CHAPTER 1

◇◇◇

An Entangled Bank

Evolutionary Relationships and Ecological Patterns

One faces the future with one's past.
—*Pearl S. Buck*

All organisms are the product of their own history. What they eat, where they live, the diseases they carry, and how they compete are all shaped by the environments and biological interactions experienced by their ancestors. For much of its scientific history the focus of ecological study was largely restricted to explaining the interactions between species and their environments; but it paid scant attention to patterns of shared evolutionary histories, and the integration of evolutionary theory into ecology began only recently. When defining the scope of ecology in the inaugural issue of the journal *Ecology* in 1920, Barrington Moore, the forester and fourth president of the Ecological Society of America, said "all life is controlled by two great forces, heredity and environment, and ecology is the science dealing with the environment." He was advocating specialization in the emergence of ecology as a scientific discipline. Perhaps the early divide between ecology and evolution was really about the differences in scientific cultures, whereby evolutionists viewed ecology as field based, primitive, and unscientific, and ecologists viewed evolutionary biologists as pursuing overly simplistic laboratory experiments and explanations (Kohler 2002). Or perhaps early ecologists sought specialization as a necessary way to develop and test key hypotheses, to craft new tools, and to create robust explanations for patterns of nature. As independent areas of research, ecological and evolutionary studies have provided profound insights into the formation and function of biological systems. For example, ecologists have developed predictions about species coexistence based on the ratios of limiting resources (Tilman 1982) and several theories of macroecology explain large-scale diversity patterns (Blackburn and Gaston 2006). Despite, or even because of this early history, researchers today are working in an era of synthesis, where Moore's two great forces are now part of unified explanations in ecology (Ricklefs 2007).

Evolutionary history has long informed our view of the world, and its consideration can be traced back in the scientific literature at least as far as Charles Darwin's *On the Origin of Species* (Darwin 1859). More recently, we have been witness to a rapid development of phylogenetic methods allowing phylogenetically informed comparisons of traits among groups of species (Felsenstein 1985, Harvey and Pagel 1991). Critically, concomitant advances in technologies have provided us with the raw material required for these new approaches—molecular sequences and phylogenetic trees. Interest in using phylogenetic information to evaluate mechanisms of community assembly and coexistence (Webb 2000) and niche conservatism (Holt 1996, Wiens and Graham 2005) increased dramatically as phylogenetic information for entire regions, clades, and communities has become readily available (fig. 1.1).
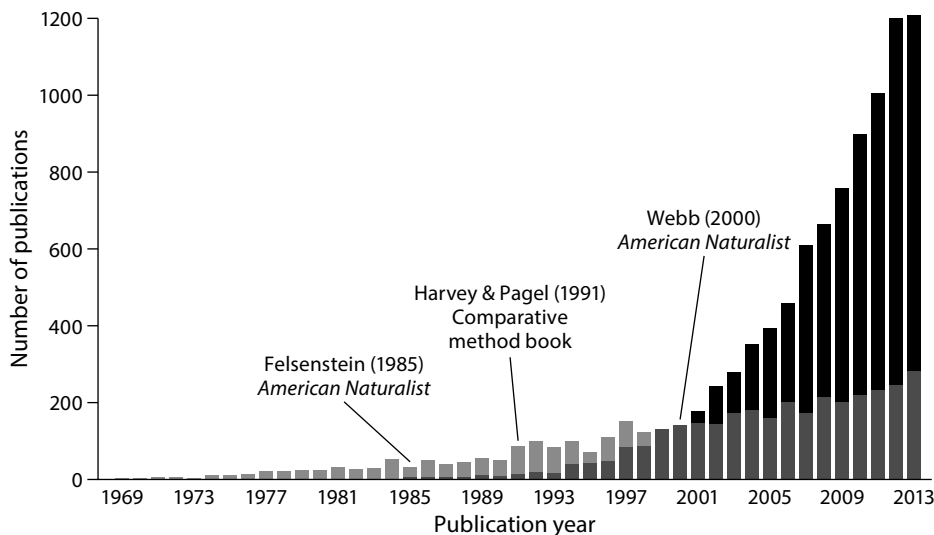
Figure 1.1. The number of publications indexed in Thompson Reuters ISI Web of Science that contain the keywords "phylogen'*'" and "community" (*dark bars*) from 1969–2013. The lighter bars transposed on top show the number of publications that contain the key word "competitive exclusion" for comparison. The dates of three classic publications that used phylogenetic information to test ecological hypotheses and thereby shaped the field are indicated.

Here we briefly review the history of the use of phylogenetics in ecology, starting with early attempts to classify the diversity of life and the development of evolutionary theory, through the rise of the comparative method, and finally to the emergence of ecological phylogenetics or *ecophylogenetics*.

1.1. SYSTEMATICS AND THE DIVERSITY OF LIFE

Even before evolutionary theory was widely accepted, early taxonomists, beginning with Carl Linnaeus, the father of modern taxonomy, have grouped species by similarity in their traits. Taxonomic ranks therefore represented a nested set of groupings—the Linnaean classification system (*Systema Naturae*)—from subspecies to the kingdoms of life. Linnaeus formally recognized only two kingdom's (Vegetabilia and Animalia) because the diversity of microscopic life was as yet largely unknown. A third kingdom was proposed to categorize the diversity of rocks and minerals, including fossils, but has fallen into disuse. We now recognize six separate kingdoms (Bacteria, Protozoa, Chromista, Plantae, Fungi, and Animalia), of which macroscopic organisms contribute just a small fraction—most of the life that surrounds us is invisible to us. A first step to categorizing the diversity of life is the placement of organisms within this taxonomic framework based upon their shared features.

The early classification presented in the 1735 *Systema Naturae* is now barely recognizable; some animal groups are still in use, including Aves (birds) and Mammalia (mammals), and although Linnaeus correctly placed bats (formerly grouped with birds) within Mammalia, whales were classified as fishes. In 1753, Linnaeus followed his *Systema Naturae* with

publication of the *Species Plantarum*, in which he formally classified all the then known plants of Europe based predominantly on their reproductive traits. Plant taxonomy has gone through numerous iterations and currently the most widely accepted, traditional taxonomic treatment (i.e., based upon shared features) for flowering plants is that of Cronquist (Cronquist 1968). The revolution in molecular methods and the subsequent rapid expansion in the volume of gene sequence data then inspired the first major revision to the taxonomic system since Linnaeus. The Angiosperm Phylogeny Group, a loosely coordinated network of research institutions led by Mark Chase at the Royal Botanic Gardens, Kew, Peter Stevens from the Harvard University Herbaria, and Kåre Bremer from Uppsala University, proposed a new classification of flowering plants explicitly based on monophyletic groupings reflecting evolutionary relationships (APG 1998). The publication of this work inspired the following headline: "A Rose Is Still a Rose, but Everything Else in Botany Is Turned on its Head" from one national newspaper (*The Independent*, 23 November 1998). Indeed, a revised taxonomy based on DNA data revealed several surprising relationships. For example, despite apparent morphological and ecological similarities with water lilies (Nymphaeaceae), the sacred lotus (*Nelumbo*) was shown to be more closely related to the plane tree (*Platanus*) and the southern-hemisphere *Protea* (Proteaceae); the traits that had led previous taxonomists to group them together reflect a remarkable case of convergent evolution, also referred to in the systematics literature as homoplasy. Nonetheless, perhaps the most astonishing, but least remarked on, feature of the new APG classification system was its broad agreement with previous classifications—many taxonomic groupings were found to represent monophyletic groups, defined by a shared common ancestor. The traits used by taxonomists were informative for describing important evolutionary relationships. Such traits are referred to as synapomorphies—derived traits that are similar due to inheritance from a shared common ancestor. In modern terminology we might now describe these traits as being phylogenetically conserved.

1.2. THE ORIGINS

As the father of evolutionary biology, Charles Darwin presented a framework to explain both modern patterns of biological form and function and gradients of biological diversity. Darwin's general theory of evolution through natural selection informed his understanding of the geographical distribution of species, as well as the outcomes of species interactions. For example, he stated that a "larger number of the very common and much diffused or dominant species will be found on the side of larger genera" (Darwin 1859, 54). That is, community dominance likely relies on inherited traits linked to species diversification, thus bridging the gap between evolutionary process and ecological pattern. Further, he noted, "if a plot of bare ground be sown with one species of grass, and a similar plot be sown with several distinct genera of grasses, a greater number of plants and a greater weight of dry herbage can thus be raised" (Darwin 1859, 113). Here the central premise is that the more closely related two individuals are, the more likely they are to be functionally and ecologically similar; this idea laid the foundations for decades of research linking diversity and productivity.

Darwin's theory predicted a simple but fundamental relationship between evolutionary divergence and ecological distance between species. He stated this elegantly when concluding his thesis on the influence of relatedness and competition:

> As species of the same genus have usually, though by no means invariably, some similarity
> in habits and constitution, and always in structure, the struggle will generally be more severe
> between species of the same genus when they come into competition with each other, than
> between species of distinct genera (Darwin 1859, 127).

Darwin thus recognized that a species' traits were important in determining its fit to the environment. Hence, closely related species sharing similar traits might be expected to be favored in similar environments; but also, competition would be greatest among species sharing similar ecologies, and patterns of evolutionary relationships serve as a representation of the ecological differences between species. The idea that ecological differences and evolutionary relatedness were critically important factors influencing species interactions became a central tenet in ecology by the mid-twentieth century. From his study of the birds in New Guinea, Ernst Mayr noted there was a general lack of co-occurrence between closely related species within habitats (Mayr 1942). Also, in a classic paper, Brown and Wilson (1956, 49) provided the slightly more formal observation on *closely related* species that have overlapping ranges. In the parts of the range where one species occurs alone, the populations of that species are similar to the other species and may even be very difficult to distinguish from it. In the area of overlap, where the two species occur together, the populations are more divergent.

Other early ecologists and biologists also saw the importance of relatedness in influencing ecological interactions and species distributions. For example, in discussing how competition structures animal communities, Charles Elton (1946, 64) commented:

> Some (though not necessarily all) genera of the same consumer level that are capable of living
> in a particular habitat at all can coexist permanently on an area; whereas it is unusual, in the
> communities analyzed, for species of the same genus to coexist there.

Biogeographers seemed to readily embrace the importance of species relatedness to explain patterns of biodiversity. In contrast, the emergence of community ecology during this time rarely considered species' evolutionary relationships directly (but see Simberloff 1970, Jarvinen 1982). That is not to say that early ecologists were not concerned about similarities; to the contrary, early conceptual and theoretical approaches placed a great deal of emphasis on how species similarities and differences influenced coexistence (Grinnell 1917, 1925; Gause 1934; Macarthur 1958). For example, Robert MacArthur, building on the concepts of Grinnell, Gause, and Volterra, stated, "if the species requirements [of co-occurring species] are sufficiently similar . . . only one will be able to persist, so that the existence of one species may even control the presence or absence of another" (MacArthur 1958, 599). From these observations, evolutionary theory would suggest that the phylogenetic pattern of species coexistence and co-occurrence should be one of overdispersion, with communities of coexisting species comprising distant relatives. Separate from this early period in ecology, and coinciding with the emergence of biogeography and community ecology, a theory was developing on why some communities might be comprised of species that are more similar than expected by chance. This contrast perhaps reflected the more general dissociation between ecology and evolution during that time (Kohler 2002).

1.2.1. Neutral Dynamics and the Lack of Species Differences

While the importance of species differentiation led ecologists to predict that less-closely related species should be more able to coexist, the importance of ecological differences itself has been repeatedly questioned. At the same time as the rise and solidification of

theories of coexistence based on niche differentiation, ecologists were observing community assemblages that were comprised of ecologically similar species (Ross 1957, Udvardy 1959) and started to explore the possibility that ecological interactions could be "neutral" (Sakai 1965). In 2001, Stephen Hubbell published *The Unified Neutral Theory of Biodiversity and Biogeography*, and introduced the neutral theory of biodiversity into mainstream ecology. Hubbell demonstrated that ecological communities could plausibly be the result of processes that rely little on interspecific differences, thus challenging the Hutchinsonian niche paradigm that had dominated community ecology over the preceding decades. The explicit assumption of neutral theory is that species have identical fitness responses to local environmental conditions. Under neutral processes one would therefore predict community assembly to be largely independent of phylogenetic history. However, neutral communities might still demonstrate strong phylogenetic structure. For example, phylogenetic structure can emerge through patterns of local dispersal. In addition, structure might reflect regional evolutionary dynamics; under a "point mutation" mode of speciation, where incipient species arise from a single individual, phylogenetic trees tend to be highly asymmetrical, dominated by one or a few species-rich clades (Mooers et al. 2007). In contrast, a fission mode of speciation, in which incipient species abundances may be large, can produce unusually symmetrical trees (Davies et al. 2011). Because tree symmetry also influences the evolutionary distances between taxa (Schweiger et al. 2008), depending upon the mode of speciation we might predict communities comprised of more or less related species, even when structured entirely by neutral processes (i.e., neither environmental filtering nor competition are important determinants of species coexistence).

1.2.2. Phylogenetic Patterns and Niche/Neutral Community Assembly

As focus shifted to consider multiple processes structuring species assemblages, the importance of spatial scale was also to become more apparent. The assembly of a local community is the result of the tension among local interactions, which are often negative, but sometimes neutral, and possibly deterministic (these interactions promote divergence and species ecological differences), as well as larger-scale environmental, dispersal constraints, or stochastic events (Berlow 1997, Belyea and Lancaster 1999, Levine 2000, Lovette and Hochachka 2006). At these larger scales, environment or dispersal limitation filters a larger pool of potential colonists into a smaller subset that includes species with the appropriate suite of traits for establishment (Keddy 1992, Weiher and Keddy 1995a), thereby promoting species similarities within communities. For this reason, larger regional species pools consisting of a broad array of functional traits should produce species assemblages with better matches between traits and local conditions compared to regions with small regional species pools (Questad and Foster 2008), perhaps shifting the balance between local versus large-scale processes.

The niche-based model of community assembly depends on phenotypic variability where niche differentiation maximizes differences among species, whereas neutral dynamics are the product of species similarities (MacArthur 1958, Schoener 1968, Fox and Brown 1993, Chesson 2000, Hubbell 2001, Chave 2004, Tilman 2004, Clark 2009). Recent work has shown that both evolutionary and ecological processes can give rise to assemblages built from processes simultaneously acting on species similarities and differences (Gravel et al. 2006, Holt 2006, Scheffer and van Nes 2006, Adler et al. 2007, Cadotte 2007, Fukami et al. 2007). Moreover, advances in coexistence theory suggest that species co-occurrence might

reflect a balance between differences in competitive abilities and niche differences (Chesson 2000, Adler et al. 2007), which may themselves have opposing effects on relatedness patterns (Mayfield and Levine 2010). In addition, niche and fitness differences may evolve at different rates, and singular predictions about phylogenetic relatedness and the strength of competition are therefore unlikely to be correct (Gerhold et al. 2015). For example, species with high niche overlap may also be better competitors relative to other species because they share a common trait that gives them a competitive advantage (e.g., tall plants); consequently, communities structured by competition can contain similar species, which some may see as evidence for environmental filtering (Mayfield and Levine 2010).

While viewing community assembly as a product of species similarities and differences may be conceptually appealing, quantifying these differences in traits or ecological niches in diverse communities has proven difficult. When we measure a limited set of traits, we run the risk of missing important traits or over-inflating the role of measured traits (Wright et al. 2006, Mokany et al. 2008, Schamp et al. 2008, Cadotte et al. 2013, Kraft et al. 2015). Given our limited ability to adequately identify and measure species differences across diverse assemblages, ecologists are increasingly employing the evolutionary history of species (represented by their phylogenies) as a surrogate measure of the overall phenotypic similarities and differences among members of a community, or at least as a starting point to identify the most important phenotypic differences for community assembly (Webb 2000; Webb et al. 2002; Cavender-Bares and Wilczek 2003; Cavender-Bares et al. 2006; Lovette and Hochachka 2006; Helmus, Savage et al. 2007; Cavender-Bares et al. 2009). Phylogenetic approaches provide two principal advantages. First, phylogenetic patterns of community assembly can lead investigators to key traits or niche differences important for coexistence. Second, and more importantly, phylogenies may better capture the multitude of ecological requirements, tolerances, interactions and proclivities that define a species niche; they therefore provide a powerful tool for interpreting community assemblage, which is likely better than any single-trait studies. That is, the phylogeny of a regional pool of species represents the product of the phenotypic and niche variation among species.

1.3. "CORRECTING" ECOLOGICAL COMPARISONS

Despite increasing awareness of the importance of phylogenetic history in informing patterns of species distributions and co-occurrences, the first rigorous use of modern phylogenetic information in ecology was in a method to account for species covariance in multispecies comparisons. A long tradition in ecological and evolutionary research has been to determine how traits correlate with one another among multiple species. For example, there may be interest in relating brain size to body mass, egg clutch size to fledging time, or plant invasiveness to seed size. However, Felsenstein identifies a potentially serious problem when analyzing such data, arising "from the fact that species are part of a hierarchically structured phylogeny, and thus cannot be regarded for statistical purposes as if drawn independently from the same distribution" (Felsenstein 1985, 1; see also Harvey and Pagel 1991). That is, invasive plants may tend to have smaller seed sizes because, say, a large family has a preponderance to be invasive while also having small seeds without a mechanistic link between the two. Thus, invasiveness and small seeds are correlated because of shared ancestry, and in fact, any autapomorphy for the clade might correlate just as strongly with invasiveness; hence correlation does not imply causality. Since Felsenstein first articulated this problem in

the statistical framework of phylogenetically independent contrasts, a number of solutions have been presented that account for phylogenetic covariance among species when measuring correlations among phenotypic traits (Felsenstein 1985, Harvey and Pagel 1991, Hansen 1997, Garland et al. 1999, Freckleton 2009, Nuismer and Harmon 2015).

1.3.1. Questioning the Relevance of Phylogenetic Corrections for Ecology

While the logic for the accounting of phylogenetic relationships in comparative analysis is persuasive, its relevance was called into question by some ecologists (Westoby et al. 1995, Ricklefs and Starck 1996). These researchers openly questioned the need for a "phylogenetic correction" of ecological analyses, for example when interpreting the importance of functional traits in different habitats. The concern was that phylogenetic corrections might erroneously weaken statistical relationships. In our above example, perhaps seed size does influence invasiveness, and applying the correction may obfuscate this relationship. The more robust approach would be to design experiments based upon the observed correlations. Philosophical considerations about the ability to infer the ecological value of traits aside, we cannot ignore the fact that failure to take phylogenetic relationships into account when examining trait correlations will lead to violations of the basic assumptions of most statistical tests (Ackerly and Donoghue 1995; Harvey et al. 1995b, 1995a; Ricklefs and Starck 1996). Rather than attempting to "correct" ecological analyses for phylogenetic nonindependence, modern comparative methods provide a statistical framework that allows phylogenetic information to be included within analyses of trait correlations (Ricklefs 1996, Ricklefs and Starck 1996, Martins and Hansen 1997, Abouheif 1999, Freckleton 2000).

1.4. THE EMERGENCE OF ECOPHYLOGENETICS

Perhaps the first published use of the term "ecophylogenetic" was by Armbruster (1992) in reference to the merging of phylogenetic information and ecological data so as to "infer the evolutionary history of ecological relationships." This early usage focused on the evolutionary origins of species' interactions. More recently the term ecophylogenetics has taken on a more specific definition, and is used most frequently to refer to phylogenetic community patterns and associated methods for assessing both potential assembly mechanisms and the key evolutionary events (e.g., trait evolution, species radiations, etc.) that influence the assembly of ecological communities (Cavender-Bares et al. 2009).

Campbell Webb's seminal paper in 2000 (Webb 2000) and the subsequent publication in 2002 of his influential review in the *Annual Review of Ecology and Systematics* (Webb et al. 2002) addressed whether communities contained species that were more or less closely related than expected by chance, and defined the modern field of ecophylogenetics. In 2000, Webb used a phylogeny of tree species found in a region of Borneo and generated null expectations for community phylogenetic patterns by creating random communities of co-occurring species. By comparing real community data to the null expectations, he showed that plots tended to contain species that were more closely related then expected by chance, inferring shared ecological preferences among close relatives. Previous work had approached similar questions using taxonomic information (Clarke and Warwick 1998, Warwick and Clarke 1998), or measured phylogenetic diversity as an alternative way to

understand biological diversity (Faith 1992a, 1994; Crozier et al. 2005). However, Webb's framework, linking trait evolution with community assembly processes to predict patterns of phylogenetic relatedness in communities, along with the increasing availability of phylogenies for entire communities, led to a rapid proliferation of studies using phylogenetic information as a lens through which to view the distribution of biological diversity and interpret the structure of ecological communities (fig. 1.1).

Within the succeeding few years, a number of other key events helped shape the field and accelerate its rapid expansion. In 2006 the highly influential journal *Ecology* published a special issue entitled "Integrating Phylogenies into Community Ecology" (Webb et al. 2006). This showcased 14 examples of ecophylogenetic studies in various systems around the world, from bacteria to tropical forests, and bird assemblages to frog parasites. Over the past few years a growing a number of researchers and working groups supported by diverse institutes, including the Long Term Ecological Research (LTER), National Center for Ecological Analysis and Synthesis (NCEAS), and the German Center for Integrative Biodiversity (iDiv), have worked toward integrating phylogeny and community ecology. Ecophylogenetics is now an established paradigm under which ecologists design studies and interpret ecological patterns. Yet despite this recognition, there are numerous limitations and critical assumptions that require research priority.

1.5. THE GOAL OF THIS BOOK

This book is meant to be a synopsis of the major concepts and methods available for ecophylogenetic analyses. We assume that the ultimate goal is to use evolutionary history to gain a deeper understanding of extant ecological patterns. Put another way, we assume that the user wishes to test ecological hypotheses (as opposed to strictly evolutionary hypotheses) ranging from community assembly and dynamics, to ecosystem function, to macroecological patterns. While highlighting the use of phylogenies in ecology, we also carefully describe important assumptions and limitations.

We point to how ecophylogenetic analyses can be applied to understanding and predicting the effects of anthropogenic environmental change. For example, distributions and community composition are responding to species invasions, climate change, eutrophication, and disturbance, and these responses are dictated in large part by the species' evolutionary history. As we move into a world full of novel habitats, species are armed only with the information contained within their genomes—information derived from evolution in past environments.

We see this book as a potential resource for students and working scientists interested in learning about and applying phylogenetic approaches to analyzing ecological data. We assume that the reader is broadly familiar with modern ecology and its core concepts (e.g., competition, predation, dispersal, etc.). Further, we hope that this book will aid researchers in developing and refining their questions, and make clear the data and analyses necessary to answer these questions.

Broadly speaking, this book has three main goals. The first is as an introduction to the utility of phylogenies for understanding ecological patterns. Specifically, we explore topics such as how different assembly mechanisms generate communities, how large-scale diversity biodiversity gradients are generated and maintained, and how to analyze the distribution of ecologically relevant traits. Our second goal is to provide an instructional manual

for the use of statistical methods for ecophylogenetic analysis. We will discuss the logic and assumptions of various tests, as well as constructing and interpreting different null modeling approaches. Our third goal is a practical one. We will detail the use of the tests in the R statistical programming language (www.R-project.org). All the examples used in this book will be fully implemented in R, and both the code and data are available on a companion website (http://press.princeton.edu/titles/10775.html). We assume that the user has a basic understanding of R, but we explain all the R code used in the examples so that even an R neophyte can implement the analyses.

1.5.1. Data Sets

Throughout this book we use several sample data sets to showcase methods and analyses. These data sets are available on the companion website. A brief note about the data sets:

1. Jasper Ridge: These data are from plant communities in 30 plots of 1×1 m^2 located at Stanford University's Jasper Ridge Biological Preserve in central California. For details of the data and the molecular phylogeny, see Cadotte et al. (2010b).
2. Phylogenetic supertree for mammals: This phylogenetic tree from Fritz et al. (2009) is a modification of the phylogenetic supertree originally published by Bininda-Emonds et al. (2007) with taxonomic updates. While a number of more recent mammal phylogenies have been published, this tree remains the only with almost complete taxonomic sampling, and is thus a valuable resource for comparative biologists. Three dated topologies are available; we use the "bestDates" version.
3. Global mammal distribution data: These data are derived from publicly available range map data made available through the IUCN (International Union for Conservation of Nature). Species ranges are depicted as polygons viewable in most common GIS software, and can be downloaded here: http://www.iucnredlist.org/technical-documents/spatial-data.
4. Mammal trait data: Tooth size data for Carnivora were collected by S. Meiri and were originally published in Davies et al. (2007). Body size data for the species community at Yotvata, Israel, were also collected by S. Meiri and were published in Davies, Cooper, et al. (2012).