multiple traits simultaneously. Additionally, insufficient attention has been paid to the distinction between coevolution and other types of interactive models, such as models where one species evolves in response to the environment and another evolves in response to the first species. The machinery to connect microscale and macroscale coevolutionary models is now in place, and such approaches could potentially be developed further to shed light on the diverse macroevolutionary theories that include species interactions and, presumably, require coevolutionary dynamics.

### **Synthesis: Three Guidelines for the Future Study of Coevolution**

Future investigations of coevolution can take advantage of advances in coevolutionary theory, computational statistics, and the availability of genomic data to build understanding in a more effective and efficient manner. We suggest that three guiding principles, all derived from the above synthesis of theory and data, will help streamline future research and fill the remaining gaps between theory and data.

#### *Develop Coevolutionary Theory That Can Be Tested Using Real, Rather than Idealized, Data*

A strong predictor of the utility of any method is whether the required data already exist or can be easily collected. This is the primary advantage of indirect methods for studying coevolution: they can be designed to use data that are relatively easy to collect and/or that already exist. In contrast, much coevolutionary theory has been aloof to empirical limitations and focused primarily on identifying unifying principles. Consequently, scientific advance depended largely on the creativity of a relatively small handful of experimentalists and field biologists who worked in systems amenable to clever manipulations (Brodie et al. 2002; Benkman et al. 2003; Zangerl and Berenbaum 2003; Toju and Sota 2006; Pauw et al. 2009). We argue that it is time to develop a more balanced portfolio of coevolutionary theory where significantly greater effort focuses on developing statistical and

computational methods that allow models of the coevolutionary process to be compared against one another using readily available data. This development will democratize coevolutionary biology by making it possible to pursue and answer fundamental questions in a wider variety of systems.

#### *Build Flexible Statistical Methods That Directly Tie Coevolutionary Models to Data*

Historically, connections between empirical data and coevolutionary theory have relied on testing qualitative predictions from theory using standard statistical analyses. Although this has led to many important advances, it is also limiting in two key ways. First, this approach enables only qualitative tests of coevolutionary theory. Thus, we are often left with simple black-and-white depictions of colorful continuous processes. Second, this classical approach does not enable us to predict or forecast how coevolution will proceed or how the process will be influenced by perturbations. In contrast, statistical methods that parameterize coevolutionary models with data open the door to predicting the future path of coevolution and allow the consequences of perturbations to be explored. We suggest that the most flexible and general approaches will be simulation-based models parameterized using methods such as ABC.

#### *Rigorously Evaluate the Performance of Model-Based Statistical Methods*

The performance of model-based statistical methods can be assessed in many ways, but the most straightforward approach relies on the analysis of simulated data. The simplest implementation is to simulate data using the same model that powers the statistical method. In this idealized scenario, we expect the statistical method to perform well, accurately estimating parameter values and type I and II errors in cases where they are relevant. When it does not, it suggests model parameters may not be identifiable from the data and/or summary statistics used to describe the data. Even when a method passes this initial test, however, it is important to evaluate how well it works when data are simulated using models that make different assumptions or have slightly different structure. Here, the goal is to get a better feeling for how well we expect the model to work when it is applied to real data—data that will inevitably violate the underlying model in some way, shape, or form. A complementary approach to evaluating the performance of the statistical method and also the explanatory power of the underlying model relies on splitting the data into “training” and “test” sets. Although generally applied within the realm of machine learning, this method can be applied to any model-based approach by estimating model parameters using the training set of the data and then simulating the values of

the data expected under the fitted model to compare with the test set. To our knowledge, this cross-validation approach has yet to be taken within coevolutionary biology and would represent an important advance.

### Conclusion

Advancing our understanding of coevolution will require new methodologies that allow the strength of coevolution to be assessed more broadly and in a way that allows cryptic coevolution to be uncovered. Although still in their infancy, model-based statistical methods that do not require direct estimation of fitness for interacting individuals hold great promise because they remove the primary obstacle to empirical investigation of coevolution, allowing a more diverse range of systems to be explored. These emerging indirect methods can capitalize on what is now a relatively comprehensive collection of mathematical models describing how coevolution structures phenotypes and genotypes over space and across both micro- and macroevolutionary timescales (Nuismer 2017). Broad application of these indirect methods will provide a robust description of the distribution of coevolution across genomes, populations, and species and enhance our ability to predict how communities will respond to increasing levels of anthropogenic disturbance.

### Acknowledgments

Support for this work was provided by grants from the National Institutes of Health (R01GM122079) and the National Science Foundation (DEB1450653) to S.L.N.

### Statement of Authorship

S.L.N. conceptualized the article, and S.L.N., L.J.H., and B.W. wrote the original draft and contributed to review and editing.

### Data and Code Availability

No data or code was generated by this work.

### Literature Cited

- Adams, D. C., and J. D. Nason. 2018. A phylogenetic comparative method for evaluating trait coevolution across two phylogenies for sets of interacting species. *Evolution* 72:234–243.
- Agrawal, A. A., G. Petschenka, R. A. Bingham, M. G. Weber, and S. Rasmann. 2012. Toxic cardenolides: chemical ecology and co-evolution of specialized plant-herbivore interactions. *New Phytologist* 194:28–45.
- Andreazzi, C. S., J. N. Thompson, and P. R. Guimaraes. 2017. Network structure and selection asymmetry drive coevolution in species-rich antagonistic interactions. *American Naturalist* 190:99–115.
- Assis, A. P. A., J. N. Thompson, P. C. Santana, P. Jordano, J. Bascompte, and P. R. Guimaraes. 2020. Genetic correlations and ecological networks shape coevolving mutualisms. *Ecology Letters* 23:1789–1799.
- Benkman, C. W., T. L. Parchman, A. Favis, and A. M. Siepielski. 2003. Reciprocal selection causes a coevolutionary arms race between crossbills and lodgepole pine. *American Naturalist* 162:182–194.
- Berenbaum, M., A. Zangerl, and J. Nitao. 1986. Constraints on chemical coevolution: wild parsnips and the parsnip webworm. *Evolution* 40:1215–1228.
- Brodie, E. D., III, and B. J. Ridenhour. 2003. Reciprocal selection at the phenotypic interface of coevolution. *Integrative and Comparative Biology* 43:408–418.
- Brodie, E. D., Jr., B. J. Ridenhour, and E. D. Brodie III. 2002. The evolutionary response of predators to dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between garter snakes and newts. *Evolution* 56:2067–2082.
- Coop, G., D. Witonsky, A. D. Rienzo, and J. K. Pritchard. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185:1411–1423.
- Cortez, M. H. 2016. How the magnitude of prey genetic variation alters predator-prey eco-evolutionary dynamics. *American Naturalist* 188:329–341.
- de Andreazzi, C. S., J. Astegiano, and P. R. Guimaraes. 2020. Coevolution by different functional mechanisms modulates the structure and dynamics of antagonistic and mutualistic networks. *Oikos* 129:224–237.
- Debarre, F., S. L. Nuismer, and M. Doebeli. 2014. Multidimensional (co)evolutionary stability. *American Naturalist* 184:158–171.
- Dercole, F., R. Ferriere, A. Gragnani, and S. Rinaldi. 2006. Coevolution of slow-fast populations: evolutionary sliding, evolutionary pseudo-equilibria and complex red queen dynamics. *Proceedings of the Royal Society B* 273:983–990.
- Dieckmann, U., and R. Law. 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology* 34:579–612.
- Doebeli, M., and U. Dieckmann. 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. *American Naturalist* 156(suppl.):S77–S101.
- Drury, J., J. Clavel, M. Manceau, and H. Morlon. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. *Systematic Biology* 65:700–710.
- Drury, J. P., G. F. Grether, T. Garland, and H. Morlon. 2018. An assessment of phylogenetic tools for analyzing the interplay between interspecific interactions and phenotypic evolution. *Systematic Biology* 67:413–427.
- Ebert, D., and P. Fields. 2020. Host-parasite co-evolution and its genomic signature. *Nature Reviews Genetics* 21:754–768.
- Ehrlich, P., and P. Raven. 1964. Butterflies and plants: a study in co-evolution. *Evolution* 18:586–608.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125:1–15.
- Gandon, S., and S. L. Nuismer. 2009. Interactions between genetic drift, gene flow, and selection mosaics drive parasite local adaptation. *American Naturalist* 173:212–224.
- Gandon, S., and S. P. Otto. 2007. The evolution of sex and recombination in response to abiotic or coevolutionary fluctuations in epistasis. *Genetics* 175:1835–1853.
- Gilman, R. T., S. L. Nuismer, and D.-C. Jhlueng. 2012. Coevolution in multidimensional trait space favours escape from parasites and pathogens. *Nature* 483:328–330.

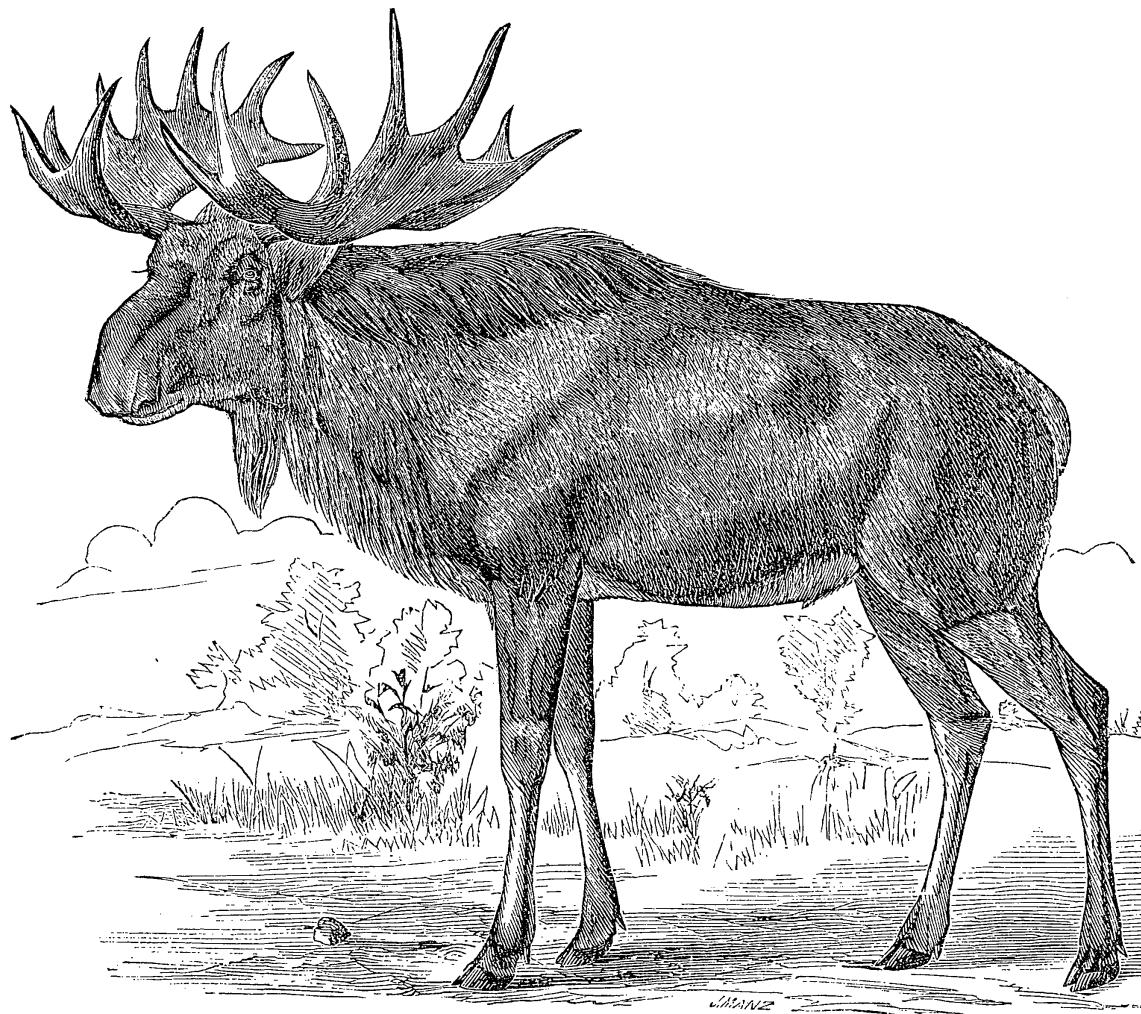
- Gómez, J. M., M. Abdelaziz, J. P. M. Camacho, A. J. Munoz-Pajares, and F. Perfectti. 2009. Local adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecology Letters* 12:672–682.
- Gomulkiewicz, R., D. M. Drown, M. F. Dybdahl, W. Godsoe, S. L. Nuismer, K. M. Pepin, B. J. Ridenhour, C. I. Smith, and J. B. Yoder. 2007. Dos and don'ts of testing the geographic mosaic theory of coevolution. *Heredity* 98:249–258.
- Guimaraes, P. R., M. M. Pires, P. Jordano, J. Bascompte, and J. N. Thompson. 2017. Indirect effects drive coevolution in mutualistic networks. *Nature* 550:511–514.
- Hanifin, C. T., E. D. Brodie Jr., and E. D. Brodie III. 2008. Phenotypic mismatches reveal escape from arms-race coevolution. *PLoS Biology* 6:471–482.
- Harmon, L. J., C. S. Andreazzi, F. Debarre, J. Drury, E. E. Goldberg, A. B. Martins, C. J. Melian, et al. 2019. Detecting the macroevolutionary signal of species interactions. *Journal of Evolutionary Biology* 32:769–782.
- Heath, K. D., and S. L. Nuismer. 2014. Connecting functional and statistical definitions of genotype by genotype interactions in coevolutionary studies. *Frontiers in Genetics* 5:77.
- Hembry, D., and M. G. Weber. 2020. Ecological interactions and macroevolution: a new field with old roots. *Annual Review of Ecology Evolution and Systematics* 51:215–243.
- Hoban, S., J. Kelley, K. Lotterhos, M. Antolin, G. Bradburd, D. Lowry, M. Poss, L. Reed, A. Storfer, and M. Whitlock. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *American Naturalist* 188:379–397.
- Janzen, D. H. 1980. When is it coevolution. *Evolution* 34:611–612.
- King, K. C., J. Jokela, and C. M. Lively. 2011. Parasites, sex, and clonal diversity in natural snail populations. *Evolution* 65:1474–1481.
- Kraft, N. J. B., W. K. Cornwell, C. O. Webb, and D. D. Ackerly. 2007. Trait evolution, community assembly, and the phylogenetic structure of ecological communities. *American Naturalist* 170:271–283.
- Lidgard, S., F. Mckinney, and P. Taylor. 1993. Competition, clade replacement, and a history of cyclostome and cheilostome bryozoan diversity. *Paleobiology* 19:352–371.
- Loeuille, N. 2010. Influence of evolution on the stability of ecological communities. *Ecology Letters* 13:1536–1545.
- Lomascolo, S. B., N. Giannini, N. P. Chacoff, R. Castro-Urgal, and D. P. Vazquez. 2019. Inferring coevolution in a plant-pollinator network. *Oikos* 128:775–789.
- Lotterhos, K. E., and M. C. Whitlock. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* 24:1031–1046.
- MacPherson, A., S. P. Otto, and S. L. Nuismer. 2018. Keeping pace with the red queen: identifying the genetic basis of susceptibility to infectious disease. *Genetics* 208:779–789.
- Maerke, H., and A. Tellier. 2020. Inference of coevolutionary dynamics and parameters from host and parasite polymorphism data of repeated experiments. *PLoS Computational Biology* 16:e1007668.
- Manceau, M., A. Lambert, and H. Morlon. 2017. A unifying comparative phylogenetic framework including traits coevolving across interacting lineages. *Systematic Biology* 66:551–568.
- Martins, E. P., and T. F. Hansen. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist* 149:646–667.
- May, R. M. 1973. Stability and complexity in model ecosystems. Monographs in Population Biology 6. Princeton University Press, Princeton, NJ.
- Memmott, J., N. M. Waser, and M. V. Price. 2004. Tolerance of pollination networks to species extinctions. *Proceedings of the Royal Society B* 271:2605–2611.
- M'Gonigle, L. K., J. J. Shen, and S. P. Otto. 2009. Mutating away from your enemies: the evolution of mutation rate in a host-parasite system. *Theoretical Population Biology* 75:301–311.
- Montoya, J. M., and R. V. Sole. 2003. Topological properties of food webs: from real data to community assembly models. *Oikos* 102:614–622.
- Nuismer, S. L. 2017. Introduction to coevolutionary theory. W. H. Freeman, New York.
- Nuismer, S. L., and M. Doeblei. 2004. Genetic correlations and the coevolutionary dynamics of three-species systems. *Evolution* 58:1165–1177.
- Nuismer, S. L., and S. Gandon. 2008. Moving beyond common-garden and transplant designs: insight into the causes of local adaptation in species interactions. *American Naturalist* 171:658–668.
- Nuismer, S. L., R. Gomulkiewicz, and B. J. Ridenhour. 2010. When is correlation coevolution? *American Naturalist* 175:525–537.
- Nuismer, S., and L. J. Harmon. 2015. Predicting rates of interspecific interaction from phylogenetic trees. *Ecology Letters* 18:17–27.
- Nuismer, S. L., C. E. Jenkins, and M. F. Dybdahl. 2017. Identifying coevolving loci using interspecific genetic correlations. *Ecology and Evolution* 7:6894–6903.
- Nuismer, S. L., P. Jordano, and J. Bascompte. 2013. Coevolution and the architecture of mutualistic networks. *Evolution* 67:338–354.
- Nuismer, S. L., and B. Week. 2019. Approximate Bayesian estimation of coevolutionary arms races. *PLoS Computational Biology* 15:e1006988.
- Nuismer, S. L., B. Week, and M. A. Aizen. 2018. Coevolution slows the disassembly of mutualistic networks. *American Naturalist* 192:490–502.
- Otto, S. P., and S. L. Nuismer. 2004. Species interactions and the evolution of sex. *Science* 304:1018–1020.
- Pauw, A., J. Stofberg, and R. J. Waterman. 2009. Flies and flowers in Darwin's race. *Evolution* 63:268–279.
- Perronne, R., F. Munoz, B. Borgya, X. Reboud, and S. Gaba. 2017. How to design trait-based analyses of community assembly mechanisms: insights and guidelines from a literature review. *Perspectives in Plant Ecology, Evolution and Systematics* 25:29–44.
- Poisot, T., P. H. Thrall, and M. E. Hochberg. 2012. Trophic network structure emerges through antagonistic coevolution in temporally varying environments. *Proceedings of the Royal Society B* 279:299–308.
- Rezende, E. L., J. E. Lavabre, P. R. Guimaraes Jr., P. Jordano, and J. Bascompte. 2007. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature* 448:925–928.
- Ridenhour, B. J. 2005. Identification of selective sources: partitioning selection based on interactions. *American Naturalist* 166:12–25.
- Segar, S. T., T. M. Fayle, D. S. Srivastava, T. M. Lewinsohn, O. T. Lewis, V. Novotny, R. L. Kitching, and S. C. Maunsell. 2020. The role of evolution in shaping ecological networks. *Trends in Ecology and Evolution* 35:454–466.
- Thebault, E., and C. Fontaine. 2010. Stability of ecological communities and the architecture of mutualistic and trophic networks. *Science* 329:853–856.
- Thompson, J. 1994. The coevolutionary process. University of Chicago Press, Chicago.

- \_\_\_\_\_. 2005a. Coevolution: the geographic mosaic of coevolutionary arms races. *Current Biology* 15:R992–R994.
- \_\_\_\_\_. 2005b. The geographic mosaic of coevolution. University of Chicago Press, Chicago.
- Toju, H., and T. Sota. 2006. Imbalance of predator and prey armament: geographic clines in phenotypic interface and natural selection. *American Naturalist* 167:105–117.
- Wang, M., F. Roux, C. Bartoli, C. Huard-Chauveau, C. Meyer, H. Lee, D. Roby, M. S. McPeek, and J. Bergelson. 2018. Two-way mixed-effects methods for joint association analysis using both host and pathogen genomes. *Proceedings of the National Academy of Sciences of the USA* 115:E5440–E5449.
- Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue. 2002. Phylogenies and community ecology. *Annual Review of Ecology and Systematics* 33:475–505.
- Weber, M. G., C. E. Wagner, R. J. Best, L. J. Harmon, and B. Matthews. 2017. Evolution in a community context: on integrating ecological interactions and macroevolution. *Trends in Ecology and Evolution* 32:291–304.
- Week, B., and S. L. Nuismer. 2019. The measurement of coevolution in the wild. *Ecology Letters* 22:717–725.
- Week, B., S. Nuismer, L. Harmon, and S. Krone. 2021. A white noise approach to evolutionary ecology. *Journal of Theoretical Biology* 521:110660.
- Yoder, J. B., and S. L. Nuismer. 2010. When does coevolution promote diversification? *American Naturalist* 176:802–817.
- Zangerl, A. R., and M. R. Berenbaum. 2003. Phenotype matching in wild parsnip and parsnip webworms: causes and consequences. *Evolution* 57:806–815.

Focused Topic Editor: Ana Paula A. Assis

Focused Topic Editor: Paulo R. Guimarães Jr.

Editor: Daniel I. Bolnick



"Persons are frequently puzzled by the reverse use of the terms 'moose' and 'elk.' The author makes it perfectly clear that the American moose is the analogue of the palmate-horned animal called 'elk' in Europe; and that the American elk is the analogue of the stag or red deer of Europe." Figured: "Scandinavian Elk." From the review of Caton's *Deer of America* by E. Coues (*The American Naturalist*, 1877, 11:354–358).