OIKOS

Forum

Phylogenetic generalized linear mixed modeling presents novel opportunities for eco-evolutionary synthesis

Amanda S. Gallinat and William D. Pearse

A. S. Gallinat (https://orcid.org/0000-0003-0397-6562) ✉ (gallinat@uwm.edu) and W. D. Pearse, Dept of Biology and Ecology Center, Utah State Univ., Logan, UT, USA. ASG also at: Dept of Geography, Univ. of Wisconsin-Milwaukee, Milwaukee, WI, USA. WDP also at: Dept of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire, UK.

Despite their interdependence, community ecology and evolutionary biology have proven difficult to synthesize empirically in studies of community assembly. This is primarily due to differing temporal and spatial scales of ecological and evolutionary dynamics, ranging from broad-scale processes like speciation and environmental filtering to local-scale past and present-day niche dynamics. Phylogenetic generalized linear mixed modeling (PGLMM) offers a solution to this problem, it can be used to integrate through time by modeling the evolution of trait-based community assembly, and across space ranging from broad-scale environmental sensitivities to local-scale co-occurrences. As such, PGLMM provides the ability to compare the relative strength of deep versus shallow-time drivers of biodiversity by including them in a single model. Despite its unique value, the application of PGLMM has been limited because statistical advances have not been adequately matched by conceptual progress. Recent expansion in the availability of cross-clade assemblage data and phylogenetic tools have increased the urgency of conceptual unification. Here we describe the potential of PGLMM for parsing the evolutionary and ecological drivers of community assembly, focusing on how three major drivers – environmental sensitivities, within-clade interactions (e.g. competition), and cross-clade associations (e.g. herbivory) – shape historical and present-day assemblages. We outline three fundamental questions that PGLMM can address, linked to each of the aforementioned drivers: 1) are species' regional-scale environmental responses evolutionarily constrained? 2) Do evolved responses to past competition minimize or enhance present-day competition? 3) Are cross-clade associations evolutionarily constrained? For each question, we review conceptual advances and opportunities, and demonstrate the application of PGLMM in

Synthesis

The synthesis of evolutionary biology and community ecology is critical to understanding species assemblages today and predicting future shifts. Here we outline how phylogenetic generalized linear mixed modeling (PGLMM) can accomplish such synthesis and provide foundational insights into the drivers of community assembly. Focusing on the evolution of environmental constraints and interactions within and across taxa, we review conceptual advances and opportunities, and demonstrate the application of PGLMM in a supplementary tutorial. We also highlight the application of these concepts and tools for conservation and natural resource management to stimulate future research with clear conceptual and practical goals.

a supplementary tutorial. We focus on the ecological and evolutionary outcomes of PGLMM and describe the value of these outcomes for conservation and natural resource management, in order to move PGLMM beyond statistical complexities and toward a future with clear conceptual and practical goals.

Bridging ecology and evolution: a challenge of scale

Evolution is the context for present-day ecological interactions, which in turn shape the evolution of species (Webb et al. 2002), although evolutionary biology and community ecology have proven notoriously difficult to integrate empirically. This has predominantly been an issue of scale, classically, ecological processes operate across smaller spatio–temporal scales than evolutionary processes (Webb et al. 2002, Cavender-Bares et al. 2009; but see, Richardson et al. 2014), but for practical reasons biodiversity data often represent *either* local *or* regional scales. Models and data sets that integrate across multiple spatial and temporal scales to capture processes dependent on dynamics that are both broad- and fine-scale in nature are difficult to generate and so rare. Yet, they are necessary to describe the interaction of evolutionary history and ongoing ecology in driving species occurrences over space and time (Pearse et al. 2018).

The lack of such cross-scale statistical models and empirical data has limited our ability to make broad inferences about the processes that determine the distribution of biodiversity. For instance, it is generally accepted that environmental filtering limits species' distributions at the regional scale, while at local scales competition and facilitation within and among clades (e.g. vascular plants, insects, mammals) additionally influence community assembly (Kraft et al. 2015). Yet it is also generally accepted that this is an oversimplification, and that disentangling the interaction of these niche-based processes requires the integration of regional and local data (Cadotte and Tucker 2017). Even the same process can have profoundly different effects when operating at different scales. For instance, dispersal limitation can mediate coexistence and priority effects when operating at a finer spatio–temporal scale, while at broader scales it can drive speciation (Vellend 2010).

The next great challenge for those who seek to understand and predict community structure is to solve this problem of parsing cross-scale processes by explicitly modeling those processes in cross-scale data (McGill 2019). Recent efforts to address this challenge have called for tools which incorporate biogeographic history and present-day ecology to test evolutionary theory, interpret ecological patterns and predict future community assemblages (Pearse et al. 2018).

Community phylogenetics as a field has made strides toward answering this call, and for bridging the ecology and evolution of community assembly (Weber et al. 2017). By incorporating phylogeny with ecological data community phylogenetics can illuminate processes regulating biodiversity across spatial and temporal scales. However, the field has been appropriately criticized for focusing on the statistical rather than the conceptual, and specifically for struggling to link statistical patterns to evolutionary process (Mayfield and Levine 2010), which has stymied theoretical advances and practical applications to ecology and conservation biology. For instance, it is now common to use community phylogenetics in studies of species' traits to statistically account for biases stemming from relatedness (Webb et al. 2002). It is far more rare for such studies to use phylogenetics to test eco-evolutionary concepts like the role of trait evolution in present-day niche dynamics. Within the eco-phylogenetic toolbox, phylogenetic generalized linear mixed modeling (PGLMM) can offer particularly valuable insight into the drivers of community assembly, including the evolution of environmental sensitivities and interactions within and across clades (Ives and Helmus 2011, Pearse et al. 2014). Despite its value, however, this method remains infrequently used in studies of community assembly, due to a lack of conceptual unification and demonstration of the tool. Without such efforts, the synthesis of evolutionary history and present-day ecology will remain limited.

Here we outline the conceptual value of PGLMM for ecology, evolution and community assembly research, and examine its past use and frontiers for future exploration. We outline three conceptual questions, each focused on how past evolution constrains present-day ecological responses, that PGLMM can address (with associated models and exemplar code). Finally, we link those questions to actionable insights for conservation and natural resource management.

Phylogenetic generalized linear mixed modeling in community assembly research

Eco-phylogenetic approaches offer a unique opportunity to incorporate the evolutionary history of species (their phylogeny) into analyses of current species occurrences to determine how the shared biogeographic history of related species shapes their present-day ecology. Within community assembly research, eco-phylogenetic analyses have traditionally focused on whether co-occurring species are more or less closely-related than would be expected by chance. Under the assumption of niche conservatism (wherein closely-related species tend to resemble one-another), the co-occurrence of close relatives is the product of environmental filtering on shared tolerances, and distantly-related species co-occur through competitive exclusion on the basis of similar traits (Webb et al. 2002, Cavender-Bares et al. 2009). This

Box 1. Past and potential uses of PGLMM

$$\text{abund}_{\text{spp,site}} \sim \text{Poisson}(\beta_{\text{spp}} \times b_{\text{site}} + a_{\text{spp}} + \alpha_{\text{spp,site}}) \qquad (1)$$

Equation 1 defines a PGLMM (Ives and Helmus 2011) case where $\text{abund}_{\text{spp,site}}$ is the abundance of species (spp) across sites (site), a_{spp} is the overall abundances of species and other terms are defined below. This qualifies as PGLMM because phylogeny can inform its hierarchical effects, $a_{\text{spp,site}}$ and β_{spp}.

$$\beta_{\text{spp}} \sim N\left(\mu_\beta, \sigma_\beta^2 \times \lambda(\text{VCV}_{\text{phy}}, \hat{\lambda})\right) \qquad (2)$$

Equation 2 models the β-trait as a draw from a multivariate Normal distribution with an estimated central tendency (μ_β) and covariance proportional to species' phylogenetic covariance matrix (VCV_{phy}) and overall variation (σ^2). By scaling phylogenetic covariance according to a linear transformation, such as Pagel's λ transformation (1999), we can estimate evolutionary constraint ($\hat{\lambda}$; in this case making μ_β the estimate of ancestral β at the root). Alternative transformations of VCV_{phy} can be used to test for different evolutionary models (Uyeda et al. 2015 for how linear transformations can be used for OU models).

$$\alpha_{\text{spp,site}} \sim N\left(0, \sigma_{\text{site}} \times I_{\text{site}} \otimes \sigma_\alpha \times \text{VCV}_{\text{phy}}^{-1}\right) \qquad (3)$$

Equation 3 accounts for the α-trait in two parts. The first accounts for each species' independent variation in site-level abundances – an identity matrix with dimensions equal to the number of sites (I_{site}) multiplied by a scalar (σ_{site}) – the second accounts for evolutionary repulsion of species' co-occurrences – the inverse of the phylogenetic covariance matrix ($\text{VCV}_{\text{phy}}^{-1}$) multiplied by a scalar (σ_α). These terms are combined through their Kronecker product, and the magnitude of σ_α^2 reflects the importance of repulsive trait evolution. In the Supporting information, we show how to fit these models to empirical data, and also how subtle changes to these equations can test the hypotheses we outline in the main text in the 'Are species' regional-scale environmental responses evolutionarily constrained?', 'Do evolved responses to past competition minimize or enhance present-day competition?' and 'Are cross-clade associations evolutionarily constrained?' sections. These include extending this model to account for the phylogenetic structure of another clade (the basis of 'Are cross-clade associations evolutionarily constrained' section).

---

simple, yet powerful, framework has driven advances in our understanding of how ecological structure varies across taxa (Swenson et al. 2006), and how it interacts with the evolution of species' traits (Rabosky et al. 2011, Pearse et al. 2019). Yet it has also led to the understanding that mapping eco-phylogenetic pattern onto particular ecological or evolutionary processes is fraught with difficulty (Mayfield and Levine 2010), mapping pattern onto process is essentially impossible if we cannot contrast the relative strengths of differing drivers to explain observed patterns. Thus, recent advances in the field of community phylogenetics have aimed to synthesize deep- and shallow-time ecological and evolutionary processes into a single model to contrast the relative strength of potential drivers of biodiversity.

Phylogenetic generalized linear mixed modeling (PGLMM; Ives and Helmus 2011) has emerged as a powerful way to quantify species' occurrences or abundances as a function of species' traits and environmental conditions (Box 1). PGLMM goes beyond describing the structure of assemblages (e.g. containing closely-related, functionally distinct species) to describe the processes structuring assemblages (e.g. divergent trait evolution followed by competition on the basis of that divergent trait). By using random effect terms to account for site- and species-level variation and measuring phylogenetic patterns among those random effects, PGLMM can be used to measure phylogenetically-patterned

differences in how species interact with each other and their environment. Multiple power analyses of PGLMM (Ives and Helmus 2011, Rafferty and Ives 2013, Pearse et al. 2014), have already demonstrated its statistical and methodological value, however, applications of PGLMM have remained limited in community assembly research.

The application of PGLMM has been limited in part by the need for high-resolution phylogenies and cross-scale ecological community data. The recent rise in accessibility of both these tools has vastly increased opportunities to leverage PGLMM for eco-evolutionary insights into community assembly. In recent years, the proliferation of available global phylogenies (Faurby and Svenning 2015) and automated tools (Smith and Walker 2019) has made the phylogenetic data required for such analyses available to ecologists. At the same time, ecology's big data revolution has increased the availability of community data that cross relevant spatial and taxonomic scales for both macro-ecological insights and local-scale applications. Empirically parsing how regional environmental and local within- and cross-clade processes motivate ecological assembly with PGLMM requires data sets that capture all three processes. Specifically, this challenge requires overlapping local-scale assemblage data for multiple clades that connect across broader (often regional) environmental gradients (Table 1). Earth observation networks like the USA's National Ecological Observatory Network (NEON;

Table 1. Conceptual and data guidelines for using PGLMM to examine broad-scale environmental responses (β-trait models), local-scale co-occurrences (α-trait models) and models of cross-clade associations. Each of three questions addresses a unique process underlying community assembly, and by building on the previous question(s), they can be used to parse the relative influences of these three processes. For each question, we describe the types of data required for PGLMM (including sampling breadth and density) and provide examples of existing data sets to stimulate future work. We also include conceptual illustrations of data collection, in which each color represents a different species or clade corresponding with the phylogenies at the right. Double arrows indicate closely-related species that have either evolved to minimize co-occurrence (top panel) or have constrained evolution such that they compete in the present (middle panel). At the regional scale (top panel), species within a clade (co-)occur based on environmental sensitivities that drive broad-scale distributions; the double arrow indicates species that have evolved not to co-occur. At the local scale (middle panel), species that overlap in distribution at the regional scale may not co-occur due to within-clade interactions; the double arrows indicate evolutionary constraints such that species compete in the present. Investigating cross-clade associations (bottom panel) requires that the regional- and local-scale data illustrated above be collected for multiple clades; for species pairs across clades, the presence of one, both or neither can be used to identify the phylogenetic patterns of association.

Question	Data Requirements	Examples	Data Collection
(1) Are species' regional-scale environmental responses evolutionarily constrained?	• Spatial extent includes an environmental gradient across which (co-)occurrence patterns vary • Assemblage sampling is sufficient to estimate phylogenetic structure of β-traits	• Global Biodiversity Information Facility (GBIF: www.gbif.org; Edwards, 2000) • eBird (www.ebird.org; Sullivan et al., 2014) • Specimen data, *e.g.* US Integrated Digitized Biocollections (iDigBio: www.idigbio.org; Matsunaga et al., 2013)	
(2) Do evolved responses to past competition minimize or enhance present-day competition?	• Requirements for **(1)** • Sampling density within each local site captures variation in species co-occurrence patterns • Assemblage sampling is sufficient to estimate phylogenetic structure of α-traits	• International Long-Term Ecological Research Network (ILTER: Mirtl et al., 2018) • U.S. Forest Inventory & Analysis (FIA: www.fia.fs.fed.us; Smith, 2002) • North American Breeding Bird Survey (BBS: www.pwrc.usgs.gov/bbs; Sauer et al., 2013)	
(3) Are cross-clade associations evolutionarily constrained?	• Requirements for **(1)** and **(2)** • Data overlap spatially and temporally for associated clades; interaction networks further inform associations of interest • Assemblage sampling is sufficient to estimate phylogenetic structure across (monophyletic) clades	• U.S. National Ecological Observatory Network (NEON: www.neonscience.org; Schimel et al., 2007) • Australia's Terrestrial Ecosystem Research Network (TERN) SuperSites (supersites.tern.org.au; Karan et al., 2016)	

Schimel et al. 2007) and Australia's Terrestrial Ecosystem Research Network SuperSites (TERN; Karan et al. 2016) employ sampling designs that meet all of these requirements, and have very recently emerged as providers of this type of nested community data. The prescribed, top–down approach of these networks pairs intentional site distribution with robust sampling protocols to capture regional environmental responses for a wide range of taxa. Nested within this regional site design, local assemblage data can be used to investigate the interplay between species historic and current niches through the phylogenetic patterns of co-occurrence within clades. Furthermore, local-scale species assemblage data for *multiple* clades provides the power to model the phylogenetic structure of co-occurrences across clades (e.g. flowering plants and their insect pollinators, or ticks and their rodent hosts) allowing researchers to test the biotic drivers of species distributions.

PGLMM can reveal insights into the ecology and evolution of regional sensitivities, within-clade and cross-clade dynamics independently of one another (using data sets that capture one or two but not all three dynamics, e.g. data from museum specimens, forest inventory and analysis surveys, eBird). However, the most sophisticated conceptual advances will come from applying PGLMM to nested, cross-clade community data to model local-scale within- and cross-clade dynamics together within the context of broad-scale environmental filtering, to compare their relative contributions to species assembly and generate predictions of future (co-)occurrences. PGLMM is well-suited to analyzing cross-scale patterns from this type of nested data because it inherently integrates broad-scale processes underlying the macro-evolution of species traits with their local-scale consequences for community assembly. Indeed, PGLMM often requires a combination of regional and local community data, since the composition of nearby sites rarely encompasses sufficient environmental variation for macro-scale conclusions. Studies of small assemblages, or large assemblages within small regions, are often insufficient to detect both environmental filtering and the evolutionary signal of environmental filtering (Swenson et al. 2006). While we have primarily highlighted data and scales
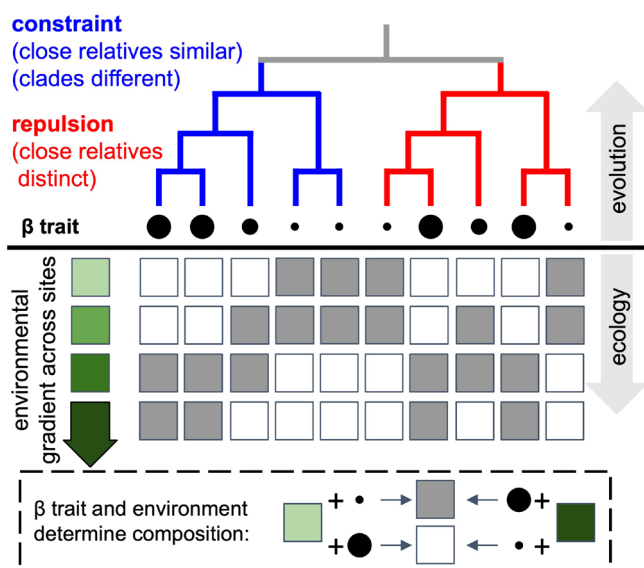
Figure 1. Conceptual overview of regional-scale and macro-evolutionary hypotheses that can be tested with PGLMM. Ten species (columns) and their presence in four sites along an environmental gradient (rows) are shown. The match of species' β-traits (circles) to the environment (green boxes) – environmental filtering – determines community membership. The β-traits themselves are shown as they would be if evolving under two distinct macro-evolutionary models: repulsion (the red clade on the right) and constraint (the blue clade on the left). We note that the inference of constraint from trait similarity is nuanced, while phylogenetic signal often results from phylogenetic inheritance of trait similarity, there are exceptions (Revell et al. 2008).

relevant to macro-organisms, we stress that researchers can use PGLMM to derive insights into community assembly in many systems with different scales of interest. For instance, community ecologists could use the same conceptual framework outlined here to explore how environmental filtering, within- and cross-clade dynamics drive community patterns in microbial assemblages across environmental gradients (Nemergut et al. 2013), or in studies of small, isolated communities (e.g. cactus yeasts; Anderson et al. 2004).

With the expansion of phylogenetic tools and cross-scale data, researchers are poised to use PGLMM to address some of our most fundamental questions about how biological communities are generated and maintained. To make the most of these statistical and technological advances, it is urgent that we define and explore the conceptual framework underlying PGLMM. This framework should be central to the design of future species assembly research as well as its applications to natural resource management.

Three eco-evolutionary questions PGLMM can answer

PGLMM provides the flexibility necessary to investigate three cross-scale processes driving species distributions: the evolution of 1) abiotic sensitivities, 2) within-clade interactions

and 3) cross-clade associations. To address these research themes, we outline the conceptual framework underlying PGLMM that integrates differing mechanisms of the macro-evolution of species' traits with both macro- and local-scale assembly processes based upon those traits (Fig. 1). In this framework, species traits evolve under constraint (in which close relatives are similar, constrained by broad optima; but see Revell et al. 2008 for alternative mechanisms behind trait similarity) or repulsion (reducing similarity among close relatives, putatively to minimize competition; Nuismer and Harmon 2015). These same traits are then filtered by the regional environment and drive local-scale competition to dictate assemblage composition. Associations such as pollination and trophic links constitute the cross-clade environment which can further limit or enable species occurrences on the basis of resource availability, mutualism and pressure from predators, parasites and pathogens. Reflecting an increasing recognition of the role of species' associations in community assembly, this framework incorporates cross-clade co-occurrence within the context of within-clade and regional dynamics to understand how species' evolution and community assembly shape, and are shaped by, these processes.

A large part of the PGLMM framework is based around the concepts of β- and α-traits, which parse the differing roles that species traits can play in community assembly (Silvertown et al. 2006). Regional habitat affiliations are determined by β-traits, which regulate species' tolerances to environmental filters and are hypothesized to evolve early and be strongly conserved through time. By contrast, α-traits are hypothesized to evolve later, and determine local-scale co-occurrence, competition and environmental filtering. Drought tolerance, for instance, is considered a β-trait at the regional scale, with soil moisture explaining differences among habitat types and thus regionally clustering related species with shared water use traits (Moeslund et al. 2013). Yet such tolerance is an α-trait when driving local-scale hydrological niche differentiation in a different system (Silvertown et al. 2015). That the same trait can be subject to different evolutionary and ecological forces in different taxa and environments is unsurprising. Indeed, we argue below that modeling traits in this fashion will allow us better generalize seemingly discordant results across systems and unpick the cross-scale drivers of community assembly.

The eco-evolutionary framework underlying PGLMM is not hypothetical, and can be applied by combining readily-available tools, including the R packages 'pez' (Pearse et al. 2015) and 'phyr' (Li et al. 2020), as well as ecological data and phylogenetic tools described above. To demonstrate the structure and utility of PGLMM, we provide examples of how to fit PGLMM to address each of the conceptual processes outlined below in the Supporting information. This code maps directly onto the overview given in Box 1.

Are species' regional-scale environmental responses evolutionarily constrained?

While it is now well-established in the comparative literature that assessing patterns in species' traits without accounting

for their shared evolutionary history can introduce bias (Freckleton et al. 2002), it is only beginning to be recognized that the same biases can affect studies of species' ecological assembly. For instance, PGLMM has been shown to reveal even stronger signals of trait-based assembly by accounting for phylogenetic non-independence (Li and Ives 2017). As in the comparative literature, however, phylogeny can play a much greater role in our understanding of community assembly than as a mere statistical correction for non-independence. At the regional scale, phylogeny can be used to investigate the evolutionary constraints on, and evolutionary processes underlying, species' environmental responses. Many mechanisms are invoked to explain the pattern of phylogenetic signal in β-traits, many of which involve the environment. It follows that a goal of species assembly modeling should be to measure the strength of niche conservatism itself as reflected in species' environmental responses, in which the responses themselves are β-traits. By incorporating random effects for species and sites, PGLMM can be used to model species' environmental responses a) as trait-based environmental tolerances, using the interaction of species' measured traits and the environment, b) as unmeasured (latent) traits for environmental responses, estimated using phylogeny (as described in Box 1) and c) as coefficients describing species-specific responses to environmental gradients. In the latter case (c), environmental responses are still modeled as latent traits, but can now be used to estimate the extent to which those latent traits are better when informed by phylogeny, and to test for evolutionary constraints on environmental sensitivity. Critically, within this approach latent traits are β-traits, and measure environmental responses. The phylogenetic signal of these responses reflects the relative influence of evolution on present-day ecology, with a Pagel's λ value of one suggesting evolutionary constraint, and a value of zero suggesting none (Pagel 1999). By virtue of phylogeny, this method can be used to estimate the evolutionary signal of sensitivities and predict those for species lacking data, though it should be noted that phylogenetic imputation is unreliable for imputing trait values below the macroecological scale (Swenson 2013, Swenson et al. 2017).

Beyond patterns of environmental response among species, a deeper understanding of the mode by which trait–environment relationships develop is critical to understanding how β-traits affect species assembly. If functional traits are driving environmental responses, PGLMM can be used to co-estimate the evolution and ecology of those traits, allowing for more nuanced inference of each component. Techniques such as divergence-order time analysis (Ackerly et al. 2006) can also be used to determine which came first: evolutionary constraint on the trait, or environmental niche. Equally, application of PGLMM allows us to take advantage of the recent explosion of macro-evolutionary hypotheses and methods, and it is now common to contrast Brownian motion (a null model of conserved evolution; Losos 2011) with Ornstein–Uhlenbeck (OU – a model of constrained evolution around an estimated trait optimum; Butler and King 2004) and accelerating/decelerating models of trait evolution

(Blomberg et al. 2003). OU models may describe the adaptive landscape of macro-evolution (Uyeda and Harmon 2014), but it remains to be seen whether their macro-evolutionary trait tradeoff surfaces apply in the literal landscapes of broad-scale community data. As described in Box 1, differing models could be fit to these data through linear transformations of species' phylogenetic variance–covariance matrices (Freckleton et al. 2002) to meet the conceptual challenge of contrasting different rates and modes of evolution for niches and traits. Critically, this approach allows researchers to estimate the impact of past evolution (e.g. detecting evidence of Brownian motion evolution on a phylogeny), although we caution that accurate rate estimates require complete or non-biased lineage sampling. Indeed, few trait-based comparative methods evade this challenge, as ecological assemblages are inherently biased or phylogenetically incomplete.

Such combined eco-phylogenetic techniques offer a new and promising approach to a pervasive challenge in the study of trait macro-evolution: the mapping of phylogenetic pattern onto process. It is well-known that phylogenetic signal, and/or support for a particular statistical model of trait evolution, does not always reveal the mechanism of evolution. For example, statistical support for Brownian motion could be the result of neutral drift or environmental constraint (Revell et al. 2008). Here again, however, the breadth of cross-scale data described in Table 1 combined with the statistical flexibility of PGLMM provide a way forward. By modeling the evolution of species' environmental responses (β-traits), we can explicitly link ecological mechanism to evolutionary model, and thus use statistical means to address mechanistic questions. For example, the hypothesis that a trait's pattern of phylogenetic signal results from evolutionary constraints on environmental sensitivity can be tested with PGLMM that quantifies both evolutionary constraint and environmental filtering in the present. While taxa with strong fossil records will continue to offer researchers the best opportunities to unpick deep-time eco-evolutionary processes (Ezard et al. 2011), there will always be a limited amount of past data from which to infer evolutionary process. Using PGLMM, researchers can naturally incorporate ecological information into a holistic, phylogenetic natural history perspective (Uyeda et al. 2018), leveraging the wealth of data available in the present to inform our understanding of the past.

Do evolved responses to past competition minimize or enhance present-day competition?

There is still great debate over how we can best measure species' niche differences, and so estimate the magnitude of competition (D'Andrea and Ostling 2016). While some clades have candidate α-traits known to impact local-scale niche structure (e.g. specific leaf area in vascular plants; Cornwell and Ackerly 2009), the latent trait approach described in the 'Are species' regional-scale environmental responses evolutionarily constrained?' section can also aid in the discovery of traits relevant for competition. PGLMM can go further by measuring not just how species' α-traits drive local-scale

assembly, but also jointly estimating how those traits have evolved. Specifically, PGLMM can be used to model species occurrence as a function of trait dispersion, phylogenetic dispersion or both to test whether the extent to which related species overlap in trait space may preclude present-day co-occurrence. To examine modes of α-trait evolution, PGLMM can also be used to model species occurrence as a function of trait dispersion combined with some evolutionary model of the trait. This need not be limited to simple competitive exclusion models of competition based around niche differences (Mayfield and Levine 2010), PGLMM could be applied to more complex models of fitness and stabilizing niche differences (Chesson 2000). Future efforts could also incorporate other explicit models of competition (Drury et al. 2016, Morlon et al. 2016) in much the same way we suggest using phylogenetic transformations to test different models of (latent) trait evolution in Question 1.

The traditional picture of excluding competition driving local-scale ecological assembly has been further complicated by an increasing understanding of the role of Neutral processes (Hubbell 2001, Vellend 2010). Within Neural theory, there is a growing disconnect between two classes of model. In the first class, ecological assembly within clades is random with respect to species' traits, and so stabilizing forces are absent and species do not compete for resources (Hubbell 2001). In the second, traits drive ecological assembly but the evolution of those traits is neutral; stabilizing forces and competition are present and act on traits evolving under drift (Rosindell et al. 2015). To examine the validity of the first class, sufficient environmental variation is required to detect differences across species' ranges, as provided by broad-scale species occurrence data, while the second class requires the integration of ecological and evolutionary dynamics provided by PGLMM.

Rather than evolution constraining close relatives' traits to be similar such that they compete, species may have evolved to reduce ecological competition. While within macro-ecology such evolution, as manifest as character displacement, is reasonably uncontroversial (Brown and Wilson 1956, Schluter and McPhail 1992), in community ecology it is controversial to invoke evolved responses to past competition, or the 'ghost of competition past', to explain an absence of present-day competition (Connell 1980). The integration of evolutionary modeling into community ecological models through PGLMM offers the opportunity to not only test for evidence this ecological ghost, but also improve macro-evolutionary models of species traits. We caution that the usual assumptions of comparative methods apply, such as, in some cases, sympatric macroevolution of interacting species.

While new and promising methods are being developed to measure and simulate the repulsive evolution of species' traits (Drury et al. 2016), it remains difficult to distinguish among different models on the basis of purely extant species' traits (Nuismer and Harmon 2015). Indeed, the signature of ecology on trait evolution can be very difficult to detect over phylogenetic timescales. And inferring evolutionary process from statistical patterns is particularly challenging in the case

of competition (Nuismer and Harmon 2015, Pearse et al. 2019). Given that this problem is well-recognized in evolutionary biology, we propose it is more productive to draw strength from observations of present-day ecology than to generate increasingly complex evolutionary models.

Only such joint eco-evolutionary models can allow us to distinguish between whether traits have evolved to minimize competition, or whether evolutionary constraint has forced species to compete in the present. In practice, the former is characterized by statistical support for traits evolving under repulsion, as measured using the inverse of the phylogenetic covariance matrix (Box 1). Such support, wherein closely-related species tend not to share similar traits, may reflect past competition on the basis of those traits, assuming those relatives co-occurred in the past. On the other hand, evolutionary constraint and present-day competition are supported by phylogenetic signal in focal traits, and phylogenetic (or trait) overdispersion across assemblages (Supporting information). Taken together, these two results would suggest closely-related species, which share similar trait values, tend not to co-occur, potentially due to competition on the basis of those traits. The interpretation and application of such results requires critical reflection and taxonomic expertise (as in the case of all eco-evolutionary models). However, parsing the influence of historical and present-day competition on species assembly is a key area where PGLMM can offer novel insight and push the integration of ecology and evolution forward.

Are cross-clade associations evolutionarily constrained?

The α- and β-trait based model of ecological assembly described above, like many filtering-based models of community assembly, ignores interactions across clades. However, species' distributions result not just from their own intrinsic cross-scale evolutionary and ecological dynamics, but also from their interactions with other groups. In the name of tractability, community ecologists often taxonomically restrict their assemblages (Ricklefs 2008, Vellend 2010) and interaction ecologists are frequently forced to overlook temporal and spatial variation in species' interactions (Poisot et al. 2015). The co-diversification of groups like plants and their pollinators or herbivores, and mammals and their parasites, have been rich and controversial fields of study. Even the most specialized mutualisms that form classic models of diversification have provided surprisingly little evidence for pure, strict coevolution by cospeciation (Hembry and Althoff 2016). And while the timing of diversification in many clades has been well-studied, there are fewer studies of the evolution of the traits driving co-occurrences and associations among these clades (but see, Rafferty and Ives 2013). Thus, modeling the role of the cross-clade environment in shaping communities and driving biodiversity is likely to provide answers to many outstanding evolutionary questions.

For species to interact, they must first be capable of co-occurring in the same space, and thus a natural starting place for modelling variation in network structure is the
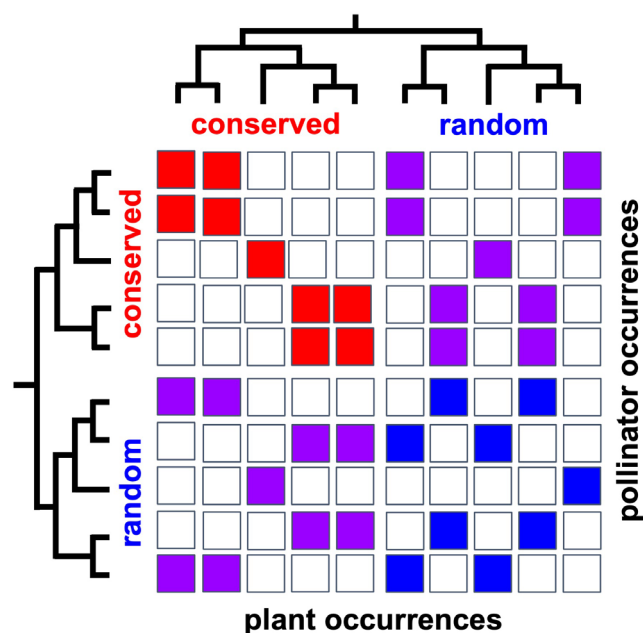
Figure 2. Cross-clade hypotheses of phylogenetic co-occurrence, shown here as plants and pollinators, that can be tested with PGLMM. Ten plant (columns) and ten pollinator (rows) species are shown, with recorded co-occurrences between them shown with colored cells. For each group, two clades, each with either phylogenetically conserved or random co-occurrences, are shown. Each combination of clades represents a different hypothesis about plant–pollinator co-occurrences.

α/β framework outlined above. Using this framework, the simplest model of co-occurrence across clades is a statistical interaction between each group's parameters. While species' responses to environmental drivers are often phylogenetically patterned, the pattern of responses to other species is unclear (Cavender-Bares et al. 2009). The PGLMM framework provides a solution by estimating the effect of the cross-clade environment on species occurrences. By comparing data describing the presence of two interacting species, as in Fig. 2, co-occurrence data can be modeled as they would be in a PGLMM of interaction network data (Rafferty and Ives 2013), incorporating two separate phylogenies to generate expected pairwise interaction rates. The responses and effects of clades could both be phylogenetically patterned (top left of Fig. 2), neither could be phylogenetically patterned (bottom right of Fig. 2), or only the effect or response could be patterned (top right and bottom left of Fig. 2), resulting in observable differences in overlapping distribution. These differences can provide insight into the contrasting drivers of species assembly for interacting clades, for instance, the phylogenetic patterns of co-occurring hosts may be a strong driver of parasite community assembly, while host assembly may be predominantly driven by environmental limits and within-clade dynamics. This method therefore builds on within-clade regional environmental and local-scale species assembly models to enable researchers to estimate the drivers and future of cross-clade interactions.

Central to the concept of the cross-clade environment is the idea that species evolve to match the environment of other species, and vice versa. Species' traits drive many interactions, including those among predators and prey, parasites and their hosts, and mutualists (Werner and Peacor 2003). The PGLMM method improves our ability to test the effect of traits on species co-occurrence (as for species assembly) through comparison among models describing separate (for each species) and combined regional and local trait- and environment-based drivers, employing the same model transformations described in sections above to test for constrained, repulsive and neutral evolutionary process. For instance, this framework can be used to test if species' interactions thought to drive plant structure are evolutionarily conserved or divergent (Cavender-Bares et al. 2009), or whether species' interactions are too diffuse to permit co-evolution (Zillio et al. 2005). Inherent to this analysis is the need for sufficient spatial breadth of species assemblage data to capture both regional and local drivers, as well as a broad sample of where species do and do not co-occur.

Applying the PGLMM framework to current and future conservation challenges

Phylogeny has already proven an informative tool for conservation planning (Isaac et al. 2007), restoration (Hipp et al. 2015) and trait prediction (Swenson et al. 2017). PGLMM, applied to cross-scale data, adds an additional dimension with which to formulate niche-based conservation approaches as related to environmental response, competitive interactions and cross-clade associations. For instance, the framework we outline above can be used to rank species in terms of their likelihood of successful colonization and/or interaction with other members of a community. Below we outline some of the ways in which such a prioritization could be produced and used.

Environmental responses (β-traits)

PGLMM allows for the prediction of whether rare, understudied or invasive species could colonize a region on the basis of phylogeny, with applications for restoration and invasive species management. For instance, if species' environmental responses are phylogenetically patterned, or if closely related species tend to (or tend not to) co-occur, a site's environmental characteristics and current species assemblage can be used to assess how successful an introduced or invasive species will be there. Using these tools, the impact of invasive species could be predicted even before species have arrived on a continent.

Competitive interactions (α-traits)

Darwin's theory of naturalization predicts that invasive species are less likely to establish near native congeneric species

due to higher competition and vulnerability to local pests and pathogens (Daehler 2001). Thus, in observable communities invasives species are expected to be both phylogenetically and functionally distant from native neighbors, however this expectation has proven somewhat controversial in practice (Thuiller et al. 2010). The PGLMM framework can be used to test the theory of naturalization across a broad range of communities and taxa. Furthermore, using freely available functional trait and phylogenetic (or at least taxonomic) data, and the estimated parameters from the PGLMM described above, it is possible to rank the likelihood of species successfully invading sites to help managers plan for invasions.

Cross-clade associations

Joint evolutionary and ecological modeling of interaction networks could aid in the prediction of pest and pathogen outbreaks, vulnerable mutualisms and shifting trophic dynamics. PGLMM can be used to determine whether species have co-diversified or whether similar environmental responses have caused them to associate. In the former case, we might expect limited host-switching and spill-over under climate change, whereas in the latter case we might expect more. Previous research identifying phylogenetic patterns in spillover among hosts (Parker et al. 2015) indicates the potential for eco-phylogenetic tools to combine the phylogenetic likelihood of host spillover with the ecological limitations of pests and pathogens as well as their hosts to predict future associations. Understanding when species are evolutionarily constrained not to co-occur will allow efforts to be targeted at those species who are only held back by ecological opportunity, which, in a changing world, may not be the case for long.

Conclusion

PGLMM has the capacity to integrate evolutionary and ecological tools, theory and their differing strengths to identify cross-scale drivers of species assembly. In macro-evolutionary biology, the patterns of species' traits fit statistical models well, but determining underlying mechanisms is hard, we simply cannot run evolutionary experiments over millions of years. A strength of ecology, on the other hand, is the ability to examine mechanisms at the local scale, but that scale limits the potential to address underlying generative processes (Peters et al. 2014). PGLMM has the unique potential to bridge these gaps, strengthening evolutionary techniques by requiring more cross-scale data to test mechanisms, and broadening ecological scope with the generality required to predict across regions. Opportunities to use PGLMM to integrate across evolution and ecology will almost certainly continue to expand as the availability of occurrence and co-occurrence data increases over space, time and taxa. Increasingly sophisticated macro-evolutionary models are only going to need more data with which to test their predictions, and as ecology becomes more predictive the temporal and spatial scale over which models will need to operate can only grow. PGLMM provides a way forward to merge both fields for mutual benefit, but will require that statistical and data advances are accompanied by conceptual breakthroughs and practical progress toward improved conservation and management.

Data availability statement

No new data are released as part of this manuscript. R code for Bayesian PGLMM is included in the Supporting information.

Author contributions

Amanda S. Gallinat: Conceptualization (equal); Software (supporting); Visualization (equal); Writing – original draft (lead); Writing – review and editing (equal). **William D. Pearse**: Conceptualization (equal); Funding acquisition (lead); Software (lead); Visualization (equal); Writing – original draft (supporting); Writing – review and editing (equal).

References

Ackerly, D. D. et al. 2006. Niche evolution and adaptive radiation: testing the order of trait divergence. – Ecology 87: S50–S61.

Anderson, T. M. et al. 2004. The relationship of phylogeny to community structure: the cactus yeast community. – Am. Nat. 164: 709–721.

Blomberg, S. P. et al. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. – Evolution 57: 717–745.

Brown, W. L. and Wilson, E. O. 1956. Character displacement. – Syst. Zool. 5: 49–64.

Butler, M. A. and King, A. A. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. – Am. Nat. 164: 683–695.

Cadotte, M. W. and Tucker, C. M. 2017. Should environmental filtering be abandoned? – Trends Ecol. Evol. 32: 429–437.

Cavender-Bares, J. et al. 2009. The merging of community ecology and phylogenetic biology. – Ecol. Lett. 12: 693–715.

Chesson, P. 2000. Mechanisms of maintenance of species diversity. – Annu. Rev. Ecol. Syst. 31: 343–366.

Connell, J. H. 1980. Diversity and the coevolution of competitors, or the ghost of competition past. – Oikos 35: 131–138.

Cornwell, W. K. and Ackerly, D. D. 2009. Community assembly and shifts in plant trait distributions across an environmental gradient in coastal California. – Ecol. Monogr. 79: 109–126.

D'Andrea, R. and Ostling, A. 2016. Challenges in linking trait patterns to niche differentiation. – Oikos 125: 1369–1385.

Daehler, C. C. 2001. Darwin's naturalization hypothesis revisited. – Am. Nat. 158: 324–330.

Drury, J. et al. 2016. Estimating the effect of competition on trait evolution using maximum likelihood inference. – Syst. Biol. 65: 700–710.

Edwards, J. L. et al. 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. – Science 289: 2312–2314.

Ezard, T. H. et al. 2011. Interplay between changing climate and species ecology drives macroevolutionary dynamics. – Science 332: 349–351.

Faurby, S. and Svenning, J.-C. 2015. A species-level phylogeny of all extant and late Quaternary extinct mammals using a novel heuristic-hierarchical Bayesian approach. – Mol. Phylogenet. Evol. 84: 14–26.

Freckleton, R. P. et al. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. – Am. Nat. 160: 712–726.

Hembry, D. H. and Althoff, D. M. 2016. Diversification and coevolution in brood pollination mutualisms: windows into the role of biotic interactions in generating biological diversity. – Am. J. Bot. 103: 1783–1792.

Hipp, A. L. et al. 2015. Phylogeny in the service of ecological restoration. – Am. J. Bot. 102: 647–648.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. – Princeton Univ. Press.

Isaac, N. J. et al. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. – PLoS One 2: e296.

Ives, A. R. and Helmus, M. R. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. – Ecol. Monogr. 81: 511–525.

Karan, M. et al. 2016. The Australian SuperSite Network: a continental, long-term terrestrial ecosystem observatory. – Sci. Total Environ. 568: 1263–1274.

Kraft, N. J. et al. 2015. Community assembly, coexistence and the environmental filtering metaphor. – Funct. Ecol. 29: 592–599.

Li, D. et al. 2020. phyr: an R package for phylogenetic species-distribution modelling in ecological communities. – Methods Ecol. Evol. 11: 1455–1463.

Li, D. and Ives, A. R. 2017. The statistical need to include phylogeny in trait-based analyses of community composition. – Methods Ecol. Evol. 8: 1192–1199.

Losos, J. B. 2011. Convergence, adaptation and constraint. – Evolution 65: 1827–1840.

Matsunaga, A. et al. 2013. A computational-and storage-cloud for integration of biodiversity collections. – In: 2013 IEEE 9th International Conference on e-Science. IEEE, pp. 78–87.

Mayfield, M. M. and Levine, J. M. 2010. Opposing effects of competitive exclusion on the phylogenetic structure of communities. – Ecol. Lett. 13: 1085–1093.

McGill, B. J. 2019. The what, how and why of doing macroecology. – Global Ecol. Biogeogr. 28: 6–17.

Mirtl, M. et al. 2018. Genesis, goals and achievements of long-term ecological research at the global scale: a critical review of ILTER and future directions. – Sci. Total Environ. 626: 1439–1462.

Moeslund, J. E. et al. 2013. Topographically controlled soil moisture drives plant diversity patterns within grasslands. – Biodivers. Conserv. 22: 2151–2166.

Morlon, H. et al. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. – Methods Ecol. Evol. 7: 589–597.

Nemergut, D. R. et al. 2013. Patterns and processes of microbial community assembly. – Microbiol. Mol. Biol. Rev. 77: 342–356.

Nuismer, S. L. and Harmon, L. J. 2015. Predicting rates of inter-specific interaction from phylogenetic trees. – Ecol. Lett. 18: 17–27.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. – Nature 401: 877–884.

Parker, I. M. et al. 2015. Phylogenetic structure and host abundance drive disease pressure in communities. – Nature 520: 542–544.

Pearse, W. D. et al. 2018. Building up biogeography: pattern to process. – J. Biogeogr. 45: 1223–1230.

Pearse, W. D. et al. 2015. Pez: phylogenetics for the environmental sciences. – Bioinformatics 31: 2888–2890.

Pearse, W. D. et al. 2014. Metrics and models of community phylogenetics. – In: Garamszegi, L. Z. (ed.), Modern phylogenetic comparative methods and their application in evolutionary biology – concepts and practice. Springer.

Pearse, W. D. et al. 2019. The interaction of phylogeny and community structure: linking the community composition and trait evolution of clades. – Global Ecol. Biogeogr. 28: 1499–1511.

Peters, D. P. et al. 2014. Taking the pulse of a continent: expanding site-based research infrastructure for regional-to continental-scale ecology. – Ecosphere 5: 1–23.

Poisot, T. et al. 2015. Beyond species: why ecological interaction networks vary through space and time. – Oikos 124: 243–251.

Rabosky, D. L. et al. 2011. Species interactions mediate phylogenetic community structure in a hyperdiverse lizard assemblage from arid Australia. – Am. Nat. 178: 579–595.

Rafferty, N. E. and Ives, A. R. 2013. Phylogenetic trait-based analyses of ecological networks. – Ecology 94: 2321–2333.

Revell, L. J. et al. 2008. Phylogenetic signal, evolutionary process and rate. – Syst. Biol. 57: 591–601.

Richardson, J. L. et al. 2014. Microgeographic adaptation and the spatial scale of evolution. – Trends Ecol. Evol. 29: 165–176.

Ricklefs, R. E. 2008. Disintegration of the ecological community. – Am. Nat. 172: 741–50.

Rosindell, J. et al. 2015. Unifying ecology and macroevolution with individual-based theory. – Ecol. Lett. 18: 472–482.

Sauer, J. R. et al. 2013. The North American breeding bird survey 1966–2011: summary analysis and species accounts. – N. Am. Fauna 79: 1–32.

Schimel, D. et al. 2007. NEON: a hierarchically designed national ecological network. – Front. Ecol. Environ. 5: 59.

Schluter, D. and McPhail, J. D. 1992. Ecological character displacement and speciation in sticklebacks. – Am. Nat. 140: 85–108.

Silvertown, J. et al. 2015. Hydrological niches in terrestrial plant communities: a review. – J. Ecol. 103: 93–108.

Silvertown, J. et al. 2006. Phylogeny and the hierarchical organization of plant diversity. – Ecology 87: S39–S49.

Smith, S. A. and Walker, J. F. 2019. Py phlawd: a python tool for phylogenetic dataset construction. – Methods Ecol. Evol. 10: 104–108.

Smith, W. B. 2002. Forest inventory and analysis: a national inventory and monitoring program. – Environ. Pollut. 116: S233–S242.

Sullivan, B. L. et al. 2014. The eBird enterprise: an integrated approach to development and application of citizen science. – Biol. Conserv. 169: 31–40.

Swenson, N. G. et al. 2017. Phylogeny and the prediction of tree functional diversity across novel continental settings. – Global Ecol. Biogeogr. 26: 553–562.

Swenson, N. G. 2013. The assembly of tropical tree communities – the advances and shortcomings of phylogenetic and functional trait analyses. – Ecography 36: 264–276.

Swenson, N. G. et al. 2006. The problem and promise of scale dependency in community phylogenetics. – Ecology 87: 2418–2424.

Thuiller, W. et al. 2010. Resolving Darwin's naturalization conundrum: a quest for evidence. – Divers. Distrib. 16: 461–475.

Uyeda, J. C. et al. 2015. Comparative analysis of principal components can be misleading. – Syst. Biol. 64: 677–689.

Uyeda, J. C. and Harmon, L. J. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. – Syst. Biol. 63: 902–918.

Uyeda, J. C. et al. 2018. Rethinking phylogenetic comparative methods. – Syst. Biol. 67: 1091–1109.

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

Webb, C. O. et al. 2002. Phylogenies and community ecology. – Annu. Rev. Ecol. Syst. 33: 475–505.

Weber, M. G. et al. 2017. Evolution in a community context: on integrating ecological interactions and macroevolution. – Trends Ecol. Evol. 32: 291–304.

Werner, E. E. and Peacor, S. D. 2003. A review of trait-mediated indirect interactions in ecological communities. – Ecology 84: 1083–1100.

Zillio, T. et al. 2005. Spatial scaling in model plant communities. – Phys. Rev. Lett. 95: 098101.