Project Description

Overview & objectives

Integrating the cross-scale drivers of the distribution of biodiversity requires an understanding of both ongoing ecology and historic evolution (Levin 1992; Chave 2013). Species' shared evolutionary history reflects the biogeographical history that shaped them, and can inform us of constraints on species' ecology in the present (Webb *et al.* 2002; Cavender-Bares *et al.* 2009). **Here I propose to integrate local-scale ecological community assembly with macro-scale regional assembly and trait evolution**. Evolution is the context within which ecological processes take place (Hutchinson 1965): there can be no macro-scale synthesis of ecology without addressing phylogenetic context. I propose to extend previous analyses of the evolution of species' environmental tolerances to test macro-evolutionary models of local-scale competition (Objective 2) and cross-guild interactions (Objective 3). While doing so, I will develop and release new datasets (Objective 1) and improve engagement in groups under-represented in STEM (Objective 4).

Objective 1: Assemble phylogenetic, functional trait, and assemblage data for focal plant, beetle, tick, mammal, and bird species. I will generate plant and beetle phylogenies for study species, and collect new functional trait data for focal plants, beetles, and ticks. Assemblage data will come from NEON (2003), FIA (2017), BBS (Sauer *et al.* 1966), and Thibault *et al.* (2011). These data will support my other objectives and be released as data products.

Objective 2: Quantify the roles of local-scale competition, regional environmental filtering, and macro-evolution of species' traits in ecological assembly. Ecological assemblages result from historical biogeography, contemporary environmental filtering, and local-scale competitive interactions. I will measure the influence of each of these three processes, and use these insights to predict future site compositions and highlight potential invasive species.

Objective 3: Contrast the importance of interactions between guilds in determining the distribution of species. There is a growing recognition that macro-scale processes can inform plant–herbivore and host–parasite dynamics. Building on the results from Objective 2, I will examine how species' traits and phylogenetic history affect species co-occurrences and interactions. These results will inform predictions of pest and pathogen outbreaks.

Objective 4: Use NEON sites as a platform to engage under-served student populations in modern eco-informatic research. Native Americans make up 70% of the student body at Utah State's Blanding campus, but getting these students to complete four-year STEM programs has been challenging. I will run an ecological data science course targeted at these students, combined with a trip to the nearby Moab NEON site and an independent study project. This yearly program will increase the STEM skills of students.

Conceptual background

Macro-scale community ecology. Biologists have long-recognized that present-day ecological processes operate within the context of past evolutionary history (Hutchinson 1965; Levin 1992; Ricklefs 2010; Chave 2013). For example, in biogeography, great progress has been made quantifying the extent to which the evolutionary origin of species affects the ecological niche(s) and distributions species occupy today (*e.g.*, Wiens & Donoghue 2004; Donoghue 2008; Albert *et al.* 2017; Cabral *et al.* 2017), but progress in community ecology has been more mixed. Community ecologists have tried to incorporate evolutionary dynamics through eco-phylogenetic (or community phylogenetic) analyses (Webb *et al.* 2002; Cavender-Bares *et al.* 2009; Pearse *et al.* 2014), but the field has had difficulty moving from correlative studies to generate mechanistic insights (discussed in Mayfield & Levine 2010; Swenson 2013). To address this challenge, we need a synthetic framework that integrates differing mechanisms of the macro-evolution of species' traits with macro- and local-scale assembly processes based upon those traits. Below I describe how regional-scale environmental filtering (β traits), local-scale competitive assembly (α traits), and cross-guild interactions, can be integrated with models of trait evolution within the eco-phylogenetic 'PGLMM' framework (described below; Helmus *et al.* 2007; Ives & Helmus 2011; Pearse *et al.* 2014; Pearse *et al.* 2015). This conceptual framework is laid out in figure 1.

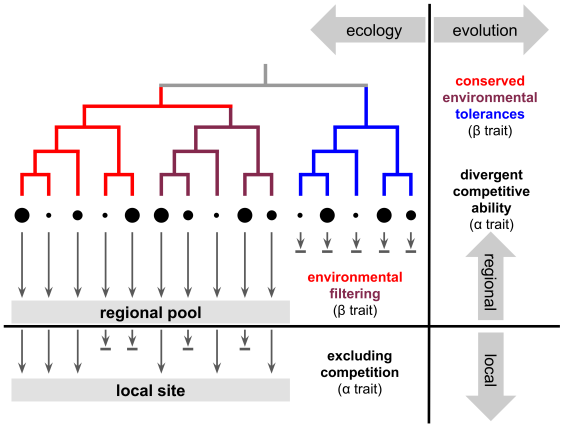

Figure 1: **Conceptual overview of the evolutionary and ecological processes governing ecological assembly.** This diagram partitions processes into four quadrants along two axes: evolution–ecology and regional–local. It shows one example of how the conserved evolution of a β trait (the 'red/purple/blue' colors) drives environmental filtering within one region. Competition on the more divergent α trait (the black circles) determines the assembly of local sites from the regional pool of species. Different regional conditions would lead to a different regional pool, and different patterns of trait evolution (*e.g.*, divergent β evolution and constrained α evolution) would lead to different regional and local-scale patterns.

Conserved evolution of continental-scale environmental filtering—β traits. Eco-phylogeneticists have distinguished between species' β-traits that determine regional habitat-affiliations, and species' α-traits that drive local-scale co-existence and competition (Ackerly *et al.* 2006; Silvertown *et al.* 2006). As I discuss below (and show in figure 1), by using this framework it is possible to measure how species' trait evolution constrains macro-scale and local processes of ecological assembly. β-traits, which regulate species' tolerances to environmental filters, are hypothesized to evolve earlier and be strongly conserved through time (Cavender-Bares *et al.* 2009). Examining such traits requires a combination of macro- and micro-scale data, since the composition of nearby sites rarely encompasses sufficient environmental variation. Studies of small assemblages, or large assemblages within small regions, are insufficient to detect both environmental filtering and the evolutionary signal of environmental filtering (see also Swenson *et al.* 2006; Kraft & Ackerly 2010). For example, Cavender-Bares *et al.* showed, in a series of influential papers focusing on oaks (2004; 2006), that the evolutionary and ecological signature of filtering on functional traits was obscured at smaller spatial scales. However, correlating the phylogenetic and trait structure of

ecological assemblages cannot distinguish between different processes of trait evolution (Swenson 2013). Thus, an integrated model of both trait macro-evolution and ecological assembly is required to test if the β-traits driving regional assembly evolved early and have been strongly conserved.

Divergent evolution of local-scale competition—α traits. The traditional picture of excluding competition driving local-scale ecological assembly has been complicated by a growing understanding of the role of Neutral processes (Hubbell 2001; Vellend 2010). Under excluding competition, species' similarities, measured as α-traits, make species less likely to co-occur, and so under selection the expectation is that species' traits will diverge through repulsive evolution (Cavender-Bares *et al.* 2009; Nuismer & Harmon 2015). This complication, which makes it difficult to infer ecological pattern from eco-phylogenetic structure (Mayfield & Levine 2010; Gerhold *et al.* 2015), means that we must test for α-trait competition both ecologically—dissimilar species tend to co-occur—and evolutionarily—species traits evolve to to minimize similarity. This will also allow us to distinguish between two classes of Neutral theory: (i) ecological assembly within guilds is random with respect to species' traits (Hubbell 2001; Rosindell *et al.* 2011), and (ii) traits do matter but the evolution of those traits are themselves random (Pigot & Etienne 2015; Rosindell *et al.* 2015). Such competition is difficult to detect without accounting for the regional context within which it operates (see figure 1).

The evolution of cross-guild interactions. The α- and β-trait based model of ecological assembly described above, like many filtering-based models of community assembly (Vellend 2010), ignores interactions across trophic levels. The co-diversification of plants and their beetle pollinators and herbivores (Farrell 1998; Hunt *et al.* 2007), along with mammals and their parasites (Barker & Murrell 2003), have been rich and controversial fields of study. 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 interactions among these clades (but see Ives & Godfray 2006; Rezende *et al.* 2007; Eklöf *et al.* 2012; Rafferty & Ives 2013). This is despite concerns over outbreaks of novel, invasive beetles, whose potential impacts on forestry are large but our data for prediction limited (Bentz *et al.* 2010). Phylogenetically-informed models could be used to make predictions for those species about which we currently know very little other than their phylogenetic placement (Mace *et al.* 2003). Modeling interactions among species is fraught with difficulty (Bascompte & Jordano 2007; Vázquez *et al.* 2009), but modeling co-occurrences need not be. Each group (*e.g.*, plants and beetles) assembles through the regional and local-scale processes described above; the simplest model of co-occurrence would simply describe a statistical interaction between each taxonomic group's parameters. In the case of parasites and their hosts, this would provide a first step towards integrating macro-ecology into parasite biology (Stephens *et al.* 2016), and suggests a way to account for variation in interaction specificity through time and space (Poisot *et al.* 2015). It also provides a framework 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).

PGLMM as a synthetic hypothesis-testing framework. Among the 'jungle' of eco-phylogenetic approaches (Pausas & Verdú 2010), there are few frameworks capable of addressing the evolution of species' traits, environmental filtering, competition, and species' interactions. The Phylogenetic Generalized Linear Mixed Model (PGLMM; Ives & Helmus 2011; Rafferty & Ives 2013; Pearse *et al.* 2015) has emerged as a flexible (*e.g.*, Kaldhusdal *et al.* 2015) way to distinguish among the various complex processes that drive ecological assembly. For example, PGLMM has been used to tease apart the macro-scale processes of dispersal, *in situ* radiation, and environmental filtering within Madagascan lemurs (Herrera 2016). A full description of PGLMM is given in Ives & Helmus (2011), but I give a non-technical description below. A PGLMM quantifies species' occurrences or abundances as a function of data on species' traits and environmental conditions. As an extension of Generalized Linear Mixed

Models (GLMM; reviewed in Bolker *et al.* 2009), a PGLMM uses random effect terms to account for idiosyncratic site- and species-specific variation. Critically, PGLMM can use these random effects to measure phylogenetically-patterned differences in how species interact with each other and their environment. Thus, PGLMM can go beyond describing the structure of assemblages (*e.g.*, "*they contain closely-related species that are dissimilar*") and describe the processes structuring assemblages (*e.g.*, "*divergent trait evolution, followed by competition on that divergent trait, structures these communities*"). Therefore, when correctly specified, a PGLMM can quantify the contributions of past trait evolution, and regional and local ecological processes, to present-day ecological assembly (as outlined in figure 1).

Modeling continental-scale ecological assembly. Above, I described a model of regional-scale environmental assembly and subsequent competitive exclusion (see figure 1). Testing this framework will advance theory about how the cross-scale processes that govern ecological community assembly evolved. There are two challenges to testing this framework: (i) the scale of the data required and (ii) the complexity of fitting models to such data. The NEON data are perfectly placed to integrate across other datasets, as they cover a wide range of regions (and so environmental conditions) and taxa. I feel that PGLMM approaches, whose computational development I have helped (Pearse *et al.* 2015), are now capable of tackling the analytical challenge presented by this framework.

Description of Research and Education Activities

I will model the evolutionary and ecological controls on the structure of North American local communities, building datasets and techniques to help other researchers, targeting Native American groups with outreach activities. I will focus on plants, beetles, birds, and mammals (and their tick parasites), using NEON sites as comparison points to integrate these diverse taxa taxon-specific datasets based around them. I will produce phylogenies for the plants and beetles of NEON (other groups already have adequate phylogenies), assemble novel functional trait databases for all of my study taxa, and integrate NEON with other ecological assemblage data (Objective 1). Using these data, I will model the assemblage structure of species across North America, and predict future compositional change (Objective 2). In Objective 3, I will modify these models and predictions to take into account plant–beetle and mammal–tick associations. This will inform and improve predictions of disease outbreaks in ecosystems, such as forests. Finally, I will reach out to under-served Native American student populations, helping them develop STEM skills (Objective 4).

Objective 1: Assemble phylogenetic, functional trait, and assemblage data

Modeling species' ecology or evolution requires knowing something about those species. To support my other objectives, I will develop detailed phylogenetic, functional trait, and assemblage databases of the plants, beetles, ticks, mammals, and birds of North America. I will openly release these data, and images used to create them, to catalyze further research.

Plants:

Assemblages. NEON (Council *et al.* 2003) collects plant percentage cover data at all core sites, and I will integrate this with Forest Inventory and Analysis (FIA 2017) data (covering over 325,000 sites). I will use data collected within the last ten years of the FIA to ease comparison with NEON data. All data will be taxonomically standardized (using Cayuela *et al.* 2012).

Phylogeny. There are ~10,200 species and ~2,800 genera recorded at NEON (recorded within ±0.2° of sites according to the Global Biodiversity Information Facility) and FIA sites. Of these species,

~5,600 species and ~2,000 genera have DNA barcode data (*rbcL* or *matK*; Hollingsworth *et al.* 2009) on *GenBank* (Benson *et al.* 2013), as found in a preliminary *phyloGenerator2* run (Pearse & Purvis 2013). To construct a phylogeny from these sequences (and additional ones NEON plans to collect), I will use *phyloGenerator2* (Pearse & Purvis 2013) to download, choose, and align sequences for species from the *rbcL*, *matK*, *atpB*, *ndhF*, *psbBTNH*, *rpoC2*, *rps16*, and *rps4* loci (following Zanne *et al.* 2014). Using these alignments, trees will be built using *ExaML* (similar to *RAxML* but optimized for large datasets; Stamatakis 2006; Kozlov *et al.* 2015), and then dated using a *BEAST* (Bouckaert *et al.* 2014) search on the alignment constrained on the topology found by *ExaML* (following Pearse & Purvis 2013). If this *BEAST* approach proves too computationally challenging, I will date the tree using parametric rate smoothing (Sanderson 2002; Sanderson 2003) with *treePL* (Smith & O'Meara 2012). Fossil calibrations for all dating will be taken from Smith *et al.* (2010) and Clarke *et al.* (2011). Both the ExaML and BEAST approaches will reveal a single 'best' phylogeny and sets of candidate phylogenies that can be used to account for phylogenetic uncertainty (see Huelsenbeck *et al.* 2001; Bollback 2005).

Traits. Much of plant functional ecology has focused on leaf, stem, and root traits (Reich 2014): I will gather Specific Leaf Area (SLA; leaf trait), flow conductivity (stem trait), and maximum rooting depth data for the species from the TRY plant database (Kattge *et al.* 2011). In the June 2014 species list to which I have access, TRY contained data on ~5,500 of the ~10,200 species, and ~2,120 of the ~2,800 plant genera, of the potential study species (see above). These TRY data will be used to carry out objectives 2 and 3, but TRY data cannot be released as part of a data package for use by others. To fulfill Objective 1 and create trait datasets for others, I will compile comparable data from other sources (requesting consent and co-authorship; Wright *et al.* 2004; Kleyer *et al.* 2008; Chave *et al.* 2009; Zanne *et al.* 2014) and release this second, public trait dataset. I will also compile a NEON-specific leaf functional trait dataset, measuring length, width, area, perimeter, and dry mass, and from these, calculate SLA. NEON will digitally scan and weigh leaves from the herbarium collections of each region (following standard protocols; Cornelissen *et al.* 2003) and email these data to me. I will then quantify the above leaf functional traits using the *stalkless* pipeline (Pearse *et al.* 2016), and release the raw images through *iDigBio* (see "*Data Management Plan*").

Beetles:

Assemblages. NEON (Council *et al.* 2003) collects beetle pitfall trap data at all core sites; as with the ticks, I am unaware of another macro-scale beetle assemblage dataset (beyond NEON).

Phylogeny. ~510 of the ~1450 species recorded as within ±0.2° of the NEON sites on GBIF have DNA barcode (*COI*) data available on *GenBank* (coverage calculated as for plants above). Additionally, the NEON team are sequencing *COI* data for beetles at the sites, and ~1250 sequences are already available through the Barcode of Life initiative (BOLD). This coverage is sufficient for objectives 2 and 3, and I will use the same protocol described for the plants (above) to download, align, and build a phylogeny from these data. To date the phylogeny, I will use the fossil calibrations and clade ages given in Trautwein *et al.* (2012) and references therein.

Traits. I know of no beetle trait databases with coverage comparable to TRY, and so I will create a trait database for the beetles in NEON. A recent review (Fountain-Jones *et al.* 2015) split beetle traits into morphological/physiological traits and effect/response traits (*sensu* Díaz *et al.* 2013). Of these traits, head width (morphological–response), leg length (morphological–effect) can be quantified from images of specimens (and others; Fountain-Jones *et al.* 2015), while food preferences (physiological–effect/response) can be inferred from taxonomy or literature. At least one specimen of each species will be shipped from the NEON beetle repositories to Utah State University, where they will be digitized and then returned to NEON's repositories. I will quantify the above functional traits from these images using ImageJ (Abràmoff *et al.* 2004), and use Fourier-based morphometric principal component analyses to

split the beetles into broad functional groups based on their overall shape (using *stalkless* and *Momocs*; Kuhl & Giardina 1982; Rohlf & Archie 1984; Bonhomme *et al.* 2014; Pearse *et al.* 2016).

Mammals:

Assemblages. NEON (Council *et al.* 2003) collects rodent trap data at all core sites, which will be augmented with data from 694 rodent trap sites distributed across the US (Thibault *et al.* 2011). All data will be taxonomically standardized (using Chamberlain & Szöcs 2013).

Phylogeny. The phylogeny of global mammals is reasonably uncontroversial at the family level (Bininda-Emonds *et al.* 2007), and I will use an existing fully-resolved phylogeny for all mammals (with posterior distributions to account for uncertainty; Faurby & Svenning 2015).

Traits. Mammal trait data are readily available for the overwhelming majority of species. I will take basic life-history traits such as body mass and length from Jones *et al.* (2009), and Eltonian aspects of niche, such as diet and foraging attributes, from Wilman *et al.* (2014).

Birds:

Assemblages. NEON (Council *et al.* 2003) collects bird count data at all core sites, which will be augmented with data from 4100 transect surveys conducted by the Breeding Bird Survey (BBS; Sauer *et al.* 1966). Only data collected within the last five years of the BBS will be used to aid comparison with NEON, and all data will be taxonomically standardized (using Chamberlain & Szöcs 2013).

Phylogeny. I will use the Jetz *et al.* (2014) global bird phylogeny, which has posterior distributions to account for uncertainty. Genomic data suggests the placement of some major bird clades is unclear; for eco-phylogenetic analysis deep phylogenetic structure is relatively uninformative (Letten & Cornwell 2015) and so this will not affect objectives 2 and 3.

Traits. I will take body mass and diet and foraging attributes from the global Wilman *et al.* (2014) dataset. If necessary, I will take additional life-history attributes from the Cornell Lab of Ornithology's "*Birds of North America*" website.

Ticks:

Interactions. NEON (Council *et al.* 2003) collects mammal-tick association data at all core sites; as with the beetles, I am unaware of a comparable macro-scale interaction dataset (beyond NEON).

Phylogeny. The taxonomy of ticks that occur on mammals is reasonably stable, but the same is not true of all genera (and so species; see Barker & Murrell 2004; Nava *et al.* 2009, for reviews). Only 15 tick species are recorded within ±0.2° of the NEON sites on GBIF; this almost certainly reflects poor sampling and so is insufficient to assess DNA coverage. Therefore, I will not make a phylogeny of the ticks found on mammals at NEON, and one is not required for the objectives in this proposal.

Traits. I am not aware of a tick trait database with broad coverage of the taxa in this proposal, and so I will create a trait database for the ticks in NEON. Specifically, I will ship at least one specimen of each species from the NEON tick repository to Utah State University, digitize and then return them. Following the methods outlined for the beetles (above), I will measure tick body and leg length, and use the overall shape of the ticks to split them into broad functional groups.

Feasibility and risk mitigation:

I have estimated the phylogenetic and trait coverage of each group, and will collect more plant, beetle, and tick data. A complete phylogeny is not necessary for meaningful inference: missing species can be added using taxonomy (Pearse *et al.* 2015) and randomly incomplete phylogenies do not bias inference (given reasonable caveats; see Purvis 2008; O'Meara 2012; Rabosky 2015). The potential impact of major taxonomic revisions is discussed above for each group, and using beetle DNA sequences

taken from vouchered NEON specimens mitigates against species re-definitions. Merging assemblage datasets as large as these is challenging, but I have successfully analyzed larger datasets (*e.g.*, Pearse *et al.* in press) and this proposal does request funds for High Performance Computing nodes.

Products from Objective 1:
- The first resolved and dated phylogenies of the NEON plant and beetle species.
- Functional trait datasets and high-resolution images for plants, beetles, and ticks.

Objective 2: Model the evolutionary and ecological drivers of ecological assembly and thus predict responses to environmental change and invasive species

I propose to model how species' communities result from competitive, trait-based interactions operating within the context of regional-scale environmental filtering. Through the use of explicit phylogenetically-informed models, I will measure the macro-evolution of species' traits and trait-based responses that determine these regional and local-scale ecological assembly processes. Below, I describe how I will integrate each of these three components—regional environmental filtering, competition, and macro-evolution—into a single cross-scale model. Finally, I describe how these models will inform ecological predictions of geographic regions and species for which I do not have data.

Data. The plant, beetle, mammal, and bird data from Objective 1 will be used for Objective 2, compiled into a GIS database with regional-scale temperature, precipitation, and aridity data (Hijmans *et al.* 2005; Trabucco & Zomer 2009) for each site. While the mammal, bird, and plant datasets come from different locations, the NEON sites are common to all taxa and so will unite these cross-taxon analyses.

Hypotheses. Environmental filtering of phylogenetically conserved β-traits is thought to drive compositional change across environmental gradients (Webb *et al.* 2002; Cavender-Bares *et al.* 2009; Vamosi *et al.* 2009; Helmus *et al.* 2010), but the traits underlying these changes are often unknown (Winter *et al.* 2013; Gerhold *et al.* 2015). I will address this challenge by (1) measuring trait change across a temperature/aridity gradient in the sites, and (2) testing for phylogenetic pattern in both species' traits and trait-based responses to the gradient. Figure 2 depicts such β-trait-based regional environmental filtering across an environmental gradient under different kinds of trait evolution. Below, I predict functional traits likely to drive compositional shifts across the aridity/temperature gradients in the sites. In plants, major macro-scale environmental drivers of distributions are temperature and aridity (Currie 1991), and greater leaf perimeter:area (Givnish & Vermeij 1976; Sack *et al.* 2003) and rooting depth (Lynch 1995; Reich *et al.* 2003; Silvertown *et al.* 2015) are associated with tolerance of aridity and heat. Thermal constraints in mammals (Ashton *et al.* 2000) and birds (Ashton 2002) are thought to drive a tendency for smaller-bodied species in hotter regions (resulting in '*Bergmann's Rule*'; see Meiri & Dayan 2003). These constraints may (or may not; Mousseau 1997) apply more severely to beetles because they are endothermic, although a recent review (Chown & Gaston 2010) suggested that there are too few insect studies to be certain of this (Diniz-Filho & Fowler 1998; Moreeteau *et al.* 2003). While there have been studies of the regional-scale distributions of species' traits, few (if any) take a cross-scale approach to also consider local-scale community composition (McNab 1971; Beck *et al.* 2012). This is despite indications of scale-dependent variation in trait-based beetle assembly (Holland *et al.* 2005). **Regional-scale filtering—β-traits:**

Analysis. I will use Phylogenetic Generalized Linear Mixed Models (PGLMM; Ives & Helmus 2011, see '*Conceptual background*') to test the above hypotheses that species' traits drive regional assembly across the temperature/aridity gradient in the study sites. I will fit three models for each taxonomic group: (i) a *null model* with random effect terms for each species and site, (ii) an *additive model* with the same random terms and additive fixed effects for species' traits and site conditions (temperature or aridity), and (iii) an *interaction model* with all of the above terms and an interaction between species' traits and site conditions. A Likelihood Ratio Test (LRT) of $\frac{additive}{null}$ will reveal if each species' richness/abundance varies across the environmental gradient, and what traits are associated with more widespread/abundant species. This test will show, for example, if smaller beetles are more abundant, or warmer sites more diverse. $LRT(\frac{interaction}{additive})$ will test my key hypothesis of regional filtering: whether species' traits *interact* with the environment, such as if smaller beetles are more abundant at warmer sites. If support for regional filtering is found, a final key test will be to determine whether regional filtering explains more variation than under the null expectation of site- and species-specific variation [*i.e.*, $LRT(\frac{interaction}{null})$]. The random effect structure of PGLMM is key to distinguishing among idiosyncratic site- or species-level differences and systematic changes driven by regional filtering. In the section '*macro-evolution of species' traits and responses*' below, I describe how I will model the evolution of these regional responses (answering question 2 above).

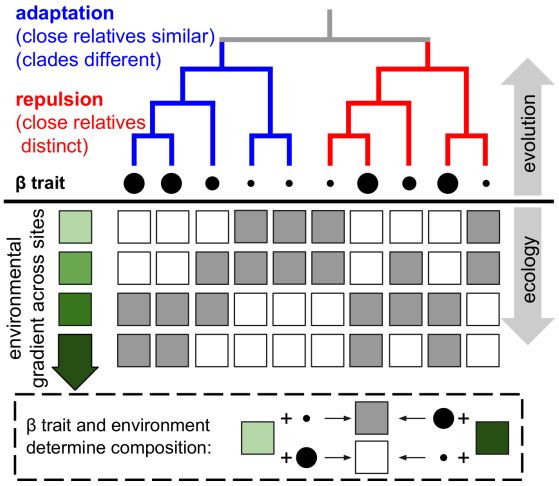

Figure 2: **Conceptual overview of regional-scale and macro-evolutionary hypotheses tested in Objective 2.** Ten species (columns) and their presence in four sites along an environmental gradient (rows) are shown. The match of species' β traits (circles) to the environment (green boxes)—environmental filtering (Kraft *et al.* 2015)—determines community membership. The β-traits themselves are shown as they would be if evolving under two distinct macro-evolutionary models: divergence/repulsion (the red clade on the right) and adaptation (the blue clade on the left; see Uyeda & Harmon 2014).

Local-scale competitive interactions—α-traits:

Hypotheses. Competitive exclusion on the basis of niche differences is thought to drive local-scale assemblage structure (Huston 1999; Chesson 2000). Yet, there is debate whether individual functional traits, functional traits combined into a multidimensional trait space (Srivastava *et al.* 2012; Laughlin 2014; Kraft *et al.* 2015), or phylogeny (Mace *et al.* 2003; Mayfield & Levine 2010; Cadotte *et al.* 2013; Swenson 2013) best-reflect species' niche differences. I will address these key questions by fitting models that explicitly contrast the importance of (1) species' individual and combined functional trait distances, and (2) contrasting functional trait and phylogenetic distance in explaining ecological assemblage structure. Below, I describe the functional traits most likely to drive competition within each taxonomic group. In plants, Specific Leaf Area (SLA) and its correlates, such as leaf dry-matter content, are frequently evenly-spaced among plant communities (reflecting competition; Cornwell & Ackerly 2009; Bernard-Verdier *et al.* 2012), as is plant height (Šímová *et al.* 2015), and rooting depth (Kraft *et al.* 2015). In some systems, there is evidence that these traits form competitive hierarchies (with

optimal trait values; *e.g.*, Kraft *et al.* 2014); such patterns would be detected by the *additive* regional models described above. Similarity in body size is often used as a proxy for competition (Wilson 1975) in mammals (Bowers & Brown 1982; Brown & Nicoletto 1991; Jonathan Davies *et al.* 2007), birds (Leyequién *et al.* 2007), and beetles (Bakker *et al.* 2006; Fountain-Jones *et al.* 2015), and I would also expect that mammals, birds, and beetles with similar feeding preferences would compete more strongly (Cody 1974; Dayan & Simberloff 1998). Note that it is likely that intra-guild competition will take place within the context of environmental filtering on similar traits (as in birds; Cotgreave 1994).

Analysis. I will contrast the roles of functional trait and phylogenetic differences in driving local-scale assembly using the PGLMM framework (Ives & Helmus 2011) for each taxonomic group. To measure competitive exclusion, I will estimate a random effect for each species-pair, drawing each random effect from a species-pair covariance matrix defined by species' functional trait (*sensu* Laliberté & Legendre 2010) or phylogenetic (branch length) distances. This covariance matrix is weighted by an estimated parameter, σ, which varies between 0 and 1 and reflects the importance of traits or phylogeny in driving species' exclusion. The optimal value of this σ parameter and its confidence intervals are estimated as part of the PGLMM fitting process (Ives & Helmus 2011; Pearse *et al.* 2014). I will fit four models for each taxonomic group: (i) a *null model* consisting the best of the regional models (see '*Regional-scale environmental filtering: Analysis*' above), (ii) a *functional model* with the same terms and a random effect for each species-pair defined by functional trait differences, (iii) a *phylogenetic model* which is the same as the functional model but parameterized by phylogenetic distance, and (iv) a *combined model* which contains the random effects from the functional and phylogenetic models. Nested LRT of the models will reveal whether phylogeny [$LRT(\frac{phylogenetic}{null})$], functional traits [$LRT(\frac{functional}{null})$], or a combination of the two [$LRT(\frac{combined}{functional})$; $LRT(\frac{combined}{phylogenetic})$] are driving local-scale competition (question 2). In the case that functional and phylogenetic distance are both important, the relative σ values will quantify the influence of each. I will contrast the importance of individual and combined functional trait distances (question 1) by comparing models using individual functional traits and all functional traits combined (*sensu* Laliberté & Legendre 2010), using nested LRT and σ comparisons as described above. It is the inclusion of local-scale drivers of *co-occurrence* within the context of regional drivers of species' *occurrence* that makes this a truly multi-scale approach. For example, it would be possible to discern competition on the basis of body size in birds (local-scale co-occurrence; measured with random effects for species-pairs) within the context of regional filtering for large-bodied birds at colder sites (regional-scale; measured with fixed effect trait responses).

Macro-evolution of species' traits and responses:

Evolution of environmental filtering. In "*Regional-scale environmental filtering*" I describe how PGLMM can be used to model species' responses to their environment. This leaves two critical questions unanswered: (1) how the traits underlying these responses evolved, and (2) whether species' responses themselves (response functions; Díaz *et al.* 2013) are evolutionarily conserved. To test (1), I will fit models to estimate the tempo and mode of each trait's evolution, contrasting Brownian motion (a null model of conserved evolution; see Losos 2011), Ornstein-Uhlenbeck (OU—a model of adaptive evolution; see Butler & King 2004; O'Meara *et al.* 2006; Uyeda & Harmon 2014), and accelerating/decelerating models of trait evolution (Blomberg *et al.* 2003; Harmon *et al.* 2010) using AIC-based model selection (Burnham & Anderson 2004; Pennell *et al.* 2014). Assuming species are adapted to their environment or constrained by niche conservatism (Wiens & Donoghue 2004; Losos 2011), I expect β-traits to show evidence of conservation (Brownian) or adaptation (OU). This question was first posed in "*Regional-scale filtering*" above, and figure 2 contrasts some of these hypotheses. To test question 2, I will fit random effects that allow the slope of the trait-environment response to vary for each species (*i.e.*, allow the regional *interaction model* to vary across species), drawing each species'

random slope from a covariance matrix defined by species' phylogenetic distances using an estimated weighting parameter (σ; a simpler form of the random effects fit in the local-scale *phylogenetic model*). The σ parameter value (0—no phylogenetic effect; 1—strong phylogenetic effect) and LRT of this model with the simpler regional *interaction model* will reveal whether the environmental response is phylogenetically patterned.

Adaptation and repulsion. Above, I focused on using phylogenetic distance among species as a proxy for missing functional traits (see "*Local-scale competitive interactions*") and to test for conserved evolution (see "*Evolution of environmental filtering*" directly above). Modern comparative phylogenetic methods allow for the explicit comparison of different models of trait evolution by LRT of observed data under different linear transformations of a phylogenetic covariance matrix. Indeed, these transformations underlie the Brownian, OU, and accelerating/decelerating models of trait evolution I will use to measure the evolution of species' traits (question 1 in "*Evolution of environmental filtering*"). I will repeat all analyses involving phylogenetic covariance matrices with them transformed to represent Brownian, OU, and accelerating/decelerating trait evolution, testing whether the tempo and mode of trait-environment interactions are constant across plants, beetles, birds, and mammals using LRT to compare the fit of transformations. Further, I will examine the effect of using the inverse of the phylogenetic covariance matrix to see whether traits and species' responses have diverged through time (phylogenetic repulsion; see Nuismer & Harmon 2015). From theory, traits that drive local-scale competition would be expected to drive repulsive evolution to maximize trait distances among species (Nuismer & Harmon 2015).

Predicting species' compositions under change:

Environmental change. The models described above are all parameterized using regional-scale environmental data (see "*Data*") for which predictions under climate change scenarios are known (IPCC 2014). Because of the good coverage of GBIF data in North America for plants, beetles, mammals, and birds, the species that could plausibly disperse into each site are reasonably well-known. Using the existing phylogenetic and functional trait data for these species outlined in Objective 1, I will predict which of these nearby species are more likely to enter and leave the sites under future climate scenarios. These predictions will be made by fitting the estimated parameters from the models described above to a dataset including both these potential species and the species already at the sites. I will measure the predicted compositional, phylogenetic, and functional trait turnover (Pavoine & Bonsall 2011; Swenson 2011) of each site as a result of environmental change, to identify those sites most at risk of profound compositional shift. This will highlight prominent at-risk ecosystems across North America.

Invasive species. The USDA and IUCN release lists of the invasive species of greatest concern to them. Since each species in the lists is reasonably well-described in terms of their functional traits (often in the lists themselves), gathering functional trait (and phylogenetic, or at least taxonomic) data on these species is straightforward. Using these data and the estimated parameters of the models fitted above, I will rank the likelihood of each species successfully invading each site to help managers plan for invasive species at study sites. I will not attempt to estimate the resulting species compositions of the entire site (*c.f.* "*Environmental change*" above), since the wide-ranging impacts of invasive species can invalidate models parameterized before invasion (invasional meltdown; Simberloff & Von Holle 1999).

Feasibility, precision, and power. PGLMM are computationally intensive (Ives & Helmus 2011; Kaldhusdal *et al.* 2015), but new algorithms (Ho & Ané 2014), software packages (Pearse *et al.* 2015; Carpenter *et al.* 2017), and the access to the High Performance Computing facilities of Utah State University (see "*Facilities, Equipment, and Other Resources*") mitigate these issues. In terms of statistical power, PGLMM performs well and can distinguish between differing ecological assembly processes in smaller datasets than I propose to study here (*e.g.*, 31 sites and 32 species; Ives & Helmus 2011; Pearse

et al. 2014).

Peer-reviewed manuscripts (products) arising from Objective 2:
1. Comparison of the contribution of environmental filtering, competition, and evolutionary history to ecological assembly in North American beetle, bird, mammal, and plant assemblages.
2. Prediction of site species co-occurrences under environmental change, and of regions of North America most likely to be invaded by non-native species.

Objective 3: Model species' associations and interactions and thus predict beetle pest and pathogen outbreaks

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 (Bascompte & Jordano 2007; Poisot *et al.* 2015). By extending the cross-scale framework of Objective 2, I will overcome these limitations and describe the evolution of plant–beetle co-occurrences and mammal–tick interactions. This will broaden our understanding of the ecological and evolutionary dynamics of species interactions, and help predict pest outbreaks.

Data and justification of scope. I cannot examine interactions and co-occurrences among all of the groups of Objective 2 within the timescale of a two-year Early Career proposal. I focus, therefore, on two systems: plant–beetle co-occurrences and mammal–tick interactions. The plant–beetle association is the most tractable of the taxa in Objective 2: the trait and phylogenetic data are rich for both taxa (see Objective 1), their evolutionary and ecological associations well-documented (if complex; Price *et al.* 1980; Farrell 1998; Ode 2006; Hunt *et al.* 2007; Winkler & Mitter 2008), and the extension of PGLMM to them methodologically clear (see below). The only disadvantage to the plant–beetle data is they describe co-occurrences (not interactions), which is not the case for the mammal–tick data. These data are straightforward to model using PGLMM (see below), and the mammal insights developed in Objective 2, combined with the rich trait data for mammals and ticks (see objective 1), make the mammal–tick interaction tractable and appealing. I do not expect to have a tick phylogeny (see Objective 1), and so I only address questions that do not require one.

Plant–beetle co-occurrences:

Hypotheses. I will test three cross-taxa drivers of plant and beetle distributions: (1) separate regional drivers of plants and beetles, (2) plant traits, and (3) conserved evolutionary associations. Due to the complexities of beetle competition's effects on plant diversity (Bakker *et al.* 2006), I will not attempt to model any indirect effects of beetles' competitive traits on plant structure. Below I outline the rationale behind each of these hypotheses that I will test using NEON data. (1) The ground beetles (*Carabidae*) NEON collect are mostly predatory, and are both directly, and indirectly (through changing prey communities), affected by plant community composition and diversity (Siemann 1998; Koricheva *et al.* 2000; Brose 2003; Schaffers *et al.* 2008). However, as few studies have considered cross-scale drivers of species compositions, it is unclear whether these results are driven by plants' and beetles' responses to common regional environmental gradients. I will therefore test whether plant–beetle co-occurrences are driven by the regional factors measured in Objective 2. (2) Plants with 'fast' life-history traits [*e.g.*, greater Specific Leaf Area (SLA) and stem flow conductivity] tend

to grow faster (Reich *et al.* 2003; Reich 2014), and so may recover more quickly from herbivory. If so, it is plausible that plants with faster traits support greater herbivorous beetle populations, and so I will test whether more predatory beetles co-occur with 'faster' plants. (3) Species' responses to environmental drivers are often phylogenetically patterned (*e.g.*, Helmus *et al.* 2007), but the pattern of responses to other species is unclear (Cavender-Bares *et al.* 2009). In the case of plants and beetles, either group's response or effect could be phylogenetically patterned (top row and left column of figure 3) or not (bottom row and right column of figure 3). Describing the interaction of these processes (see figure 3) will help explain contrasting patterns of eco-phylogenetic structure found in other systems (Cavender-Bares *et al.* 2009) and predict ecosystem vulnerabilities (Díaz *et al.* 2013).

Analysis. To model species' co-occurrences, a small modification of both the input data and the PGLMM approach outlined in Objective 2 is required. I will represent each plant–beetle potential pairing at each site as a binary variable: 1 for "both present" and 0 otherwise. These data can then be modeled exactly as they would be in a PGLMM of interaction network data (Ives & Godfray 2006; Rafferty & Ives 2013). To test hypothesis 1, I will compare the first of two models using a Likelihood Ratio Test (LRT): a *null model* with separate random effect terms for each plant species, beetle species, and site, and a *regional model* with the same random effect terms and whatever regional terms are found to drive plant and beetle distributions in Objective 2. If the *regional* model significantly explains more variation, this will support hypothesis 1 and suggest regional processes influence plant–beetle co-occurrences. To test hypothesis 2, I will compare the best-fitting of the above models with two models containing additional fixed effects for plant SLA and stem flow. If LRT show either model explains more variance, it will support hypothesis 2. To examine question 3, I will compare the fit of four models using nested LRT: the best-fitting of the four models described above (the *null model*), and models with random effects drawn from a covariance matrix parameterized by beetle and plant phylogenetic distances (as in Objective 2's *Local-scale competitive interactions*), called the *beetle* and *plant phylogenetic models*, respectively. The fourth *combined phylogenetic model* will have random effects drawn from the Kronecker product of both the beetle and plant phylogenetic covariance matrices (following Ives & Godfray 2006; Rafferty & Ives 2013): intuitively, this resulting covariance matrix will 'expand' the interaction between the two matrices. $LRT(\frac{plant}{null})$ and $LRT(\frac{beetle}{null})$ will reveal if beetle or plant co-occurrences have phylogenetic structure—the top-right and bottom-left of figure 3, respectively. If $LRT(\frac{combined}{beetle})$, $LRT(\frac{combined}{plant})$, and $LRT(\frac{combined}{null})$ are all significant, this will reveal that both plant and beetle co-occurrences are phylogenetically structured (top-left of figure 3).

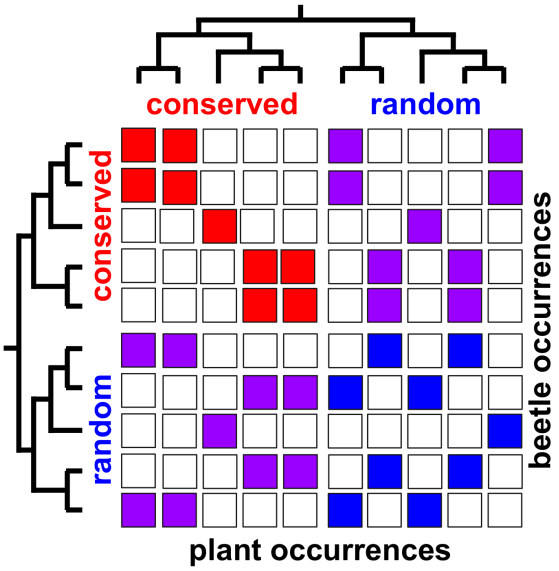

Figure 3: **Conceptual overview of hypotheses of plant–beetle phylogenetic co-occurrence.** Ten plant (columns) and ten beetle (rows) species are shown, with recorded co-occurrences between them shown with colored cells. For each group two clades, each with either phylogenetically conserved or random co-occurrences, are shown. Each clade combination represents a different hypothesis about plant–beetle co-occurrences (see text).

Mammal–tick interactions:

Hypothesis. It is increasingly understood that not all parasites are specific to a single host (Ewald 1983), and there is some evidence that host-specificity in mammal–tick associations is exaggerated by poor sampling (Klompen *et al.* 1996). Historically, work on mammalian parasites has focused on co-evolution (Barker & Murrell 2003), but there is growing concern that ignoring the macro-ecology (Stephens *et al.* 2016) and local-scale abundance of hosts (Farrell *et al.* 2015) leads to an incomplete picture of the drivers of host-parasite associations. I hypothesize that if ticks have diversified to match their mammal hosts, then closely-related mammals should share tick parasites. However, I further hypothesize that (2) tick associations will also be driven by ecological opportunity. Ecological opportunity can take two forms: mammals with similar traits (thus occupying similar niches) may have tick parasites in common, and variation in environment may limit/alter ticks' abilities hunt hosts (Schulze *et al.* 2001; Crooks & Randolph 2006; Brunner & Ostfeld 2008). I emphasize that I will not be examining tick diversification due to a lack of tick phylogenetic data: since my hypothesis is that mammal–tick interactions are conserved and specific to particular hosts, it can be verified using mammalian phylogenetic data alone.

Analysis. I will model mammal–tick associations across sites as a binary variable: 1 for a tick species observed on a mammal species and 0 otherwise (including when a mammal species is not observed at a site; following Rafferty & Ives 2013). I will then fit four models to test the two hypotheses outlined above: a *null model* with separate random effect terms for each mammal species, tick species, and site, and whatever regional terms are found to drive mammal distributions in Objective 2, a *phylogenetic model* with the same terms as the null model and an additional random effect for each tick species drawn from a covariance matrix parameterized by mammalian phylogenetic distances (analogous to the *plant phylogenetic model* in the plant–beetle section above), a *traits model* with the same terms as a the null model and an additional random slope measuring the effect of mammal traits on each tick species, and finally a *combined model* with all the above terms. I will then test hypothesis 1 [$LRT(\frac{combined}{phylogenetic})$ and $LRT(\frac{phylogenetic}{null})$] and hypothesis 2 [$LRT(\frac{combined}{traits})$ and $LRT(\frac{traits}{null})$] using nested LRT. If more than one mammalian trait is found to predict structure in Objective 2, I will fit separate *traits* models to test the importance of each trait in predicting mammal–tick interactions.

Predicting species co-occurrences and interactions:

Beetle outbreaks. There is a growing concern that the frequency of beetle outbreaks is increasing in the US (Raffa *et al.* 2008), and managing natural communities to maintain predator diversity is a well-established way to manage pest outbreaks (Landis *et al.* 2000). Objective 3 will build on the predictions of beetle distributions from Objective 2, and will additionally reveal how plant diversity is linked with beetle diversity. Plant diversity surveys are often quicker and easier to perform than beetle surveys, and these results will highlight what facets of plant community structure can be used to identify communities with fewer predatory beetles. Using the same methods as Objective 2, I will predict the regions most like to undergo the greatest beetle compositional changes, and so potentially at greater risk of pest outbreaks. I will generalize these predictions beyond NEON, using the outputs from Objectives 1 and 2 to identify sites outside of NEON that resemble at-risk NEON sites.

Pathogen spillover. Many tick diseases can infect humans (such as Lyme disease; Parola & Raoult 2001; Goodman *et al.* 2005). Despite this public health risk, tick sampling is labor-intensive and difficult to maintain over long periods. The models I will develop can be used to make predictions of the likelihood of tick presence, not just on the basis of mammal compositions that were observed at NEON, but also in species that were not observed on the basis of phylogeny. Using forecasts of future climate, I will predict what NEON sites (and so regions) are likely to see gains or losses of tick species, and so changes in the likelihood of infection by these disease vectors. This will highlight regions in