

**Principle Investigators:**

**Luke J. Harmon**, Associate Professor, University of Idaho: PI on the full proposal. PI Harmon will take the lead in creating computer software that implements novel comparative methods and testing their statistical properties, and will collaborate in developing mathematical models, obtaining and analyzing empirical data, and writing manuscripts.

**Scott L. Nuismer**, Associate Professor, University of Idaho: PI on the full proposal. PI Nuismer will take the lead in developing mathematical models, and will collaborate in methods testing and writing manuscripts.

**Senior Personnel:**

**Michael E. Alfaro**, Associate Professor, University of California, Los Angeles: Senior personnel on the full proposal. Alfaro will assist in obtaining and analyzing empirical data, and will collaborate in creating and testing software and writing manuscripts.

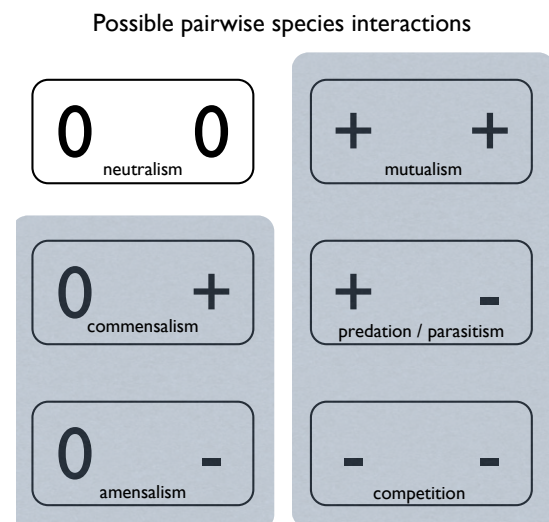
**Response to previous reviews:** The previous panel made some critical comments and recommended several key improvements that we have implemented. First, we now make a concentrated effort to more thoroughly capture the complexities of geographic range evolution - in particular, when species are sympatric or allopatric - when modeling species interactions. Second, we now better describe the method and its novelty in relation to previous work. A key component of this revision is an attempt to clarify the data that are required for the methods and the interpretation of model parameters. We believe that these modifications satisfy the main concerns reflected in the panel summary.

**Background:** Species interactions - including predation, competition, and mutualism - shape evolution (Thompson 2005; Futuyma and Agrawal 2009a). Most current research focuses on short-term effects of species interactions (e.g. Burdon and Thrall 1999; Nuismer and Gandon 2008). However, almost all major theories of macroevolutionary change postulate a central role for species interactions (e.g. escape-and-radiate [Erich and Raven 1964; Futuyma and Agrawal 2009b], the red queen [Van Valen 1973], and adaptive radiation [Simpson 1959; Schluter 2000; Yoder et al. 2010]). Rapid increases in our understanding of the tree of life (e.g. Delsuc et al. 2005; Thompson and Shaffer 2010) provide a potentially rich source of information about how species interactions shape evolution over long time scales. **However, there is a lack of analytic methods that can directly fit coevolutionary models of species interactions using phylogenetic comparative data.** In fact, all commonly used methods for fitting comparative models to trait data assume that evolutionary changes along each branch in a phylogenetic tree are independent of events on every other branch (e.g. Harmon et al. 2010) - an assumption that precludes interactions among species (Fig. 1; but see Ives and Godfray 2006). In this grant, we will combine mathematical analysis, statistical software development, and empirical analyses to fit coevolutionary models of species interactions to phylogenetic comparative data. These novel methods will allow comparative studies of species interactions (e.g. Page 2003) to leverage the power of model-fitting approaches, and consequently gain new insights into the dynamics of coevolution in deep time.

The proposed research will create a unified theoretical framework and analytic software for studying coevolution over macroevolutionary time using the phylogenetic relationships among species. This framework will expand spatial models of coevolution developed by PI Nuismer and colleagues (Nuismer et al. 2010; Yoder and Nuismer 2010) to the case where species share evolutionary history according to a phylogenetic tree. We will place these models in a framework for Bayesian comparative methods developed by PIs Harmon and Alfaro to quantify coevolution within diverse clades of interacting species. Our statistical framework is flexible and accounts for phylogenetic uncertainty in model parameter estimates (see Eastman et al. 2011; Slater et al. 2012; Slater et al. in press). We will then apply these new methods to a range of empirical datasets. This work is a follow-up from a previous NSF proposal to PIs Alfaro and Harmon. This grant, currently in its final year of funding, has thus far generated >20 publications in a range of high-impact journals.

Research questions: This work is guided by the following two questions:

**Question 1:** *How do species interaction networks evolve?* To address this question, we will measure the rate of formation and dissolution of novel interactions, investigate whether these rates vary through time and across clades, and determine whether related species tend to have similar sets of interactions.



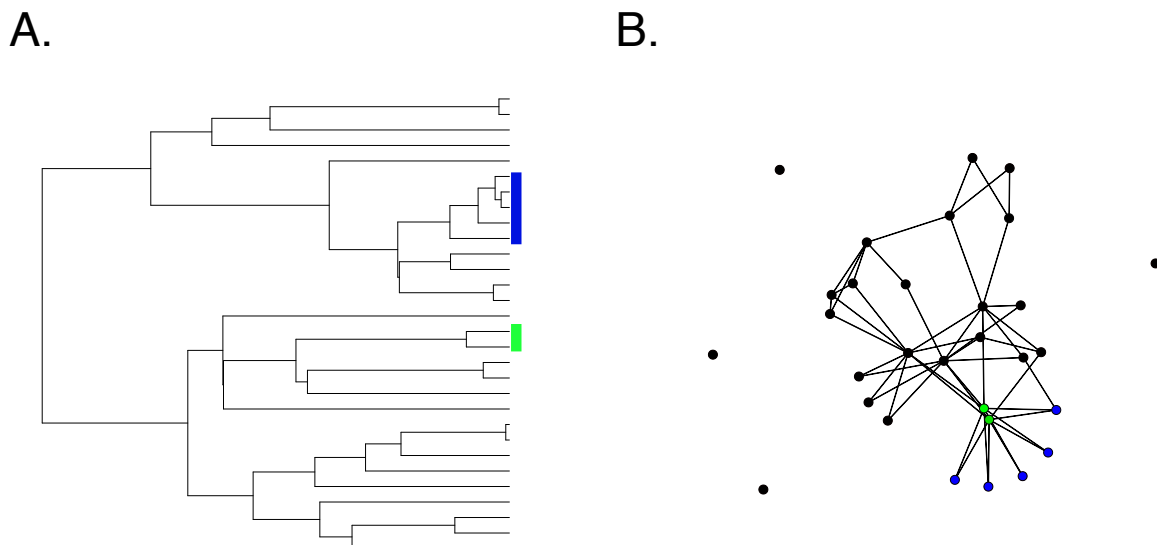
**Figure 1.** Possible coevolutionary interactions between pairs of species. Symbols represent the effect (positive, negative, or neutral) that individuals of one species have on the population growth rate of the other species. Models of trait evolution in the grey box cannot currently be fit to phylogenetic comparative data.

**Question 2:** *Can we detect the influence of species interactions on patterns of trait evolution?* To address this question, we will compare the fit of models that include species interactions to models that do not using phylogenetic comparative data.

**Research approach:** To answer **research question one**, we will develop software to analyze the evolution of species interaction networks. These networks summarize the set of pairwise interactions - either positive or negative - among a set of species (Fig. 2; Bascompte and Jordano 2007; Rezende et al. 2007, 2009; Vázquez et al. 2009a,b). The historical relationships among species on phylogenetic trees provide information about the evolution of species interaction networks. For example, one might expect shared interactions within clades of related species (Cattin et al. 2004; Ives and Godfrey 2006; Fig. 2).

We have developed an analytic model of the evolution of an interaction network on a phylogenetic tree with branch lengths. Our model can be described with two transition rate parameters:  $q_{01}$ , which is the “connection rate,” the rate at which new connections form between existing species; and  $q_{10}$ , the “disconnection rate” at which existing connections between species are lost. At speciation, both daughter lineages inherit the interaction traits of their ancestor. Our model also includes a parameter  $p_s$  that determines whether sister species have a connection immediately following the speciation event. Our model can be applied to either a single clade of evolving species that interact with each other, or to two clades of interacting species (for example, if species in clade A interact with species in clade B). Our approach also incorporates the geographic dynamics of coevolving species. This is crucial, because species with completely allopatric distributions cannot interact. To do this we rely on stochastic character mapping to reconstruct geographic ranges of species through time (Ree et al. 2005), then restrict interactions to occur only among species with overlapping ranges.

We have also implemented a dynamic programming algorithm that can calculate the exact likelihood for a particular interaction network given model parameters and a phylogenetic tree. **We can use this likelihood to fit models of species interactions to comparative data.** We require that end users will provide an ultrametric phylogenetic tree (or a distribution of trees reflecting phylogenetic uncertainty), an interaction network among extant species, and species’ geographic distributions. Our likelihood calculation allows a wide range of applications, from ML parameter estimation to Bayesian model fitting. We have already implemented a Bayesian MCMC algorithm to fit our model to empirical data. Using simulations, we have confirmed that our algorithm gives accurate estimates of parameter values for



**Figure 2.** Simulated phylogenetic tree and associated interaction network. We generated a phylogenetic tree with 30 species under a pure-birth model (panel A). We then evolved an interaction network on this tree (Panel B). We have highlighted two clades in this figure to illustrate that evolved networks have phylogenetic structure: note how the blue and green clades (panel A) interact with one another (Panel B). This phylogenetic structure (and our preliminary simulation results) imply that one can extract information about the evolution of interactions from phylogenetic comparative data.

datasets of reasonable size ( $> 40$  species). This software is written in R (R Core Development Team 2011) and will be a part of the upcoming release of GEIGER 2.0 (Harmon et al. 2008).

We will develop extensions of this method based on reversible-jump MCMC (Green 1995). For example, we can apply an MCMC model where either single parameter ( $q = q_{01} = q_{10}$ ) or two parameters ( $q_{01} \neq q_{10}$ ) can be used to describe the gain and loss of pairwise interactions across the tree. The reversible-jump chain can “jump” between these two different parameter spaces (Eastman et al. 2011). At equilibrium, the time that the chain spends in these two different spaces reflects the support in the data for a two-rate model. This allows us to test the hypothesis that either gains and losses of interactions happen more frequently. We will use a similar approach to test for different network dynamics in different parts of a phylogeny (following Eastman et al. 2011).

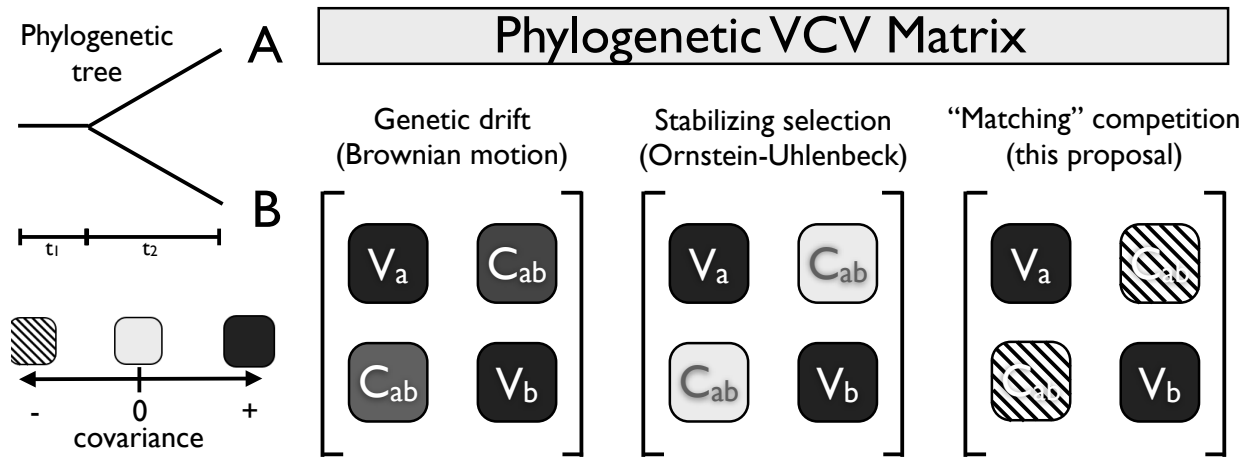
Analyses of network evolution will focus on four datasets: *Heliconius* butterflies (Beltran et al. 2007), labrid fishes (Alfaro et al in prep), marine food webs (Rezende 2009), and plant-animal mutualistic networks (Rezende 2007). In all of these datasets, we will estimate parameters of our interaction model. Initial analyses will target three fundamental questions: first, *how often do new network connections form and dissolve?*, second, *do newly formed species tend to interact with one another or not?*; and third, *do network dynamics vary through time and across clades?*

To answer **research question two**, we will analyze the evolution of interacting species’ traits. **Unlike all current phylogenetic comparative methods, our model will allow interactions among lineages to influence patterns of trait evolution.** We will base our method on coevolutionary quantitative genetic models that explicitly describe species’ interactions (Nuismer et al. 2010). Interactions can be positive or negative and are strongest when traits are most similar (Nuismer et al. 2005). Changing the sign of the interaction parameter allows us to study all common forms of ecological interaction, including parasitism, mutualism, and competition.

The behavior of coevolutionary models is well-known from previous research (e.g. Lande 1976; Nuismer et al. 2005, 2010). However, current analytical approximations of microevolutionary models apply only to the case when a number of populations diverge simultaneously from a common ancestor (a “star” phylogeny). For this work, we will extend these analytic results to the case where species evolve on a phylogenetic tree. We have developed a general method, based on solving systems of differential equations for the time intervals between speciation events, that can be used to derive the expected distribution of species’ trait values given a set of model parameters and a phylogenetic tree. Our method assumes that end users will supply an ultrametric, time-calibrated phylogenetic tree and data on traits and geographic ranges of extant species. Users can also incorporate knowledge of microevolutionary parameters (e.g. trait heritability, effective population size) by using informative priors in a Bayesian framework. **This model differs from all current comparative models (e.g. Brownian motion, Ornstein-Uhlenbeck) in that it can predict negative covariances in the trait values of interacting species**, for example under a competition model (Fig. 3). These negative covariances are dependent on the model parameters and the time since the species separated from a common ancestor. Our interaction model predicts that species’ traits will be drawn from a multivariate normal distribution, so we can calculate likelihoods and use standard ML or Bayesian methods for parameter estimation and model selection. Again, we use stochastic character mapping of geographic ranges to restrict interactions to occur only among species with overlapping geographic ranges.

**We will use a model-fitting approach to compare the fit of different types of interaction models to trait data.** For example, if there are no interactions in our model, species will evolve under Brownian motion or, if species are under stabilizing selection around an “optimal value”, they will evolve under an OU model. We can then compare the fit of models with and without species interactions. For example, we can test for (-,-) interactions using a model where competition is strongest when species have identical traits, and decreases as species’ traits become more different (Fig. 3). More complex models can include multiple types of interactions. We will then compare the fit of a set of potential models to data using Bayes factors and AIC. We will test the performance of this method using extensive simulations.

Analyses of trait interactions will encompass a broad diversity of datasets spanning a range of animal species for which trait and phylogenetic data are available and where species interactions have been hypothesized as drivers of trait evolution. We will analyze the full dataset from Harmon et al. (2010), which analyzed body size evolution in many groups. These data include phylogenies and trait data for a



**Figure 3.** Illustration of how species interactions affect patterns of trait evolution on a tree. For the simple phylogenetic tree (upper left), we show the expected trait variance-covariance matrix for species A and B. Covariances between species A and B are positive and large under Brownian motion, positive and small under Ornstein-Uhlenbeck, and negative under a matching competition model.

number of putative adaptive radiations like Darwin's finches and Hawaiian *Tetragratha* spiders (Harmon et al. 2010). We will also analyze patterns of trait evolution across megaphylogenies of bony fishes (Alfaro et al. in prep) and mammals (Meridith et al. 2011; Ventitti et al. 2011). Finally, we will analyze several other datasets where species interactions are thought to drive trait evolution (*Anolis* lizards [Mahler et al. 2010]; coral reef fishes [Alfaro et al. 2009, Dornburg et al. 2011], darters [Near et al. 2011], turtles [Jaffe et al. 2011]). We already have software pipelines set up for these analyses, and PI Harmon has extensive experience is assembling, dating, and analyzing large phylogenetic trees from previously published data sets (Alfaro et al, 2007, Alfaro et al, 2009, Santini et al, 2009). Analyses of all of these groups will ask whether models incorporating species interactions fit phylogenetic comparative data better than a suite of commonly used alternatives. Our analyses will address a key question for these groups: *do species interactions affect patterns of trait evolution over macroevolutionary time scales?*

**Broader impacts:** We will have three main areas of broader impacts: *graduate and postdoctoral training*, *an R short course*, and *K-6 outreach*. For **graduate training**, we will enforce a system where all graduate students are coadvised by two of the three co-PIs of the grant. We will also have a student exchange between UCLA and Idaho, where each student will spend one semester at the partner institution. We will also continue teaching our **R short course** (held annually since 2009 as a broader impact from a previous NSF grant to PI Harmon and senior personnel Alfaro). The course is extremely well received and popular and has already directly resulted in several collaborative relationships among our students. We feel that the course is approaching the point where it will be self-sustaining, but plan to requesting money to provide for six travel scholarships per year. These scholarships will target underrepresented students in the UC system and from Central and South America, a group that has expressed tremendous interest in our course (in 2011, we had many applications from South America and several such students enrolled).

For **K-6 outreach**, we will partner with two local charter schools in Idaho and Los Angeles and offer one-day "Academy Days" in evolution. The "Academy Day" is a one-day workshop program that is already established and running at Moscow Charter School (Idaho). Typical enrollment is 50 K-6th grade students. Our one-day workshop, called "Trees and Webs of Life," will teach students about the importance of phylogenies and species interactions using a variety of hands-on activities. For example, we will have students construct phylogenetic trees based on characters that they will observe themselves by looking at mammal skulls. They will also act out species interactions on the playground using a tag-like game that actually mimics the processes of interactions described in our coevolutionary models. Activities created as a part of this grant will also be posted on the web and freely available to anyone.

These broader impacts, which span training, teaching, and outreach, will substantially expand the reach and scope of our work, positively impact the field and society, and specifically target underrepresented groups.