Convergence, Natural History, and Big Questions in Biology

A symposium organized by Anurag Agrawal

Toward a Predictive Framework for Convergent Evolution: Integrating Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life
Anurag A. Agrawal, pp. S1–S12

Pattern and Process in the Comparative Study of Convergent Evolution

Convergently Evolved Toxic Secondary Metabolites in Plants Drive the Parallel Molecular Evolution of Insect Resistance
Georg Petschenka, Vera Wagschal, Michael von Tschirnhaus, Alexander Donath, Susanne Dobler, pp. S29–S43

Convergent Phenotypic Evolution despite Contrasting Demographic Histories in the Fauna of White Sands

Convergence and Divergence in a Long-Term Experiment with Bacteria
Richard E. Lenski, pp. S57–S68

Evolutionary Scenarios and Primate Natural History
Harry W. Greene, pp. S69–S86

Convergence, Consilience, and the Evolution of Temperate Deciduous Forests
Erika J. Edwards, David S. Chateau, Bo-Chang Chen, Jin Yao Ong, Shuichiro Tagane, Hironobu Kanemitsu, Kazuki Tagawa, Kentaro Teramoto, Brian Park, Kuo-Fang Chung, Jer-Ming Hu, Tetsukazu Yahara, Michael J. Donoghue
pp. S87–S104

Geographical Variation in Community Divergence: Insights from Tropical Forest Monodominance by Ectomycorrhizal Trees
Symposium

Toward a Predictive Framework for Convergent Evolution: Integrating Natural History, Genetic Mechanisms, and Consequences for the Diversity of Life*

Anurag A. Agrawal†

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853; and Department of Entomology, Cornell University, Ithaca, New York 14853

ABSTRACT: A charm of biology as a scientific discipline is the diversity of life. Although this diversity can make laws of biology challenging to discover, several repeated patterns and general principles govern evolutionary diversification. Convergent evolution, the independent evolution of similar phenotypes, has been at the heart of one approach to understanding generality in the evolutionary process. Yet understanding when and why organismal traits and strategies repeatedly evolve has been a central challenge. These issues were the focus of the American Society of Naturalists Vice Presidential Symposium in 2016 and are the subject of this collection of articles. Although naturalists have long made inferences about convergent evolution and its importance, there has been confusion in the interpretation of the pattern of convergence. Does convergence primarily indicate adaptation or constraint? How often should convergence be expected? Are there general principles that would allow us to predict where and when and by what mechanisms convergent evolution should occur? What role does natural history play in advancing our understanding of general evolutionary principles? In this introductory article, I address these questions, review several generalizations about convergent evolution that have emerged over the past 15 years, and present a framework for advancing the study and interpretation of convergence. Perhaps the most important emerging conclusion is that the genetic mechanisms of convergent evolution are phylogenetically conserved; that is, more closely related species tend to share the same genetic basis of traits, even when independently evolved. Finally, I highlight how the articles in this special issue further develop concepts, methodologies, and case studies at the frontier of our understanding of the causes and consequences of convergent evolution.

Keywords: adaptation, comparative biology, constraint, evolutionary ecology, phylogenetic ecology, plant-insect interactions.

Introduction

The search for convergent evolution and its causes is one way to make sense of the wonderfully bewildering biological diversity on our planet. The study of convergence is part of a program to discover repeated patterns and general principles that govern evolutionary diversification. Indeed, with the advent of non-model-omics, the study of convergent evolution is enjoying a new surge of interest, and this symposium and special issue of the *American Naturalist* brings together a superb group to address classic and novel questions on the topic.

Here I define convergence as the independent evolution of similar phenotypes. As has been noted many times, the intrigue of convergence was not lost on one of our greatest naturalists, Charles Darwin, when he identified traits such as luminescent organs in seemingly distantly related insects and pollen packages in distantly related plants. Darwin (1859, p. 193) wrote: "although the general appearance and function of the organ may be the same . . . some fundamental difference can generally be detected . . . Natural selection . . . has sometimes modified in very nearly the same manner two parts in two organic beings, which owe but little of their structure in common to inheritance from the same ancestor." Darwin recognized that convergent traits are not necessarily identical in all respects and that their evolution was largely independent, although not completely so, given that all organisms ultimately share a common ancestor.

Most convergent evolution falls under the umbrella of what may be considered “constrained adaptation”—evolution that is limited by the strength of natural selection, genetic architecture, and fitness costs and benefits, all of which narrow the number of possible evolutionary outcomes. Although convergence is often interpreted as evidence of both adaptation and constraint (sagaciously reviewed by Losos [2011]), the two processes are intertwined and thus can be difficult to separate. In this context, constraint has been broadly defined as “restrictions or limitations on the course or outcome of evolution” (Arnold 1992) or, more generally, the “unequal probability of outcomes in evolution” (Schwenk 1994/1995). Issues of constraint will be important in my discussion of convergent evolution, but because of the varied historical use and...
abuse of the term (Futuyma 2010), I will frequently use the term “bias” to simply refer to the pattern of unequal probability of outcome in evolution (table 1).

As Losos (2011) pointed out, repeated trait-environment associations are suggestive of convergent adaptation, and measures of natural selection and trait function can elucidate the adaptive nature of convergent traits (box 1). Nonetheless, such measures do not address the extent to which adaptation may be constrained. Many functional traits can be under selection and yet may be limited in how they can evolve. As will be discussed later in this article, a pattern of phylogenetic bias in both the degree of convergence and the underlying mechanistic basis of convergent traits is reshaping our view of constrained adaptation. The challenge of studying convergence in our era is thus interpreting patterns in natural history on phylogenies, with repeated or divergent genetic mechanisms helping to elucidate general issues in evolution. This was the focus of the American Society of Naturalists Vice Presidential Symposium in 2016.

In this introductory article, I aim to (1) introduce general issues and concepts in the study of convergence and in particular link the study of convergence to its roots in natural history, (2) provide a framework for addressing modern questions in convergence and summarize the state of the field in terms of progress and open questions, and (3) highlight themes addressed in the subsequent articles in this special issue. The current growth in studies of convergence has been fueled by increasingly sophisticated modern analytical tools of comparative biology, the ability to pinpoint genetic mechanisms of convergence using molecular biology, and the expansion of observational (as opposed to experimental) science in the past few decades. While perhaps less common, experimental approaches and novel statistical methods in the study of convergence are helping to explain patterns with process (Lenski 2017; Mahler et al. 2017). We are beginning to study the cellular and molecular mechanisms leading to convergence in distantly related organisms subject to the same selection, a critical approach to understanding phylogenetic biases in evolution (Petsch-enka et al. 2017; Rosenblum et al. 2017).

Convergence has also been at the center of thinking about major ecological patterns and their causes. For example, Edwards et al. (2017) seek to understand the causes of deciduousness in woody plants, a highly convergent trait that defines several global biomes. In some cases, it has been hypothesized that convergence occurs at the community level, driving greater similarity of species within a community than would be predicted from their phylogenetic relatedness (Schluter 1986). In other cases, dissimilarity evolves due to the convergent evolution of community members into different niches (Gillespie 2004; Losos 2009). Identifying traits that may be involved in community assembly are critical, as such traits are the link between evolutionary history and ecological outcomes. Together, these approaches represent an exciting frontier in the study of convergent evolution whereby we are able to study the causes and consequences of repeated evolutionary change integrating from genes to communities.

### Natural History: The Roots of the Study of Convergent Evolution

Natural history concerns the description of organisms (including their scientific name), their traits (any phenotype or description of their genome), location and distribution, and interactions with the biotic and abiotic environment. Although natural history serves as a foundation for biology, there continues to be controversy about the role that natural history should play in advancing knowledge in modern biology (Greene 2005). Yet clearly “reciprocal relationships among the growth of robust theory, experimentation, and accurate natural history” (Greene 2005, p. 23) are essential for advancement in our understanding of evolutionary biology. Natural history continues to be an important source of inspira-

<table>
<thead>
<tr>
<th>Table 1: Glossary of terms relevant to this article</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adaptive phenotype</strong></td>
</tr>
<tr>
<td><strong>Constraint</strong></td>
</tr>
<tr>
<td><strong>Biased convergence</strong></td>
</tr>
<tr>
<td><strong>Contingency</strong></td>
</tr>
<tr>
<td><strong>Convergent adaptation</strong></td>
</tr>
<tr>
<td><strong>Convergent evolution</strong></td>
</tr>
<tr>
<td><strong>Environment</strong></td>
</tr>
<tr>
<td><strong>Mechanism</strong></td>
</tr>
<tr>
<td><strong>Phylogenetically conserved convergence</strong></td>
</tr>
<tr>
<td><strong>Unbiased convergence</strong></td>
</tr>
</tbody>
</table>
tion, is critical for knowing and conserving what we have on the planet, and, most germane to this article, is central to the conceptual advancement of evolutionary biology in the broadest sense. Using his classic approach of natural history and phylogenetic thinking, Greene (2017) addresses convergence in animal behavior, with implications for understanding our human selves as animals.

Natural history in the context of comparative biology has historically been a critical hypothesis generator, which has often yielded general patterns and sometimes pushed the envelope of theory—especially in our understanding of convergent evolution. Two examples from my own corner of biology, plant-herbivore interactions, illustrate this point.

By the early 1960s, the birth of chemical ecology as a discipline yielded possible explanations for chemical diversity in plants and hypothesized consequences for major patterns in biological diversification (Fraenkel 1959; Ehrlich and Raven 1964). Key hypotheses about the function of plant chemistry were informed by careful natural history observations on distantly related groups of organisms (Fraenkel 1959). In a classic study on the "raison d'être of catnip," Thomas Eisner (1964) reasoned that the chemical components (including the terpene nepetalactone) causing catnip's effects, feline euphoria, were a defense of plants against herbivorous insects. His logic followed not from the incidental effects on cats but from the fact that nearly identical compounds were produced convergently by several insects, were insecticidal, and in some cases were ejected by the animals in response to risk of predation. Eisner was an observer who poked around and followed his nose. The ejection of nepetalactone by molested beetles suggested an explanation for the function of this compound in plants.

Since then, many other defensive compounds have been found to be produced by both animals and plants, including alkaloids and cyanides. The genetic basis of such convergent evolution in defenses continues to be revealed (Jensen et al. 2011; Denoeud et al. 2014), and the hypothesis generation has not stopped with identifying the functional or mechanistic basis of these defensive traits. Conceptual developments starting in the 1970s and advancing with phylogenetic thinking since 2000 have radically improved our understanding of which plant traits repeatedly evolve together to produce multivariate strategies of defense (Feeny 1976; Kursar and Coley 2003; Agrawal and Fishbein 2006; Fine et al. 2006; Mooney et al. 2010; Johnson et al. 2014; Mason et al. 2016). In each of these studies, either closely related species (many of which inhabit different habitats) or coexisting species (which span a broad swath of phylogenetic diversity) have been studied and arrayed along multiple axes of growth and defense. A clustering of phenotypes has suggested repeatedly evolving syndromes that can be tested for their ecological effects and associations with particular environments.

Another, perhaps more grandiose academic pursuit employing convergence and natural history involved the search for general patterns or rules that might govern nature. Alfred Russel Wallace (and other contemporaries, including J. W. Slater and E. Haase) posed hypotheses about bright coloration, toxicity, and the acquisition of noxious substances from the host plants of butterfly caterpillars. Each of these naturalists, through their travels, observations, and records, noted a set of associations that led to a hypothesis: brightly colored lepidopteran larvae obtain toxins from their host plants and use them in defense, typically against vertebrate predators. Subsequently, in the 1950s there was a race to demonstrate such sequestration of toxins by caterpillars, which, once successful, gave rise to the search for general rules of sequestration, aposematism, and mimicry (Reichstein et al. 1968). The monarch butterfly in particular was the first species shown to sequester toxins from its host plant and became a model to address such general rules (Agrawal 2017).

In a series of studies led by Miriam Rothschild between 1967 and 1973, the association between aposematism and sequestration was cemented by studies of several distantly related insect groups that feed on related host plants (von Euw et al. 1967, 1971; Rothschild et al. 1970, 1973), ultimately leading her to study 23 aposematic insect species from six insect orders that sequestered the same toxic compounds from their host plants in the Apocynaceae. Such findings opened the door to more rigorous statistical analyses, conceptual development, and theory on defense, sequestration, aposematism, and mimicry (Ruxton et al. 2004). Although the frequent association of sequestration, aposematism, and mimicry now seems obvious, it was the reciprocal interaction between natural history observations, studies of convergence, hypothesis testing, and theory development that led to the paradigm. Contemporary work addresses the genetic basis of convergent associations such as specialization and sequestration. In at least some cases, a diverse set of distantly related organisms utilizes the same mechanism to achieve adaptation to the same environmental challenge (Dobler et al. 2012; Bramer et al. 2015; Petschenka and Agrawal 2015).

Mechanisms of Convergence: A New Road to Assessing the Role of Constraint?

Traits are hierarchical in nature, from complex phenotypes (sometimes referred to as coarse-grained traits) down to the products of gene expression (i.e., fine-grained traits, beginning with messenger RNA). Higher-level traits in the hierarchy are typically dependent on several traits at lower levels (Conner and Hartl 2004), and the interpretation of convergent evolution has different meanings at these different hierarchical scales (Currie 2013). At the highest scale (or coarse grain), an ecological outcome of several traits may
be the trait (or extended phenotype) that exhibits convergent evolution. For example, flight has evolved at least four times, in insects, dinosaurs, birds, and mammals. Additionally, the highly similar ecological niches employed by placentals versus marsupial mammals on different continents also exemplifies convergent evolution of coarse-grained traits. If we zoom in, however, the physiological, morphological, and behavioral mechanisms that underlie convergent niches in these radiations may or may not be the same (Luo 2007).

Another such coarse-grained trait is the phenomenon of indirect defense, where protection of an organism from enemies is achieved through the attraction of, or patrolling by, animal bodyguards. The specific traits and conditions underlying indirect defense may be manifold, and these occur across diverse species and systems (e.g., ant-plants [Heil 2008], phloem-sucking hemipteran bugs [ant-tending; Styrsky and Eubanks 2007], leaf-chewing caterpillars [also ant-tending; Pierce et al. 2002], and even some fish and other vertebrates [Poulin and Grutter 1996]). The convergent evolution of such indirect defense as a high-level trait involves many distinct fine-grained traits, from housing structures and food rewards to volatile attractants, traps, and undiscovered mechanisms. Given that indirect defense is composed of multiple traits, we can begin to move down in scale to examine specific traits that generate indirect defense. In plants, a mechanism of indirect defense (one step lower in the hierarchy) may be extrafloral nectaries. We might next examine the chemical composition of extrafloral nectar secreted or, more mechanistically, the genes that code for the nectary and its chemical constituents. These finer-grained traits are mechanisms that generate coarse-grained traits.

Understanding underlying mechanisms is key to determining the causes of convergent evolution. For example, if convergent coarse-grained traits have distinct mechanisms, especially within a lineage, we may conclude that the evolution of those traits was less constrained. In other words, even within the backdrop of conserved genetic architecture, traits, and ecology common to a clade, the same
evolutionary outcome was achieved by distinct means. Conversely, if mechanisms of convergence are themselves phylogenetically conserved (the mechanisms are shared among close relatives that have independently evolved the trait), then convergence likely reflects constraint. I emphasize that I am using the term "constraint" here to simply indicate bias in the outcome of evolution, here driven by shared traits in a lineage (table 1; box 1).

**Phylogenetic Bias in the Mechanisms of Convergence**

Several examples illustrate phylogenetic bias in the mechanisms of convergent traits. C₄ photosynthesis in plants evolved over 60 times, and several different physiological means have been employed as carbon dioxide–concentrating mechanisms in C₄ plants (Sage et al. 2011). Recent evidence demonstrates that the oxygen afflux concentrating mechanisms (or cell lines) to address why such mutations are not realized. The biasing effect of genetic architecture is the tip of the iceberg (Natarajan et al. 2016).

Biases in convergent evolution can also be studied experimentally. Comparing what has evolved in the natural world with experimental populations has the potential to reveal the extent of constraint (Weinreich et al. 2006; Stern 2013; van Ditmarsch et al. 2013). For example, both mutagenesis screens for particular phenotypes and experimental evolution studies can reveal a greater number of possibilities than exist in nature, while their natural counterparts typically reveal biased outcomes. A next step will be to reveal why certain possible mechanisms are not realized. The biasing effect of genetic architecture is the tip of the iceberg (Natarajan et al. 2016).

In some cases, advantageous mutations to groups of genes involved in adaptation may be quite limited, even in highly controlled and benign laboratory conditions. In these cases, convergence may simply be the product of constrained possibilities. However, in mutagenesis studies of microbes, plants, and animals (mostly insects), screens have typically revealed many more potentially functional mutations than those that are realized through the natural evolutionary process (reviewed in Stern 2013). For example, in the case of molecular adaptations of animals to toxic cardiac glycosides (produced by plants and some animal prey), there are relatively few mutations that have repeatedly evolved (across many orders of insects and in some vertebrates) in the target site (e.g., the ubiquitous animal enzyme or the sodium-potassium ATPase; Dobler et al. 2012, 2015; Ujvari et al. 2015).

The mechanisms by which vertebrates adapt to high-elevation environments, typically involving the evolution of altered oxygenation properties of hemoglobin, have been widely studied (Storz and Moriyama 2008) and represent another case of phylogenetically conserved mechanisms in convergent evolution. In a remarkable recent study, Natarajan et al. (2016) studied 28 phylogenetic pairs of bird species and found that the oxygen affinity of hemoglobin in highland bird species was consistently higher than in closely related lowland species. Although the altered genes involved were consistent, the specific amino acid sites of substitution substantially varied. Thus, evolution was convergent at the level of individual genes but not within genes. Within hummingbirds, which had multiple origins of the highland habit, the genetic substitutions were nearly identical. Moreover, Natarajan et al. (2016) demonstrated that those genetic substitutions were functional only in the hummingbird’s genetic background, not in that of more basal species. Thus, phylogenetic bias in the mechanisms of convergent evolution may be driven by the genetic background of a clade.

Phylogenetic bias in the mechanisms of convergence appears to be a general result, one that should be quantitatively evaluated in the coming decades. Understanding this bias and its causes (such as a constraining effect of conserved genetic architecture) will lead to a greater understanding of convergence across scales. A potential exception to the pattern of phylogenetic bias in the mechanisms of convergence may be loss of functions, which can presumably occur through many mechanisms (i.e., mutations) that have the same ultimate consequence (Manceau et al. 2010; Smith et al. 2013). Thus, when the mechanisms of convergence are phylogenetically conserved (i.e., a given mechanism repeatedly evolves within a lineage but distinct mechanisms evolve between lineages), a biasing effect of genetic architecture, life history, or some other attribute of the lineage is suggested.
Contingency, Convergence, and Parallel Evolution: 
New Takes on Classic Questions

Contingency and convergence are sometimes thought to play mutually exclusive roles in evolutionary biology. In Gould’s influential book, Wonderful Life (1990), he made the case for the importance of contingent or chance events that have had a profound effect on the diversification of life. Yet despite being a hero in Gould’s story, Simon Conway-Morris subsequently took issue with several of Gould’s theses. In fact, in two subsequent books, Conway-Morris provided many examples of convergent evolution that he argued provided evidence against Gould’s conclusion of the importance of contingency (Conway-Morris 2003, 2015). Although it might be an attractive proposition that contingency and convergence are alternatives, is that truly the case? No. Convergence can be found at many levels, and yet contingent events also shape patterns of life on Earth. The polarization of the contingency and convergence debate has taken hold conceptually but is largely a false dichotomy.

Gould’s contingency was conceptualized at a deep scale of chance events that shape the future (e.g., extinction of major lineages such as the dinosaurs). If certain traits evolved only once and those lineages were subject to a contingent catastrophe, then clearly contingency rules. But the importance of such contingent events is not under debate. Rather, what is at issue is whether convergence overrides the importance of these chance events, because ultimately organisal traits would return to the convergent state. In all likelihood, the truth lies somewhere in between. While chance events can reset or change the course of evolutionary history, many traits have and will continue to exhibit convergent evolution.

Contingent events create evolutionary opportunities for convergent and divergent evolution. The geographic isolation of large groups of placental and marsupial mammals allowed for remarkable convergence as these groups diversified. Alternatively, certain other contingencies, such as mass extinctions, can reset or change the course of evolutionary history, many traits would return to the convergent state. In all likelihood, the truth lies somewhere in between. While chance events can reset or change the course of evolutionary history, many traits have and will continue to exhibit convergent evolution.

In experimentally tractable systems, if one has a hypothesis for what causes the predisposition, this can be addressed by introducing specific changes while controlling for other aspects of genetic architecture. In some cases, a particular gene may cause the predisposition (i.e., epistasis), and in other cases it may be a constellation of genetic factors (including gene duplication; Riehle et al. 2001; Stern 2013). Even the ecology or life history of a group may cause these events, and such predispositions may well be evident in phylogenies (Marazzi et al. 2012). In either case, a somewhat infrequent event may predispose a lineage to convergently evolve subsequent adaptations. As discussed above and in box 1, the same mechanism may repeatedly (convergently) evolve within lineages, but different mechanisms may be convergently employed in other lineages, all toward the same end of adapting to a particular selective agent. The examples of phylogenetically conserved convergences discussed above (C₄ photosynthesis, red flower color in plants, animal adaptation to high elevation, and animal resistance to cardiac glycosides; Dobler et al. 2012; Christin et al. 2015; Ujvari et al. 2015; Natarajan et al. 2016; Ng and Smith 2016) are outstanding candidates to study the joint role of biases and selection in adaptation.
What Is the Expected Level of Convergence?

Convergent evolution is inherently intriguing because at some level it seems unexpected, especially when convergent traits have the same genetic underpinnings. Nonetheless, where possible, the extent of convergence should be contextualized against a null model or hypothesis of what might be expected, both by chance and owing to selective processes. Stayton has argued for the importance of genetic drift in developing null models for the extent of convergence (Stayton 2008). Below I consider ecological opportunity, genetic architecture, loss versus gain of functions, and genes of small versus large effect when generating expectations for the level of convergent evolution.

In a qualitative sense, the likelihood of convergence may be driven by ecological opportunity, which may be based on what resources are in excess and can be easily employed. Viewed through this lens, it is not surprising that plants have evolved extrafloral nectaries many times (Weber and Keeler 2013). Sugar is often not limiting to plants, predacious ants often live nearby, and plant genomes share many genes. Similarly, as discussed above, many phloem-sucking hemipteran bugs have evolved indirect defense by ants who drink their sugary honeydew excretions (Styrsky and Eubanks 2007). Their honeydew is a waste product, but nonetheless, there are closely related Hemipterans that are ant tended or not, and the evolution of tending typically involves specific morphological traits as well as changes in the amount or composition of honeydew (Völk et al. 1999; Fischer et al. 2002; Shingleton and Stern 2003; Shingleton et al. 2005). Thus, the Hemiptera apparently have an ecological resource in excess, which, with some modification, may be highly sought after by an interaction partner.

In contrast, one group of leaf-chewing insects, the Lycaenid butterflies (and some close relatives) have evolved a nectar-producing gland and recruit ants as defensive bodyguards as well (Pierce et al. 2002). Within the Lepidoptera this may well be the only case of the evolution of defense by ants, and to my knowledge it has not been reported for beetles or flies. Some gall wasp lineages have repeatedly evolved the ability to induce nectar production on their galls (produced by their host plants), and they gain protection from patrolling ants (Inouye and Agrawal 2004; Nicholls et al. 2016). As far as I know, no vertebrate produces nectar to reward tending ants. Thus, among animals, the high ecological opportunity for the evolution of defense by ants appears to have widely spurred ant-hemipteran mutualism (and the traits that support it), whereas what caused the limited (single?) evolution of the same strategy in Lepidoptera is more difficult to know. Perhaps it was a chance event or circumstance in the history of the Lycaenids. The challenge for us now is to somehow quantitatively address the expectation for such convergences, perhaps depending on both extrinsic attributes such as ecological opportunity and intrinsic attributes such as genetic architecture.

Given the prevalence of epistasis, gene duplication, and effects of genetic background on the fitness advantages of particular mutations (e.g., Riehle et al. 2001; Weinreich et al. 2006; Tenaillon et al. 2012; Kryazhimskiy et al. 2014), the null expectation of convergence should be informed by knowledge of such genetic effects. Given a particular genetic background, a new mutation may be more compatible or functional and result in having higher fitness than without that specific background. As discussed above, we are at just the beginning of being able to understand biases in the extent of genetic convergence, but given that they exist, we should work toward a predictive framework that would include null expectations based on genetic architecture.

It has been suggested that loss of function mutations may on average be less mechanistically convergent than gain of functions (Arendt and Reznick 2008; Manceau et al. 2010; Smith et al. 2013). It is certainly the case that convergent loss of pigmentation, which often occurs through mutations in the same gene, occurs by distinct mutations at different sites (Protas et al. 2006). The rampant convergent loss of the gas (swim) bladder among teleost fishes also occurs by many distinct mechanisms; even losses within a single species, wild-caught zebra fishes, occurred by over 20 distinct mechanisms (McCune and Carlson 2004). In a sense, loss of function is analogous to a highly polygenic trait, since genetic modifications at many locations will result in the same phenotypic outcome.

Are convergent phenotypic traits that are controlled by many genes also less likely to have a common mechanistic basis? The genetic basis of body size is highly polygenic, and yet geographical clines (e.g., across latitude) in phenotype are highly repeatable across species and convergently evolve in native and introduced populations (e.g., see Lomolino 2005 for a review on vertebrates). Nonetheless, the genetic basis of such convergent phenotypic clines appears to occur by divergent mechanisms (reviewed for Drosophila in Gilchrist and Partridge 1999; Huey et al. 2000; Arendt and Reznick 2008). A polygenic basis for adaptation to high temperature may also underlie convergence in experimental studies of Escherichia coli. For example, experimental adaptation to high temperature in over 100 replicate lines revealed that any given pair of lineages shared relatively few nonsynonymous point mutations (2.6%), but at a higher hierarchical level, modified genes and operons were much more likely to be shared (>20%; Tenaillon et al. 2012). In the adaptation of fitness in 65 closely related yeast genotypes, Kryazhimskiy et al. (2014) demonstrated that although fitness evolved in highly repeatable trajectories, the specific mutations were highly variable and dependent on epistasis based on past fitness gains. In contrast to these studies on polygenic trait evo-
lution, in cases of resistance to particular toxins or diseases with highly specific target sites (with genes of major effect), much stronger convergence at the molecular level has been expected and observed (French-Constant et al. 1998; Ashfield et al. 2004; Dobler et al. 2012; Farhat et al. 2013; Yang et al. 2013; Brodie and Brodie 2015; Ujvari et al. 2015).

Beyond Mechanism: Evolutionary and Ecological Consequences of Convergence

Thus far, this article has focused on interpreting patterns of convergent evolution and integrating studies to understand the causes of convergence. As discussed above, the evolution of particular genes or traits may predispose a lineage to converge on a particular solution in response to natural selection. In addition, when convergence occurs (at whatever mechanistic level), are there predictable outcomes for the evolution and ecology of a lineage?

Perhaps the greatest interest among evolutionary biologists in consequences of convergence has come in the form of searching for macroevolutionary key innovations, those traits that evolve and allow organisms to interact with the environment in new ways and increase the net lineage diversification rate (Hunter 1998). Although there are now sophisticated analytical tools to detect shifts in diversification rates on phylogenies, even if they occur only once (Rabosky 2014), single occurrences have little ability to provide generality in terms of how traits impact diversification. Even though we have little understanding of why particular traits may facilitate speciation (or retard extinction; Heard and Hauser 1995; Futuyma and Agrawal 2009), recent work on plant-insect interactions suggests that such traits exist (Fine et al. 2004; Kaminski et al. 2010; Forister et al. 2011).

Indeed, our best cases of traits acting as key innovations, especially from traits that have repeatedly evolved, come from the defensive traits of plants: latex (Farrell et al. 1991) and extrafloral nectaries (Weber and Agrawal 2014). In both cases, the traits have convergently evolved in numerous plant families, the traits are associated with defense against herbivores, and lineages that have evolved these traits have higher diversification rates than closely related lineages lacking the traits. Convergent evolution of these defensive traits has had profound and predictable consequences for plant diversification. Remarkably, on the coevolutionary flip side, herbivory as a trait in insects is also an iconic case of a convergent trait (feeding strategy) that is widely associated with elevated diversification rates (Mitter et al. 1988; Wiens et al. 2015). Future work will certainly identify other convergent key innovations, but more importantly, we must address how and why the traits impact speciation or extinction.

From an ecological perspective, there is tremendous interest in understanding the species and traits that may cause predictable outcomes in community assembly and composition. Such processes may occur over long periods of time (which include evolutionary change) or over shorter periods of time (e.g., following disturbance or during succession). Do particular species and their traits, once present in a community, fundamentally change the course of assembly? In the classical cases of adaptive radiations in bounded communities (i.e., lakes and islands), there is often an association between convergently evolved animal ecomorphs and assembly of a fauna of related species (Losos et al. 1998; Gillespie 2004; Turner 2007). Nonetheless, cause and effect between the convergent traits and assembly of the community can be difficult to discern. The consideration of traits that modulate positive and negative species interactions is an especially important frontier, as both can impact community structure, though in different ways. Fukami et al. (2017) take on the issue of convergence in plant mycorrhizal associations and how these may have shaped predictable outcomes of community structure in diverse tropical forests.

In a general sense, convergent traits may strongly impact the process of community assembly through a few different processes. One metric of community assembly is the phylogenetic structure of a community, defined as a nonrandom pattern of evolutionary relatedness among species (Kraft et al. 2007). In the simplest case where habitat filtering is critical, convergence in traits among species will result in a pattern of even (sometimes called over-dispersed) communities, those where species are less closely related than would be expected by chance (Cavender-Bares et al. 2004). Conversely, if traits are phylogenetically conserved (e.g., tolerance of some stress), habitat filtering will result in phylogenetically clustered communities. However, when traits are evolutionarily conserved and yet there is selection for dissimilarity within a community (because of competition or other negative species interactions among close relatives), the outcome will result in an even community. Finally, for the case where some habitat filtering occurs as well as where species interactions are important and species’ traits are convergent (likely the most typical scenario), the community outcome is unclear (Kraft et al. 2007). Thus, quantitatively parsing out the impact of convergent evolution on community structure is a critical frontier in understanding the role of deep evolutionary processes in community assembly.

Looking Back and Looking Forward

There are two related issues for why convergence provides such an important and compelling approach to biology, especially when describing phenomena or taking a natural...
history approach. As pointed out by Felsenstein (1985), independent evolution provides statistical power in comparative biology. Correcting for phylogeny, though sometimes having a pejorative connotation, is simply about making robust statistical arguments about the association between two phenomena. Whether two traits show correlated evolution, a trait repeatedly evolves in a particular environment, or the repeated evolution of a trait is consistently associated with increased diversification, convergence in all cases provides power for stronger inference than if the association occurred only once. This is not to downplay the importance of contingent events that occurred only once; indeed, these may be some of the most critical events in evolutionary history. Nonetheless, to gain insight into whether there are general rules in biology, there is no substitute for evolutionary replication. Evolutionary replication allows us to come closer to understanding the cause of a particular association because, if repeated in independent lineages, the genetic background, life history, and ecology of the distinct lineages are likely to be different, and yet the association still stands.

The current renaissance in descriptive biology has grown from exploring the natural histories of the genomes of many organisms (e.g., Parker et al. 2013), engaging in new forms of discovery, and comparative analyses. There has been tremendous growth in the analysis of ecological gradients and understanding the natural pattern of species traits, distributions, and interactions along latitude, altitude, and other gradients. Part of this revolution has come from a renewed interest in trait-based ecology, the availability of phylogenetic information, and climatic databases. Given this resurgence, I conclude with four hopeful messages. First, let’s embrace this interest, as natural history is an important basis of inspiration, discovery, and the conservation of species. Second, the use of convergence can provide rigor and replication to address some of the biggest questions, ranging from understanding constraints to diversification. Third, there is tremendous potential to understand the underlying drivers of convergence by addressing the extent to which the mechanisms of convergent traits are biased by the phylogenetic lineage in which they evolve (box 1). Last, the combined use of comparative biology and experimentation (ranging from mutagenesis screens to reciprocal transplant experiments) will be critical in advancing biology (Weber and Agrawal 2012).

Acknowledgments

Funding and stimulation to think about this topic was provided by the John Templeton Foundation and especially P. Wason. Funding for this symposium and the speakers was provided by the American Society of Naturalists. For discussion, I am grateful to L. Arcila Hernandez, K. Böröczky, J. Elias, H. Greene, A. Hastings, K. Holmes, H. Inamine, P. Jones, M. Kirkpatrick, R. Lenski, A. McCune, K. Niklas, R. Petipas, B. Reed, and S. Scheiner. This article was improved by comments from J. Bronstein, J. Conner, J. Losos, J. Maron, A. McCune, J. Thompson, and M. Weber.

Literature Cited


Editor: Judith L. Bronstein

Pattern and Process in the Comparative Study of Convergent Evolution

D. Luke Mahler,1,† Marjorie G. Weber,2 Catherine E. Wagner,3 and Travis Ingram4

1. Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada; 2. Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824; 3. Biodiversity Institute and Department of Botany, University of Wyoming, Laramie, Wyoming 82071; 4. Department of Zoology, University of Otago, Dunedin, Otago 9016, New Zealand

**abstract:** Understanding processes that have shaped broad-scale biodiversity patterns is a fundamental goal in evolutionary biology. The development of phylogenetic comparative methods has yielded a tool kit for analyzing contemporary patterns by explicitly modeling processes of change in the past, providing neontologists tools for asking questions previously accessible only for select taxa via the fossil record or laboratory experimentation. The comparative approach, however, differs operationally from alternative approaches to studying convergence in that, for studies of only extant species, convergence must be inferred using evolutionary process models rather than being directly measured. As a result, investigation of evolutionary pattern and process cannot be decoupled in comparative studies of convergence, even though such a decoupling could in theory guard against adaptationist bias. Assumptions about evolutionary process underlying comparative tools can shape the inference of convergent pattern in sometimes profound ways and can color interpretation of such patterns. We discuss these issues and other limitations common to most phylogenetic comparative approaches and suggest ways that they can be avoided in practice. We conclude by promoting a multipronged approach to studying convergence that integrates comparative methods with complementary tests of evolutionary mechanisms and includes ecological and biogeographical perspectives. Carefully employed, the comparative method remains a powerful tool for enriching our understanding of convergence in macroevolution, especially for investigation of why convergence occurs in some settings but not others.

**Keywords:** convergence, phylogenetic comparative methods, adaptive radiation, evolutionary process, adaptation.

---

* This issue originated as the 2016 Vice Presidential Symposium presented at the annual meetings of the American Society of Naturalists.
† Corresponding author; e-mail: luke.mahler@utoronto.ca.

**ORCIDs:** Mahler, http://orcid.org/0000-0001-6483-3667; Weber, http://orcid.org/0000-0001-8629-6284; Wagner, http://orcid.org/0000-0001-8585-6120; Ingram, http://orcid.org/0000-0003-0709-5260.

DOI: 10.1086/692648

**Introduction**

The phenomenon of phenotypic convergence plays a fundamental role in the study of organic evolution. Although convergence itself is not necessarily indicative of any particular evolutionary process (Losos 2011a; Speed and Aruckle 2016), the repeated appearance of similar forms in disparate lineages stands in apparent contrast to the expected pattern of divergence over time during evolution and thus demands an explanation (Wake et al. 2011). Convergent evolution has been attributed to a great diversity of causes, at times being invoked as evidence for the importance of multiple and sometimes opposing evolutionary processes. As such, confusion persists around the connection between patterns of convergence, mechanisms of evolution, and modeled processes in comparative methods (see box 1 for definitions of these terms).

Our aim is to clarify these concepts and show how they relate to common assumptions in the comparative study of convergence, recommending best practices and advocating integrative ways to link convergent patterns with mechanistic hypotheses. We begin by reviewing the comparative study of phenotypic convergence in continuously valued traits and the factors that make it uniquely challenging to study deductively. We then discuss the relationship between pattern and process in the comparative study of convergence, giving special attention to the facts that (1) all comparative tools for studying convergence make inductive inferences about convergence and thus assume an underlying model of evolution and (2) despite the assumption of a modeled process, there can often be a many-to-one mapping of real evolutionary processes to modeled processes. Finally, we argue that because these limitations can hinder interpretation, integration of comparative phylogenetic models of convergence with other forms of inference is needed to increase confidence in links between pattern and process. We suggest several research directions to improve future prospects for gaining meaningful insights about evolutionary pro-
cesses from the comparative study of convergent patterns, including extended model development, pairing comparative study with other approaches for studying the evolutionary process, and better incorporating fossil information.

### The Potential Causes of Macroevolutionary Convergence

The investigation of convergence plays an important role in evolutionary study. To many, instances of convergence are interesting as evidence for adaptation or for the importance of ecology in the evolution of phenotypic diversity. For example, the evolution of pale dorsal coloration in numerous vertebrates and invertebrates inhabiting White Sands National Monument indisputably reflects adaptation for increased crypsis (Rosenblum et al. 2010), and the repeated evolution of cold tolerance in distantly related conifers is clearly attributable to adaptation to similar environmental conditions (Yeaman et al. 2016). Large-scale convergence between whole faunas occurring in similar ecological communities...
and environments suggests that evolution can in some circumstances exhibit a surprising degree of predictability and that ecological factors can repeatedly and predictably shape macroevolutionary diversification (Nevo 1979; Conway Morris 2010; Mahler et al. 2013; Esquerré and Keogh 2016; Moen et al. 2016).

Alternatively, convergence has been taken as evidence for constraints on the production of variation (Haldane 1932; Maynard Smith et al. 1985; Schluter 1996), manifested at one or more hierarchical levels of biological organization (Wake and Larson 1987; Wake et al. 2011). Such constraints can arise from biased mutation, pleiotropic gene networks, structural limitations, or limits on phenotypic variation imposed by ontogenetically nested developmental sequences (Gould 1980; Alberch 1982; Oster and Alberch 1982; Arnold 1992; McCune and Carlson 2004; Brakefield 2011; Streifeld and Rausher 2011; Stern 2013). For example, Bright et al. (2016) concluded that the vast majority of craniofacial variation in predatory birds was attributable to conserved patterns of allometry and covariation between traits (phenotypic integration), with only a small amount of residual variation explained by feeding ecology. Convergence resulting from such factors suggests that genetic or developmental constraints play an important and perhaps dominant role in shaping the evolution of phenotypic diversity (Gould 1980, 2002; Alberch 1982; Wake and Larson 1987).

Finally, some degree of evolutionary convergence may be an expected outcome of chance, especially for traits with low dimensionality (Wagner 2000; Stayton 2008). These possibilities are not mutually exclusive, of course, and convergence due to chance may be most likely in scenarios in which constraints on the generation of variation restrict evolutionary outcomes to a small set of phenotypes with comparable fitness (Losos 2011a; Spor et al. 2014). Likewise, some authors have uncovered phylogenetic patterns suggesting that certain convergent adaptations are hierarchically constrained by the presence or absence of preadaptations (Mara et al. 2012; Beaulieu et al. 2013). For example, the convergent evolution of arboreal adaptations in oribatid mites appears contingent on the prior evolution of sexual reproduction and strong sclerotization (Maraun et al. 2009).

### Challenges to Measuring and Studying Convergent Evolution

Despite its importance in evolutionary inquiry, convergence is difficult to directly identify, and once identified, it is difficult to ascribe to underlying mechanisms with certainty (Maynard Smith et al. 1985). A principal challenge in studying convergence is that it is rarely possible to study using deductive inference due to the difficulty of observing both ancestor and descendant phenotypes. Empirically, this is generally possible only in exceptional circumstances, such as when the fossil record preserves unambiguous ancestor-descendant sequences (e.g., Bell 1987; McCune 1987), when source populations that gave rise to convergent daughter populations are still extant (e.g., Hoekstra et al. 2006; Rosenberg et al. 2010), and when convergence is particularly rapid (e.g., Pascoal et al. 2014), including in laboratory experiments on organisms with very short generation times (e.g., Meyer et al. 2012; Spor et al. 2014). Studies documenting convergence in this way represent some of the best and most cherished evidence for convergent evolution, but they are uncommon and are limited in what they can tell us about the evolution of convergent patterns across the tree of life.

Convergence has a long history of study despite the limited opportunity for deductive inference, but aside from exceptional cases such as those described above, much historical study of convergence has been qualitative in nature. Most of the canonical examples of convergence described in introductory biology textbooks, such as the streamlined profiles of fast-moving pelagic vertebrates or the winter pelage of Arctic foxes and snowshoe hares, are so visually striking and occur in such distant relatives that there can be little question about their convergent origins. What was historically lacking was a cohesive quantitative framework for inferring the trajectories of convergence; the lack of such a structure imposed formidable limits on the study of the ultimate drivers of convergent evolution. This framework emerged with the union of the comparative method with an explicitly phylogenetic perspective in the 1980s and 1990s (Felsenstein 1985; Harvey and Pagel 1991).

### Phylogenetic Approaches to Studying Convergence

The advent of phylogenetic comparative approaches to studying trait evolution expanded the scope of convergence studies, making it possible to test quantitative hypotheses about convergent evolution in any group with sufficient phylogenetic and phenotypic information. This development effectively opened the quantitative study of convergence, previously limited to exceptional cases, to the entire tree of life. When viewed within a phylogenetic comparative framework, repeated convergence of any kind provides researchers with a degree of statistical replication rarely afforded to students of the explicitly historical science of evolution. Regardless of the question of interest, the repeated evolution of similar phenotypes in disparate lineages provides independent replicates in a grand, unplanned evolutionary experiment. The ability to repeatedly query putative cause and evolutionary effect allows investigators to overcome the risks of “just so” storytelling (Gould and Lewontin 1979) in the study of adaptation, structural constraint, and even chance (Maynard Smith et al. 1985; Harvey and Pagel 1991; Losos 2011a).

The development of tools for investigating convergence accelerated especially rapidly during the last decade (Speed
and Arbuckle 2016). New inductive tools provide a wide variety of methods for quantifying convergence—including its occurrence, frequency, extent, and historical trajectory. Such tools come in a variety of forms, the details of which are described in depth elsewhere (see Mahler and Ingram 2014; Stayton 2015a; Arbuckle and Speed 2016; Speed and Arbuckle 2016 for more in-depth descriptions). Most, however, can be classed into one of three categories: (1) statistical indices, which measure an expected emergent feature of convergent phenotypic evolution on a phylogeny; (2) ancestor reconstruction methods, in which ancestral phenotypes are estimated under some assumed evolutionary model and then used to identify and quantify convergence patterns; and (3) model-fitting approaches, in which evolutionary models explicitly incorporating processes expected to cause convergence are parameterized and compared to models in which any convergence occurs by chance.

All existing comparative methods for studying convergence have particular limitations and weaknesses, as we will discuss below. More importantly, however, is that all of these tools make inductive inferences about convergence, an approach that has practical consequences for the design and interpretation of comparative studies. Regardless of the method used, the results must be interpreted in light of an assumed model of the evolutionary process; as a result, it is not possible to decouple the quantification of convergent pattern from the study of evolutionary process using comparative data, as has been recently recommended (Stayton 2015a). This issue is compounded by the fact that many widely used evolutionary process models may plausibly represent multiple underlying evolutionary mechanisms—a potential many-to-one mapping of mechanism to model. At the heart of these issues is the complex relationship between pattern and process in comparative biology.

**Pattern and Process**

An important property of any evolutionary phenomenon is the extent to which it represents a pattern versus a process (box 1). For example, the term “adaptation” is meaningfully defined as both a process (e.g., adaptation occurs when a population evolves greater fitness via natural selection) and a pattern (e.g., an adaptation is a trait that increases an individual’s fitness compared to individuals without that trait; Gould and Vrba 1982; Futuyma 2005). In the case of evolutionary convergence, we feel that in almost all scenarios convergence is best defined as a pattern (Stayton 2015a). Convergence is less meaningful as a process, because convergent evolution is nearly always an emergent outcome of evolutionary processes operating independently in multiple lineages rather than any intrinsically convergent processes. For example, short-limbed twig specialist anoles on different Antillean islands are similar because they have adapted to similar arboreal substrates (Williams 1983; Losos 2009), not because they have been selected to resemble one another. There are a few exceptions where species are directly selected to converge with one another, such as mimicry complexes (Endler 1981) or character convergence driven by competition for nonsubstitutable resources (MacArthur and Levins 1967; Abrams 1987; Fox and Vasseur 2008), but otherwise the evolutionary processes responsible for increasing similarity in a converging lineage are blind to the phenotype of the lineage to which it is converging. As standard evolutionary theory can explain the processes by which independent lineages evolve to be more similar, there is no need for a special theory of convergence (Speed and Arbuckle 2016), and convergence is thus best defined as a pattern.

Although we agree with several recent reviews that convergence should be defined as a pattern, we argue that when using the comparative approach, convergence must be studied with the evolutionary process in mind. We therefore reject arguments that comparative analyses of convergence should proceed in a two-step manner—first testing for convergent pattern, and then investigating potential evolutionary processes responsible for this pattern (Stayton 2015a, 2015b; Speed and Arbuckle 2016). In a recent review, Stayton (2015a) made a case for this two-step approach, specifically criticizing the use of process-based comparative tools for identifying convergence. Stayton’s concern is defensible—in applying an evolutionary model to comparative data (e.g., multiple optimum Ornstein-Uhlenbeck [OU]; Butler and King 2004), the investigator assumes that the process being modeled is an appropriate representation of trait evolution in lineages of interest. In the absence of appropriate model comparison, this could bias an investigation toward a particular evolutionary explanation for convergence. The risk of bias is arguably greatest for adaptive explanations (sensu Gould and Lewontin 1979), and Stayton marshaled evidence for adaptationist bias in convergence definitions provided in many prominent biology texts (Stayton 2015a). To safeguard against adaptationist biases, Stayton recommended that comparative studies first employ process-neutral statistical tools to identify and measure macroevolutionary convergence and then, patterns in hand, test alternative hypotheses about the evolutionary processes that may have given rise to these patterns.

While the goal to investigate convergence without making assumptions about process is a worthy one, in phylogenetic comparative biology it is impossible to study evolutionary pattern and process independently (Pagel and Harvey 1989; Harvey and Pagel 1991; Freckleton et al. 2011; Hunt 2012). The purpose of phylogenetic comparative methods is to study patterns that inherently arise as a consequence of evolutionary processes, both to understand how history has shaped these patterns and to infer
the processes associated with this history. As referenced above, phylogenetic comparative biology is largely an inductive science, due to the usual lack of direct observational evidence of historical processes and patterns. The comparative method provides a framework for testing hypotheses about these phenomena, but an essential component of this framework is the assumption of evolutionary process models.

**Evolutionary Process Models Underlying Tests of Convergent Pattern**

In the case of convergence, the assumption of an evolutionary process model is required at some step (often implicitly) by all available comparative tools. The most widely assumed model is Brownian motion, which is used to model trait evolution in a variety of contexts. Brownian motion is a very simple model that represents the expectations of continuous phenotypic evolution under neutral genetic drift (Lande 1976; Felsenstein 1988). Like most models used in phylogenetic comparative methods, Brownian motion is a phenomenological model of the evolution of mean species-level characters, reflecting macroevolutionary expectations but not incorporating microevolutionary mechanisms. It is often used to represent a hypothesis of neutral evolution (especially as a null model), but some have argued for its utility in representing other evolutionary mechanisms such as fluctuating directional selection or adaptive radiation on a dynamic adaptive landscape (although with important caveats; Felsenstein 1988; O’Meara et al. 2006). In studies of adaptation, a popular generalization of Brownian motion is the Ornstein-Uhlenbeck model (Hansen 1997; Butler and King 2004; O’Meara and Beaulieu 2014). The OU model includes a Brownian drift term as well as a parameter describing the strength of attraction to some optimum value. Extensions allow different lineages to be attracted to different optima, which may be interpreted as peaks on an adaptive landscape. Unlike Brownian motion, specific OU models can model processes consistent with deterministic evolutionary convergence, though it is important to note that both are evolutionary process models (box 1). Recently developed Lévy process models represent an alternative generalization of Brownian motion in which a Brownian drift process is punctuated by large, instantaneous shifts in trait value (i.e., evolutionary jumps; Eastman et al. 2013; Landis et al. 2013). Lévy process models lack parameters specifically expected to produce convergence but can be used to test hypotheses about the frequency of convergent jumps in groups for which the phenotypic similarity of certain species has already been established (Eastman et al. 2013).

Although all comparative methods for studying convergence employ evolutionary models in one fashion or another, the role and potential impact of the model vary across approaches. Index methods implicitly use process models to generate a frame of reference with which to compare putatively convergent evolutionary patterns. For example, the Wheatsheaf index (Arbuckle et al. 2014) uses pairwise phylogenetic distances as a yardstick for evaluating and then weighting observed pairwise trait differences, to be contrasted with the correspondence between pairwise phylogenetic and trait differences expected under Brownian motion. Index tools have been described as “process-neutral” or “process-free” (Stayton 2015a; Speed and Arbuckle 2016), which is accurate in the sense that convergent patterns are not assumed to have evolved under any given evolutionary mechanism. However, these measures are useful only in reference to expectations under a particular evolutionary process model, which is sometimes unspecified but most often Brownian motion. Furthermore, because no underlying historical process model is applied to the data themselves, these tools can be limited by an inability to distinguish convergence from other causes of evolutionary similarity between distant relatives, such as a simple lack of divergence (Stayton 2015a; Speed and Arbuckle 2016). We suggest that additional insights may be gained by comparing statistical indices to distributions simulated under alternative models of the underlying process (sensu Slater and Pennell 2014).

Ancestral state reconstruction (ASR) methods critically rely on an assumed model for quantification of both the frequency and the strength of convergence. Although several kinds of ASR methods have been used to assess convergence, they all model a historical trajectory of evolution under an assumed model (almost always Brownian motion) and then analyze estimated ancestral phenotypes to detect or quantify convergence (reviewed in Stayton 2015a, 2015b; Arbuckle and Speed 2016; Speed and Arbuckle 2016). ASR methods have been used to study convergence since the dawn of the comparative methods era (e.g., Donoghue 1989; Brooks and McLennan 1991; Losos 1992) but have gained renewed popularity with the recent development of phylomorphospace tools for visualizing evolutionary trajectories and quantifying the frequency and strength of convergence (Sidlauskas 2008; Stayton 2011, 2015b). Stayton (2015b) and Speed and Arbuckle (2016) classified available ASR methods as process-free on the grounds that they do not assume that convergent evolutionary patterns were the result of adaptive mechanisms. These approaches are not truly process-free, however—they rely on parameter estimates from a model that assumes the observed patterns evolved under a process consistent with Brownian motion, such as genetic drift (as well as some, but certainly not all, alternative evolutionary mechanisms; Felsenstein 1988; Hansen and Martins 1996; O’Meara et al. 2006). ASR methods can employ alternative macroevolutionary process models, including models with explicitly adaptive processes, so long as it is possible to reconstruct ancestral states under such...
models (e.g., Elliot and Mooers 2014; Uyeda and Harmon 2014). However, this is rarely done, perhaps because tools for carrying out ASR have not kept pace with the rapid development of methods for fitting alternative models of trait evolution.

Naturally, evolutionary process models play a prominent role in methods for studying convergence that are explicitly based on model fitting. Such tools take one of two approaches. In the first, an investigator parameterizes an evolutionary process model in a way that explicitly incorporates hypothesized convergence events, fits the model to data, and then compares the fit to that of an alternative model lacking convergence. This is most commonly done using multiple-optimum OU models to represent the evolution of a clade on an adaptive landscape, with occasional peak shifts in which a lineage escapes the influence of its historical adaptive peak and is attracted to another (Hansen 1997; Butler and King 2004; Bartoszek et al. 2012; Beaulieu et al. 2012; O’Meara and Beaulieu 2014). Because it is straightforward to design OU models that permit independent lineages to evolve toward a shared adaptive peak, they provide a natural framework for the investigation of adaptive convergence. Generalized OU models permit a great deal of flexibility in parameterization, including multiple attraction strengths or rates of Brownian drift in addition to multiple optima (Beaulieu et al. 2012) and multivariate OU (Bartoszek et al. 2012), although highly complex OU models can suffer from parameter identifiability issues (Ho and Ané 2014; O’Meara and Beaulieu 2014; Cressler et al. 2015; Cooper et al. 2016b; Khabbazian et al. 2016). Most methods using OU models require a prior hypothesis for the phylogenetic placement of adaptive peak shifts, which limits their utility in estimating the frequency of convergence and precludes tests for convergence in clades where putatively convergent taxa have not been identified. These limitations can be avoided by using a second type of modeling approach in which the number or rate of evolutionary peak shifts is estimated as a model parameter. To achieve this, some such tools simply extend the multiple-peak OU modeling framework by automating the evaluation of candidate peak shift configurations (Ingram and Mahler 2013; Uyeda and Harmon 2014; Khabbazian et al. 2016), while alternative methods assume a Lévy process model of punctuated evolution (Eastman et al. 2013; Landis et al. 2013). Both techniques allow assessment of the frequency of evolutionary shifts, including convergence events (Eastman et al. 2013; Ingram and Mahler 2013).

**Why Treating Comparative Approaches as Process-Free Can Lead to Problems in the Study of Convergence**

For each of the comparative approaches to studying convergence outlined above, the resulting inferences about convergent evolution critically depend on the assumed model of the underlying evolutionary process. Quantitative measures of convergent pattern from a given method may differ markedly if an alternative model of the evolutionary process is assumed. This can be particularly problematic for Brownian motion–based ASR methods described as process-free which may overestimate or underestimate the frequency of convergence in groups evolving on rugged adaptive landscapes and yield inaccurate estimates of the strength or extent of convergence in any circumstance in which Brownian motion is a poor model of the true evolutionary process.

To illustrate, we consider a clade diversifying on an adaptive landscape in which several subclades have undergone peak shifts to much larger phenotype values but without convergence to the same optima (fig. 1). The reconstruction of ancestral states assuming Brownian motion imposes an averaging effect on ancestral phenotype estimates that is most pronounced at the root of the tree. This reconstruction substitutes the true pattern of iterated divergence from small to large phenotype values with a pattern in which at least four major lineages converge from intermediate to small values. Several ASR-based convergence metrics suggest substantial convergence in this clade (fig. 1A). In contrast, if the true (i.e., generating) Ornstein-Uhlenbeck model is instead assumed when carrying out ASR, the same pattern-based metrics accurately capture the lack of true convergence (fig. 1B).

This is a particularly striking example but we suspect not an unrepresentative one, due to the well-documented tendency of Brownian motion–based ASR methods to infer increasingly intermediate ancestral states for deeper nodes (Schluter et al. 1997; Oakley and Cunningham 2000). Similar problems can be anticipated any time ASR methods are used to investigate a group for which Brownian motion is not a reasonably good representation of the evolutionary process, and model misspecification can likewise lead to failure to identify convergence events or grossly inaccurate estimates of the strength or extent of convergence.

The issue we discuss here is shared across phylogenetic comparative biology. Hunt (2012) made similar points about the measurement of evolutionary rate, showing that measures based on process models were more meaningful across evolutionary timescales than traditional interval-based (and process-free) rate measures. However, these rate estimates were accurate only if the investigator assumed the correct model of evolution, due to the complex and model-specific relationship between evolution’s tempo (i.e., change over time) and mode (the process underlying this change). Due to the inseparability of tempo and mode, Hunt argued that the two must be considered in concert in studies of the evolutionary rate. A similar consideration applies to comparative studies of lineage diversification rates among clades, which are more accurately estimated using methods that assume an underlying diversification
model than with process-free methods that simply control for elapsed time (Rabosky 2012). These examples reflect a general truth—pattern and process are inseparable in the study of phylogenetic comparative data, and it is not possible to make inferences about evolutionary patterns without assuming something about the evolutionary process (Pagel and Harvey 1989; Freckleton et al. 2011; Hunt 2012). Not all comparative inferences are equal, of course, and some may be more robust than others to violations of model assumptions. However, convergence metrics that explicitly incorporate model-based reconstruction of ancestral phenotypes are likely to be particularly sensitive to violations of assumptions about the underlying evolutionary model (Oakley and Cunningham 2000).

Given the need to assume an evolutionary process to study convergence, how can one avoid the potential for adaptationist bias? The potential for such bias is irrefutable (Hansen 2014; Stayton 2015a, 2015b), but this is a risk that can be managed in part through conscientious consideration of alternative evolutionary models during analysis and careful interpretation of results (see box 2 for discussion of best practices). The fit of adaptive models should always be compared to nonadaptive alternatives, and interpretation should favor results obtained under the best-performing model or, in the absence of a single best model, results obtained under all plausible candidate models. However, applying these best practices can only take one so far in avoiding adaptationist pitfalls, due to the issue that evolutionary process models may fundamentally alter comparative inference about convergent evolution, we contend that there is no such thing as a process-free comparative measure of convergence.

Figure 1: Models assumed when reconstructing ancestral states can dramatically affect inferences about the frequency and strength of convergent evolution. Here, data for 20 species were simulated using a known phylogeny under a three-optimum Ornstein-Uhlenbeck (OU) process with strong selection and no convergent evolution ($a = 4$; $\sigma^2 = 1$; $\theta = 0, 3, 5$; total tree length = 1; colors represent correspondence to phenotypic optima). We reconstructed ancestral states assuming a standard Brownian motion model (A) and a three-optimum OU model (B) that closely represents the true evolutionary process under which the data were generated (the number of optima and phylogenetic locations of shifts between optima were known a priori; all other parameters were estimated). We then used ancestral state reconstruction–based comparative methods from Stayton (2015a) to estimate the frequency (C5) and magnitude (C2) of convergent evolution for the set of species with small phenotypes (species a–j; several related measures of the magnitude of convergence yield similar results but are not shown). C5 tallies the number of independent lineages that cross into the phenotype space occupied by the focal set of species (here this space is simply defined as the range of extant phenotype values for this set). C2 indicates the phylogenetic distance closed by evolution for this set of species and is calculated as the average value of ($D_{\text{max}} - D_{\text{max}}$) for all pairs of species in the focal set, where $D_{\text{max}}$ is the maximum phenotypic difference between a pair of species since their divergence and $D_{\text{int}}$ is the phenotypic difference between the same pair of species in the present. Note that assuming Brownian motion in this example (A) leads to an overestimate of the frequency and strength of convergence events in this subset of species (from intermediate to small phenotype values). If we assume the true OU model (B), we correctly infer no convergences (C5 = 0) and two divergences from small to large phenotype. Because assuming a different evolutionary model can fundamentally alter comparative inference about convergent evolution, we contend that there is no such thing as a process-free comparative measure of convergence.

Many-to-One Mapping of Real Evolutionary Processes to Modeled Processes

Most available phylogenetic comparative methods employ phenomenological models of the evolutionary process that
Box 2: Best Practices

The comparative method has important limitations that must be taken into account in any study (Losos 2011b; Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Cooper et al. 2016a), including studies that focus on convergence. Here we recommend several considerations that we think are essential to making strong inferences about historical patterns and processes of convergence.

Study Design

The comparative approach investigates patterns that result from uncontrolled natural processes rather than experimental manipulation (Freckleton et al. 2011). Nonetheless, the considerations that guide experimental design equally apply to comparative analyses.

Statistical replication is necessary for addressing most questions about convergence, although the relevant form and degree of replication can depend on the question being asked. For example, if one simply wishes to test whether two taxa have in fact converged, this can be tested with a simple model comparison or ancestral state reconstruction (ASR)-based test, although the scope of inference will be limited. Other study objectives will require relatively diverse clades to have any statistical power. For example, testing whether a shift to a new habitat is consistently associated with convergence will require a system containing numerous habitat shifts. It will almost never be possible to determine the cause of single convergence events using comparative methods alone because of the inability to rule out the possibility that an observed correlation is spurious (Gould and Lewontin 1979; Maddison and FitzJohn 2015).

Study design should also involve consideration of phylogenetic scale. Many evolutionary models can be useful at some scales but inadequate at others (Estes and Arnold 2007; Hunt 2012). For example, larger clades are more likely to have been shaped by a more heterogeneous mixture of evolutionary processes (Beaulieu et al. 2013). At the very smallest phylogenetic scales, it may not be possible to distinguish complex evolutionary models from simpler alternatives (Boettiger et al. 2012), even if the former better reflect reality.

Model Comparison

Alternative evolutionary process models can differ profoundly in how they reconstruct historical patterns (fig. 1), making it essential to compare models in any phylogenetic study of convergence. While model comparison has become routine in many areas of comparative biology (Posada and Crandall 1998; Harmon et al. 2010; Morlon 2014; Pennell et al. 2015), it is often neglected in phylogenetic studies of trait convergence. This may be because common tools for measuring convergence assume Brownian motion evolution; investigators interested in assuming alternative models must customize existing tools to do so. This is especially true for ASR methods (but see Elliot and Mooers 2014; Uyeda and Harmon 2014). Because ancestral state estimates can differ so profoundly under alternative evolutionary models, we suspect that a greater appreciation for the importance of model comparison might result from the development of more flexible ASR tools.

Although model comparison is essential for studying convergence, we caution against discussing results from alternative models on equal footing when some models clearly outperform others. Comparative studies commonly report and interpret the results of several alternative methods, with results from different methods regarded as complementary. This is to be encouraged when methods are internally consistent with one another but can be misleading when they assume different models of evolution, especially if these differences lead to meaningful differences in the quantification of convergent evolution. For example, if Brownian motion is found to yield a much worse fit to data than a multiple-peak Ornstein-Uhlenbeck (OU) model (such as in the toy example in fig. 1), the use of comparative methods that assume Brownian motion, such as most ASR and index methods, does not meaningfully contribute to our understanding of convergence and may even undermine it. Care should be taken that results that rely on alternative evolutionary process models are themselves regarded as fundamentally distinct (and potentially incompatible) rather than complementary per se.
Model Parameters and Model Adequacy

The parameters of fitted evolutionary models can be richly informative with respect to convergence, and many of the interesting features of evolutionary convergence that have inspired pattern-based tools may be effectively captured by the parameters of evolutionary models. For example, the strength of attraction in an OU model can be interpreted as a rate of adaptation in lineages that converge on a shared adaptive peak and may represent the balance between historical constraint and adaptation (Hansen 1997, 2012; Beaulieu et al. 2012; Collar et al. 2014).

Inspection of model parameters can also be used to identify when models provide a poor fit to data. For example, multi-optimum OU models frequently return estimates for some optima that fall outside the observed range of species trait data. This may reflect ongoing adaptation toward an extreme phenotype (Hansen 1997) or a mismatch between model assumptions and reality (Mahler and Ingram 2014). Some data sets cannot be fit well by multi-optimum OU models, highlighting the importance of testing whether a model can adequately reproduce key patterns in the data rather than simply assessing which model from a set of candidates fits best (Pennell et al. 2015). Simulation-based approaches can help ensure robust inference in virtually any scenario, including empirical conditions for which model performance may yet be unknown (Boettiger et al. 2012; Mahler et al. 2013; Elliot and Mooers 2014; Slater and Pennell 2014; Pennell et al. 2015; Clarke et al. 2017).

may plausibly represent more than one kind of evolutionary mechanism—that is, a many-to-one mapping of true evolutionary mechanism to modeled macroevolutionary process (Hansen and Martins 1996; O’Meara et al. 2006; Revell et al. 2008; Hansen 2012; Pennell 2014). For example, a single-peak OU model may represent adaptive evolution in a clade that has already reached a phenotypic optimum (Hansen 1997), or it could represent evolutionary stasis due to a constraint on the production of variation (e.g., Harmon et al. 2010). Many-to-one mapping is possible for a diversity of evolutionary process models, from Brownian motion to early burst and saltational macroevolutionary models (Freckleton and Harvey 2006; O’Meara et al. 2006; Mahler et al. 2010; Venditti et al. 2011; Pennell et al. 2014). Although many such models were introduced with specific microevolutionary mechanisms in mind, they describe variation at a comparatively coarse macroevolutionary scale and contain no direct link to such fine-scale mechanisms. The lack of mechanistic detail in these models presents another formidable limitation to the interpretation of macroevolutionary convergence. The investigator must consider such possibilities in any analysis of convergence—as we will argue below, such considerations can be aided by investigation of the study group using complementary approaches that allow more direct tests of mechanism and by considering the natural history of the organisms and their environments.

Integrative Approaches to the Study of Macroevolutionary Convergence

A complete understanding of any evolutionary phenomenon requires knowledge of both the detailed mechanisms by which evolutionary change occurs (i.e., the how) and the circumstances that ultimately bring about such change or shape its course (i.e., the why). Replicated convergence provides a powerful framework for both avenues of inquiry. The power of this replication has been harnessed in combination with high-throughput sequencing technologies in the last decade to greatly increase our understanding of the molecular mechanisms behind convergent phenotypic change (Elmer and Meyer 2011; Stern 2013; Rosenblum et al. 2014). By comparison, somewhat less progress has been made in understanding the ecological and phylogenetic context in which convergence occurs.

The phylogenetic comparative method can be especially useful for addressing questions about causes of convergent evolution because it is well suited for investigation at the large spatial and temporal scales at which such factors shape the evolution of biodiversity. In many cases, though, comparative models on their own may fail to distinguish among alternative hypotheses about the causes of convergent evolution, even when carefully applied. This is an intrinsic feature of the comparative approach that results from the many-to-one mapping of process to pattern in macroevolution (Hansen and Martins 1996; Pennell 2014). Thus, the comparative approach will often be much more powerful when integrated with complementary avenues of investigation. Here we discuss ways in which comparative inferences about the processes underlying convergent evolution can be strengthened by the incorporation of (1) more diverse causal mechanisms, (2) biogeography, and (3) fossil data into comparative approaches to studying convergence.

Incorporating a Greater Diversity of Causal Mechanisms into Evolutionary Process Models

A key goal in the study of evolution is to understand how microevolutionary mechanisms shape macroevolution, but elucidating this link has proved challenging (Uyeda et al.
2011; Rosindell et al. 2015). The comparative study of convergence can help to evaluate this relationship to the extent that we can make meaningful connections between potential causative factors and convergent pattern. Candidate factors that may result in macroevolutionary convergence include developmental mode (e.g., Wake 1982), genome architecture (e.g., Stern 2013), changes in climate (e.g., Yen et al. 2016), mutualistic interactions that involve phenotype matching (e.g., Hoyal Cuthill and Charleston 2015), repeated antagonistic interactions (e.g., Siepielski and Benkman 2007), and competition for nonsubstitutable resources (e.g., Abrams 1987; Scheffer and van Nes 2006). However, despite ongoing research interest in these areas, we still lack answers to basic questions such as, Do evolutionary shifts in reproductive system change the likelihood that clades will exhibit convergence? Is convergence due to abiotic selection more or less common than convergence due to biotic selection? and How likely is convergence to occur, persist, or break down under antagonistic or mutualistic selection?

One source of improvement may come from renewed attention to model development. As the scope of phylogenetic comparative methods has grown in recent years, models that focus on constrained or bounded evolution have received somewhat less attention than those inspired by explicitly adaptive mechanisms. New work in this area could help to diversify the scope of comparative inquiry (e.g., Boucher and Démery 2016). Future developments in the field should also work to clarify what kinds of macroevolutionary patterns we expect to arise from different ecological processes. Recently developed comparative models based on simple species interactions provide a welcome first step in this direction (Yoder and Nuismer 2010; Pennell and Harmon 2013; Nuismer and Harmon 2015; Drury et al. 2016; Clarke et al. 2017). Although these methods do not explicitly model convergence, the ecological processes underlying them may result in convergent phenotypes. In addition, recent years have seen the rapid development of a more sophisticated theory of species coexistence and community ecology (Chesson 2000; Hubbell 2001; Pennell and Harmon 2013; Vellend 2016), and future modeling efforts would do well to identify a set of expected evolutionary outcomes that reflects current ecological thinking. Combining modern ecological theory with the phylogenetic replication made possible in comparative studies of convergence will make for a powerful approach to studying the macroevolutionary signature of ecological mechanisms. We speculate that a principal roadblock to the development of more diverse and detailed macroevolutionary models has been the difficulty (or impossibility) of representing such models as closed-form likelihood expressions. Simulation-based approaches (including approximate Bayesian computation methods) may be required to fully diversify the models available in our macroevolutionary tool kit (e.g., Elliot and Mooers 2014; Clarke et al. 2017).

Even the development of improved models may not overcome the issue of many-to-one mapping of mechanism to evolutionary process model. However, hypotheses and models are not the same thing, and the utility of the comparative method depends on the ability of the investigator to use natural history knowledge and comparative tools together to craft specific mechanism-inspired hypotheses that can be tested using comparative data. In addition to improvements associated with integrating more diverse causal mechanisms into comparative methods, the study of convergence will be aided by designing studies that creatively use ecological experiments to test for specific ecological mechanisms acting to produce patterns observed at the clade level (Weber and Agrawal 2012). Ecological hypotheses about the drivers of convergence generally contain predictions about the relationship between trait similarity, abundance, and the relative fitness of species in a given community (table 1). For example, in instances where convergence is hypothesized to result from selection for Müllerian mimicry (i.e., selection for greater phenotype matching among aposematic species), experiments manipulating trait similarity in contemporary communities can be paired with phylogenetic tests of convergence (in relation to timing of sympatry). The prediction in this case is that (1) the reduction of phenotypic similarity decreases species fitness by increasing predation and (2) convergence occurred when species were in sympatry, not before. This integrative framework can be applied to antagonistic hypotheses as well and represents a powerful approach to testing adaptive hypotheses about the ecological drivers of convergence.

Integrating Biogeographic Approaches into Comparative Studies of Convergence

Integrating an understanding of species and clade biogeography can greatly enhance attempts to link ecological process to macroevolutionary pattern. Evolving lineages can directly interact only when they co-occur, and accounting for co-occurrence patterns will be essential if we are to distinguish the influence of species interactions on convergence from alternative factors. Furthermore, biogeographic perspectives can disentangle the influence of species range overlap from abiotic factors such as climate or soil type. The biogeographic approach is ripe for application to convergence generally, and in table 1 we outline several ways in which phylogenetic comparative methods may be combined with such an approach to augment their resolving power when investigating the ecological and evolutionary processes underlying convergent patterns.

The incorporation of a biogeographic perspective has yielded new insights in recent studies of replicated adap-
Table 1: Examples of hypothesized mechanisms driving patterns of convergence and predictions from integrated analysis

<table>
<thead>
<tr>
<th>Hypothesized underlying mechanism</th>
<th>Biogeographic predictions</th>
<th>Predictions for patterns of coexistence of convergent forms</th>
<th>Ecological predictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence due to chance</td>
<td>Convergence is independent of biogeography</td>
<td>Probability of convergence is independent of community context</td>
<td>Manipulating the density of a trait in a community does not change the selective value of the trait</td>
</tr>
<tr>
<td>Convergence due to selection driven by physical environment (e.g., climate, light)</td>
<td>Convergence is correlated with shared physical conditions</td>
<td>Probability of convergence is independent of range overlap with convergent species</td>
<td>Manipulating the density of a trait in a community does not change the selective value of the trait</td>
</tr>
<tr>
<td>Convergence due to competition resulting in niche partitioning and character displacement</td>
<td>Convergence may or may not be correlated with physical environment</td>
<td>Convergent forms evolve in allopatry but only where they are sympatric with a competitor</td>
<td>Negative density-fitness relationship: the abundance of a phenotype in the community decreases the selective value of that phenotype. Convergent communities should exhibit similar patterns of niche partitioning</td>
</tr>
<tr>
<td>Convergence due to competition for nonsubstitutable resources</td>
<td>Convergence may be correlated with particular abiotic conditions across phylogeny (e.g., an essential nutrient)</td>
<td>Convergent forms evolve in sympathy</td>
<td>Resource limitation leads to selection for greater similarity in the shared phenotype: supplementing resource decreases selection on this trait</td>
</tr>
<tr>
<td>Convergence due to facilitation/mutualism</td>
<td>Convergence may or may not be correlated with physical environment</td>
<td>Probability of convergence is independent of range overlap with convergent species</td>
<td>Average fitness is a positive function of the abundance of the mutualist and a negative function of the abundance of similar competitors</td>
</tr>
<tr>
<td>Convergence due to commensalism</td>
<td>Convergence may or may not be correlated with physical environment</td>
<td>Probability of convergence is independent of range overlap with convergent species</td>
<td>Average fitness is a positive function of the abundance of the commensal host</td>
</tr>
<tr>
<td>Convergence due to predation/parasitism</td>
<td>Convergence may or may not be correlated with physical environment</td>
<td>Convergent forms evolve in allopatry or sympathy but only where they are sympatric with an antagonist</td>
<td>Average fitness is a negative function of the abundance of the antagonist</td>
</tr>
</tbody>
</table>

Note: Linking pattern to process is a central challenge in comparative biology. However, in some cases, integrating multiple forms of inference can help researchers narrow possible pattern-to-process links. Here we provide several examples of how this framework could be applied to phylogenetic comparative studies of convergence. It is important to note that while inferring past causation with absolute certainty is impossible, explicitly considering the predictions of alternative hypotheses can help researchers identify mechanisms consistent with observed patterns. We provide several examples of mechanistic hypotheses, which are not necessarily mutually exclusive (a fact that should be accounted for in the design of a comparative study).

tive radiations in which entire well-structured ecological guilds have evolved convergently (Schluter 2000; Mahler and Ingram 2014). The existence of numerous replicated adaptive radiations suggests an important and deterministic role for interspecific competition and subsequent character displacement as a cause of convergence into the same set of niches (e.g., Frédérich et al. 2013; Gründler and Rabosky 2014; Esquerré and Keogh 2016; Moen et al. 2016). An alternative possibility, however, is that such convergence results more from biomechanical trade-offs involved in specializing on particular resources than from competitive interactions per se and that such specialists may have emerged via mechanisms other than interspecific competition. Here, information about range overlap can be informative. If interspecific competition played a role in the replicated evolution of ecological specialists, we would expect the convergent species to occur allopatrically and to have evolved only a single time in a given region, but we would have no such expectation if competition were not important in this divergence. This pattern is observed in replicated Greater Antillean Anolis radiations and in concert with experimental studies of both competition and character displacement (Pacala and Roughgarden 1982; Leal et al. 1998; Stuart et al. 2014), strongly suggests a role for competition in contributing
to the repeated evolution of similar ecomorphs on different islands (Mahler et al. 2013).

Other ecological mechanisms will leave distinct patterns in biogeographic patterns of convergent taxa. Cichlids in African lakes are well known for replicated radiations in different lakes (Wagner et al. 2012), but Muschick et al. (2012) showed that within Lake Tanganyika, numerous convergent species co-occur within the lake. The pattern of sympatric convergent taxa (see also Kozak et al. 2009; Ingram and Kai 2014) challenges the hypothesis of competition-driven character displacement and might instead suggest competitive character convergence (MacArthur and Levins 1967; Abrams 1987; Scheffer and van Nes 2006). While a great deal of work is needed to validate the hypothesis that competition can drive convergence between coexisting taxa across entire clades, the possibility highlights the need for an increased understanding of the expected biogeographic and macroevolutionary consequences of a broader range of ecological processes.

Revisiting the Fossil Record

The overwhelming majority of comparative phylogenetic studies are conducted using only extant species. Inferences from comparative analyses therefore suffer an unfortunate temporal asymmetry, whereby estimates of both historical pattern and process are associated with increasing levels of uncertainty as one looks further back in time (Schluter et al. 1997; Cunningham et al. 1998; Oakley and Cunningham 2000; Losos 2011b). For this reason alone, fossil data can make very large marginal improvements to the accuracy of comparative inference, and the incorporation of fossil information into comparative studies of convergence promises to help distinguish among alternative evolutionary models and refine parameter estimates of key evolutionary processes. Recent years have witnessed encouraging progress in the merging of paleontology and comparative phylogenetic methods, both with the development of integrative new models for phylogenetic inference and divergence dating (e.g., Ronquist et al. 2012; Heath et al. 2014; Drummond and Stadler 2016; Zhang et al. 2016) and for the fitting of comparative models of continuous trait evolution (Slater et al. 2012; Slater and Harmon 2013; Hunt and Slater 2016).

There has been enough progress to demonstrate that fossil data can dramatically improve comparative inference and in some cases shift the weight of evidence to alternative hypotheses (Slater et al. 2012; Mitchell 2015; Slater 2015; Hunt and Slater 2016). In studies of convergence, even one or a few fossil data points indicating trait values of ancestors may be critical in distinguishing between alternative scenarios (e.g., the histories depicted in fig. 1A, 1B). By anchoring ancestors in trait space, the incorporation of fossil data can potentially separate convergent pattern from process and allow more powerful hypothesis tests than can be achieved using information solely about extant species. The integration of fossils into molecular phylogenetic frameworks remains far from routine, however, and a key hurdle is the often fragmentary nature of fossil data and the considerable uncertainty often associated with phylogenetic placement of such fossils. Future efforts to address these challenges should yield large dividends for the comparative study of convergence.

Conclusions

The phylogenetic comparative method provides a rich set of tools for answering questions about convergence, including many that are otherwise inaccessible to biologists. Models of the evolutionary process are at the core of all comparative methods, however, and these models can profoundly influence the outcomes of comparative studies of convergence. This intrinsic link between pattern and process in phylogenetic comparative methods can be a liability if ignored but can be a powerful asset when models are explicitly used to test hypotheses about the ecological and evolutionary processes that give rise to phenotypic patterns. Future progress in the comparative study of convergence should result from development of more realistic models of ecological and evolutionary processes, better integration of comparative study with complementary research on biogeography and ecology, and the growing incorporation of fossil information into phylogenetic investigations currently dominated by extant taxa.

Acknowledgments

We thank A. Agrawal for inviting us to contribute to this special issue and for providing thoughtful feedback on early drafts of our manuscript. We also wish to thank T. Stayton as well as the presenters and many attendees of the 2016 American Society of Naturalists Vice Presidential Symposium for discussing these matters with us following our meeting presentation. Finally, we thank B. O’Meara, an anonymous reviewer, and members of the Mahler lab for insightful comments on our manuscript.

Literature Cited


Symposium Editor: Anurag A. Agrawal

“There are in all about twenty species of Draco, inhabiting the East Indies. The power of flight is not very great, but probably exceeds that of Ptychozoön.” Figured: Draco volans. From "Volant Adaptation in Vertebrates" by Richard S. Lull (The American Naturalist, 1906, 40:537–566).
Convergently Evolved Toxic Secondary Metabolites in Plants Drive the Parallel Molecular Evolution of Insect Resistance*

Georg Petschenka,1 Vera Wagschal,2 Michael von Tschirnhaus,3 Alexander Donath,4 and Susanne Dobler2,†

1. Institute for Insect Biotechnology, Justus-Liebig-Universität Giessen, Giessen, Germany; 2. Molecular Evolutionary Biology, Institute of Zoology, Universität Hamburg, Hamburg, Germany; 3. Fakultät Biologie, Universität Bielefeld, Bielefeld, Germany; 4. Zoologisches Forschungsmuseum Alexander Koenig, Zentrum für Molekulare Biodiversitätsforschung, Bonn, Germany


ABSTRACT: Natural selection imposed by natural toxins has led to striking levels of convergent evolution at the molecular level. Cardiac glycosides represent a group of plant toxins that block the Na,K-ATPase, a vital membrane protein in animals. Several herbivorous insects have convergently evolved resistant Na,K-ATPases, and in some species, convergent gene duplications have also arisen, likely to cope with pleiotropic costs of resistance. To understand the genetic basis and predictability of these adaptations, we studied five independent lineages of leaf-mining flies (Diptera: Agromyzidae). These flies have colonized host plants in four botanical families that convergently evolved cardiac glycosides of two structural types: cardenolides and bufadienolides. We compared each of six fly species feeding on such plants to a phylogenetically related but nonadapted species. Irrespective of the type of cardiac glycoside in the host plant, five out of six exposed species displayed substitutions in the cardiac glycoside-binding site of the Na,K-ATPase that were previously described in other insect orders; in only one species was the gene duplicated. In vitro assays of nervous tissue extractions confirmed that the substitutions lead to increased resistance of the Na,K-ATPase. Our results demonstrate that target site insensitivity of Na,K-ATPase is a common response to dietary cardiac glycosides leading to highly predictable amino acid changes; nonetheless, convergent evolution of gene duplication for this multifunctional enzyme appears more constrained.

Keywords: cardiac glycosides, Na,K-ATPase resistance, target site insensitivity, Agromyzidae, phylogeny, gene duplication.

Introduction

Evolutionary responses to common selective pressures are ideally suited to study the predictability of molecular adap-

tations, a major question in evolutionary genetics (Storz 2016). Recent studies of species that acquired resistance against toxins in their diet or the environment revealed striking levels of convergent molecular adaptations. Here the interaction of a noxious substance with a well-defined target site of a specific protein might facilitate the occurrence of molecular convergence, as specific amino acid substitutions can prevent docking of the toxin. Accordingly, across several classes of toxins, it has been shown that resistance was repeatedly achieved by substitutions of a few amino acids at specific positions within the target site of the receptor. Examples include predatory snakes that consume toxic amphibians and have convergently evolved resistance to tetrodotoxin (Feldman et al. 2012), poison frogs that acquire toxic alkaloids from arthropods to defend themselves against predators (Tarvin et al. 2016), and diverse insects that have evolved resistance to cyclodiene insecticides (Andreev et al. 1999) or pyrethroids (Rinkevich et al. 2013). An emerging model system for convergent evolution that has been studied in unprecedented detail is resistance to cardiac glycosides in insects as well as in vertebrates. Cardiac glycosides are potent toxins that inhibit the ubiquitous animal cation carrier Na,K-ATPase. They typically occur in plants but are also produced by some animals (e.g., toads, leaf beetles, and fireflies; Dobler et al. 2011; Agrawal et al. 2012). Cardiac glycosides are composed of a steroidal skeleton with a five-membered lactone ring at C17 in cardenolides and a six-membered lactone ring in bufadienolides (fig. 1). Both types either occur as glycosides with one or several sugars attached to C3 or as genins (Malcolm 1991; Agrawal et al. 2012). By binding to the Na,K-ATPase, they block the enzyme in the phosphorylated state and thus impair various physiological functions such as maintenance of an electrochemical gradient across the cell membrane and generation of neuronal action potentials (Horisberger 2004). Extensive mutagenesis experiments with the mammalian Na,K-ATPase and recent crystal structures revealed the specific
amino acid residues that are necessary for cardiac glycoside binding (Price and Lingrel 1988; Croyle et al. 1997; Qiu et al. 2005; Laursen et al. 2013, 2015). These data point to the prominent role of several amino acids in the transmembrane segments of the enzyme’s α-subunit that directly interact with the cardiac glycoside molecule and stabilize it in the extracellular binding pocket (fig. 1).

Resistance to cardiac glycosides often relies on substitutions of amino acids in the binding pocket of the Na,K-ATPase. The first evidence was found in Danaus plexippus, the monarch butterfly (Vaughan and Jungreis 1977), which sequesters toxic cardenolides from its host plants (Asclepias species) as a defense against predators. Later, the molecular basis was unraveled (Holzinger et al. 1992), and more recently it was shown that target site insensitivity in the monarch butterfly is caused by two amino acid substitutions (Aardema et al. 2012; Dalla et al. 2013) that emerged consecutively across the phylogeny of milkweed butterflies leading to a stepwise increase of resistance (Petschenka et al. 2013a). Target site insensitivity of the Na,K-ATPase has now been reported in at least six orders of insects with multiple evolutionary origins (Al-Robai et al. 1993; Dobler et al. 2012).

Figure 1: A, Structure of the pig Na,K-ATPase with bound ouabain (Yatime et al. 2011; Protein Data Bank ID 3N23). Here, α, β, and γ refer to the subunits of the Na,K-ATPase, of which γ has so far been described only for vertebrates. M = transmembrane domains, A = actuator domain, N = nucleotide-binding domain, P = phosphorylation domain. B, Enlarged and rotated view of the ouabain-binding cavity with relevant amino acids. A = alanine, E = glutamic acid, F = phenylalanine, G = glycine, I = isoleucine, N = asparagine, Q = glutamine, T = threonine, V = valine. C, Structure of cardiac glycosides. Top, hellebrin, a bufadienolide from Helleborus species; middle, ouabain, a standard cardenolide used for the physiological assays; bottom, calotropin, a common cardenolide from Asclepias species.

S30 The American Naturalist

This content downloaded from 132.236.027.111 on July 22, 2017 05:14:35 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).
2015; Zhen et al. 2012). While resistance can involve substitutions at least at five amino acid positions, the identical substitution of histidine for asparagine at position 122 occurred at least six times independently (Dobler et al. 2015). A comparable level of molecular convergence was found in cardiac glycoside-producing toads and their vertebrate predators (Ujvari et al. 2015).

In insects, it was furthermore shown that resistance-conferring amino acid substitutions may be associated with duplications of the Na,K-ATPase gene (Zhen et al. 2012). Such duplications might alleviate negative pleiotropic effects associated with increased resistance to cardiac glycosides. While resistance-conferring mutations may compromise the pumping activity of the Na,K-ATPase (Dalla and Dobler 2016), other important functions of this enzyme might not be impaired by the observed amino acid substitutions. Increasing evidence points to important roles of the Na,K-ATPase as a signal transducer and morphogenetic trigger, not only in vertebrates but also in insects (Liang et al. 2007; Paul et al. 2007; Li and Xie 2009). A balance between more or less substituted copies of the Na,K-ATPase and tissue-specific expression patterns may allow adjusting between effective ion transport and these other functions. Therefore, duplications of the Na,K-ATPase gene in cardiac glycoside-adapted insects have been interpreted as an evolutionary strategy to reduce negative pleiotropic effects (Zhen et al. 2012; Dalla and Dobler 2016).

Although cardiac glycosides are known to occur in 18 plant families (Krenn and Kopp 1998; Agrawal et al. 2012), almost all work on resistance has focused on insect herbivores feeding on plants in the family Apocynaceae. Here we studied the parallel adaptation of five independent lineages of leaf-mining flies (Agromyzidae) adapted to cardiac glycoside-producing plants (six fly species adapted to cardiac glycoside plants, including one generalist species) compared with five closely related fly species feeding on plants without cardiac glycosides. The host plants of these flies include species from four botanical families (Apocynaceae, Brassicaceae, Plantaginaceae, and Ranunculaceae) in which cardiac glycosides have themselves independently evolved. Thus, our data set comprises multiple independent comparisons of animal host use and plant toxin production, a convergently evolved interaction from the perspective of each partner.

Using our five independent contrasts of fly species, we tested how often amino acid substitutions evolved that have previously been assumed to confer resistance of the Na,K-ATPase. We were especially interested to test for repeated evolution of target site insensitivity by identical amino acid substitutions. To screen for such substitutions, we first analyzed Na,K-ATPase gene sequences amplified by reverse transcription polymerase chain reaction (RT-PCR). In species possessing putatively resistance-conferring amino acid substitutions, we further tested for the existence of duplicated Na,K-ATPase genes using a transcriptomic approach. Specifically, we wanted to investigate how strongly substitutions are associated with gene duplications. For three fly species that independently adapted to cardiac glycoside plants and two outgroup species, we additionally assessed resistance of their Na,K-ATPase in vitro and thereby confirmed the functionality of the observed molecular substitutions.

**Material and Methods**

**Natural History of Flies**

We compared six species of agromyzid taxa feeding on cardiac glycoside plants to related species whose host plants are not known to contain these toxins. The species pairs were chosen according to a priori assumptions about phylogenetic relationships based on morphology.

*Phytomyza digitalis* Hering, 1925, versus *Phytomyza crassiseta* Zetterstedt, 1860.—We compared *P. digitalis*, whose host plants are restricted to the genus *Digitalis* (Plantaginaceae), including *D. grandiflora*, *D. lutea*, and *D. lanata* (Hering 1927, 1951; www.bladmineerders.nl), which all produce cardenolides (Luckner and Wichtl 2000), to *P. crassiseta*, a specialist on various species of *Veronica* (Plantaginaceae; Hering 1927; www.bladmineerders.nl) not known to produce cardenolides.

*Phytomyza hellebori* Kaltenbach, 1872, versus *Phytomyza fallaciosa* Brischke, 1880.—We also investigated *P. hellebori*, which exclusively uses *Helleborus* (Ranunculaceae) species as hosts (Hering 1951). Several species of *Helleborus* (Ranunculaceae) are known to produce bufadienolides (e.g., hellebrin; Wisser and Kating 1974). While *H. foetidus* seems to be the preferred host plant at some locations (Ludwig 1907; www.bladmineerders.nl) and was reported to lack bufadienolides, this fly species also uses other *Helleborus* species as hosts (e.g., *H. niger*, *H. viridis*; Mortelmans et al. 2014; table 1) known to contain bufadienolides (Wisser and Kating 1974; Glombitza et al. 1989). The lack of cardiac glycosides in *H. foetidus* is questionable, as we have a clear indication of the presence of such compounds based on inhibition of Na,K-ATPase (G. Petshenka, unpublished data). As a comparison for *P. hellebori*, the *Ranunculus* feeder *P. fallaciosa* (Spencer 1990; www.bladmineerders.nl) was included.

*Napomyza scrophulariae* Spencer, 1966, versus *Napomyza lateralis* Fallén, 1823.—With respect to the other species we investigated, knowledge on the natural history of *N. scrophulariae* is very limited. The species seems to be specialized on *Digitalis purpurea* but may also use other hosts (Spencer 1966; www.ukflymines.co.uk). Nonetheless, the only specimens that were unambiguously determined as *N. scrophulariae* were all from *D. purpurea* (as were the specimens investigated here; table 1); accordingly, we refer to this species as a specialist. As a phylogenetic comparison, we employed *N. lateralis*, a polyphagous species that has a preference for Asteraceae (Spencer 1990; www.bladmineerders.nl).
<table>
<thead>
<tr>
<th>Fly species</th>
<th>Host plant</th>
<th>Collection site, year</th>
<th>Use</th>
<th>GenBank accession no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromatomyia horticola</td>
<td>Erysimum cheiranthoides</td>
<td>Pevestorf, Germany, 2011</td>
<td>RT-PCR (1)</td>
<td>LT795081 LT795089 LT795087 LT795098</td>
</tr>
<tr>
<td>Liriomyza eupatorii</td>
<td>Eupatorium cannabinum</td>
<td>Giessen, Germany, 2015</td>
<td>Transcriptome</td>
<td>LT795109 LT795097 LT795105 LT795101</td>
</tr>
<tr>
<td>Napomyza scrophulariae</td>
<td>Digitalis purpurea</td>
<td>Taunusstein, Waldkatzenbach, Germany, 2011</td>
<td>RT-PCR (2)</td>
<td>LT795080 LT795091 LT795086 LT795099</td>
</tr>
<tr>
<td>Napomyza lateralis</td>
<td>Asteraceae sp.</td>
<td>Tangstedt, Germany, 2011</td>
<td>RT-PCR (2)</td>
<td>LT795079 EF104710 EF104884 EF104796</td>
</tr>
<tr>
<td>Phytomyza crassleta</td>
<td>Veronica sp.</td>
<td>Waldkatzenbach, Germany, 2011</td>
<td>RT-PCR (2)</td>
<td>LT795078 EU367549 LT795085 EU367638</td>
</tr>
<tr>
<td>Phytomyza digitalis</td>
<td>Digitalis grandiflora</td>
<td>Mehrstetten, Germany, 2011</td>
<td>RT-PCR (2), in vitro assay, transcriptome</td>
<td>LT795083 LT795093 LT795106 LT795103</td>
</tr>
<tr>
<td>Phytomyza fallaciosa</td>
<td>Ranunculus sp.</td>
<td>Taunusstein, Germany, 2011, 2015</td>
<td>RT-PCR (2), transcriptome</td>
<td>LT795082 LT795094 LT795107 ...</td>
</tr>
<tr>
<td>Phytomyza ilicis</td>
<td>Ilex aquifolia</td>
<td>Giessen, A, Hamburg, A, B, Germany, 2011, 2016</td>
<td>RT-PCR (1), in vitro assay</td>
<td>LT795077 EU367541 LT795084 EU367630</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>(oregonR)</td>
<td>Laboratory culture</td>
<td>In vitro assay</td>
<td>AR044974 ... ... ... ...</td>
</tr>
</tbody>
</table>

Note: When specimens were obtained at different locations, letters in superscript indicate where specimens used for reverse transcription polymerase chain reaction (RT-PCR a), Na,K-ATPase in vitro assays (b), or transcriptomes (c) originated from. Numbers of biological replicates for reverse transcription polymerase chain reaction–based sequencing of Na,K-ATPase genes are given in parentheses.
**Liriomyza asclepiadis** Spencer, 1969, *versus Liriomyza eupatorii* Kaltenbach, 1873.—Another comparison involved *L. asclepiadis*, a specialized feeder on *Asclepias* species (Spencer 1990), and an unidentified *Liriomyza* species (referred to as *L. sp. Mexico*) collected from *Asclepias angustifolia* that is differentiated in its cytochrome oxidase I (COI) sequences from *L. asclepiadis* by 12.5% sequence divergence (Dobler et al. 2012). The *Asclepias* host plants of both species contain cardenolides (Agrawal et al. 2012). While *L. asclepiadis* is specialized on *Asclepias*, we have no information about the host range of the unidentified *Liriomyza* species. As it seems to be closely related to *L. asclepiadis* and no generalist leaf miners are known to use *Asclepias* species (Spencer 1990), we refer to it as a specialist. We compared these two *Liriomyza* species to *L. eupatorii*, which uses several genera of Asteraceae as hosts (www.bladmineerders.nl).

**Chromatomyia horticola** Goureau, 1851, versus **Phytomyza ilicis** Curtis, 1846.—Last, we sequenced the Na,K-ATPase gene of *C. horticola*, a highly polyphagous species (www.bladmineerders.nl), which we have found on the cardenolide-producing Brassicaceae *Erysimum cheiranthoides* (Lei et al. 1996). While all pairwise comparisons listed above were based on an a priori assumed phylogenetic relationship between fly species adapted to cardiac glycosides and nonadapted species, here we arbitrarily selected the holly leaf miner, *P. ilicis*, a specialist on *Ilex aquifolium*, as a comparison for *C. horticola*.

**Collection of Flies and Determination**
Species of Agromyzidae were determined by means of the morphology of mines and host plant species (www.bladmineerders.nl, www.leafmines.co.uk; Herding 1927, 1935). In addition, species identities were confirmed based on adult characters and verified by sequence data compared to reference sequences in GenBank. Host plant leaves containing mines were collected in the field (table 1) and kept under room conditions until the emergence of flies. Upon hatching and wing hardening, flies were frozen and stored at −80°C. In some cases, field-collected flies were also used for rearing to propagate specimens. Rearing was carried out in mesh cages under ambient conditions on potted host plants or cut branches in water. *Drosophila melanogaster* (Drosophilidae, strain oregonR) was kindly provided by the Max Planck Institute for Developmental Biology, Tübingen, Germany.

**RNA Extraction, RT-PCR, and Sequencing**
For all fly species mentioned above, sequences were obtained for the mitochondrial COI gene and the nuclear 28S and cinnamyl alcohol dehydrogenase (CAD) genes to include them in phylogenetic analysis and for their Na,K-ATPase genes. All sequencing was based on RNA extractions, as the Na,K-ATPase 1α gene consists of multiple exons separated by large intronic sequences (e.g., in *D. melanogaster*, the coding sequence of the commonest isofrom of the Na,K-ATPase α stretches over 10.195 bp and 8 exons). For RT-PCR, flies were decapitated and only heads were used, as their brains represent a rich source of Na,K-ATPase (Lebovitz et al. 1989). RNA was extracted with the RNase kit (Qiagen, Hilden, Germany) and reverse transcribed (Supercript III, Invitrogen, Carlsbad, CA). Na,K-ATPase sequences were amplified in standard PCR using the primers given in Dobler et al. (2012). This yielded a fragment of the Na,K-ATPase covering the amino acids 89–805. The numbers refer to the mature protein of the pig (Sus scrofa) to enable comparison with the extensive literature on vertebrate Na,K-ATPase including crystal structures of the protein (Yatime et al. 2011; Laursen et al. 2013, 2015). In addition, gene sequences for COI, 28S, and CAD were obtained for *C. horticola*, *L. sp. (Mexico)*, *N. lateralis*, *N. scrophulariae*, and *P. ilicis* (see table 1) using primers given in Scheffer et al. (2007) and additional primers for COI given in Maus et al. (2001).

**Transcriptome Analyses**
Transcriptomes based on mRNA preparations were obtained for *L. asclepiadis*, *L. eupatorii*, *P. digitalis*, *P. fallaciosa*, and *P. hellebori*. Since *Ranunculus* sp. and *Eupatorium canadrum* is hosts for several species of leaf-mining flies, identification of *L. eupatorii* and *P. fallaciosa* based on host plant and mine morphology was verified by barcoding individual flies that were later pooled for transcriptome analyses. For this purpose, DNA of 2–3 legs of each individual was extracted following the method detailed in Carvalho et al. (2009). A PCR product for COI was generated with primers S1634 and A2188 and sequenced to confirm species identity before transcriptome sequencing. RNA extractions of roughly 10 flies of *L. eupatorii*, *P. digitalis*, *P. fallaciosa*, and *P. hellebori* and of a single individual of *L. asclepiadis* were obtained as above. mRNA-specific cDNA library construction and paired-end sequencing on an Illumina HiSeq 2000 (Illumina, San Diego, CA) were carried out by StarSeq (Mainz, Germany) and yielded roughly 50 million reads for each species. All transcriptomes, except for *P. hellebori*, were assembled with Trinity 2.1.1 (Grabherr et al. 2011) using internal trimming with modified trimming options (SLIDINGWINDOW:4:30 LEADING:30 TRAILING:30 MINLEN:50). Reads of *P. hellebori* were first trimmed with sickle, version 1.29 (Joshi and Fass 2011; min-length: 50 bp), and then assembled using trinityrnaseq_r20140413p1 (Grabherr et al. 2011). Local Blast searches (Blastn 2.2.27 [Zhang et al. 2000] or Tblastn 2.2.27 [Altschul et al. 1997]) were used to identify gene copies of the Na,K-ATPase α-subunit and the genes needed for the phylogenetic analysis. Na,K-ATPase sequences were also checked for alternative gene cop-
ies with CLC Genomics (Qiagen) by mapping all trimmed reads against reference sequences identified in the Trinity assemblies. Paralogs of the Na,K-ATPase are usually easy to recognize since variation between gene copies by far exceeds allelic variation observed between individuals. Intra-individual variation accounts for only a few silent substitutions, and we extremely rarely observed nonsynonymous substitutions between individuals, the few cases being restricted to the large intracellular loop of the enzyme. Nevertheless, we acknowledge that we can recognize only gene copies that differ by at least a few substitutions and are expressed in the adult flies.

**Phylogenetic Analyses**

To test whether the chosen species pairs of Agromyzidae were suitable phylogenetic comparisons, we used COI, 28S, and CAD sequences from Scheffer et al. (2007) and Winkler et al. (2009) and extended the data set with sequences of the flies investigated here (see table 1 for GenBank accession numbers). From the original very extensive data set of Scheffer et al. (2007), we included all species of *Liriomyza*, *Napomyza*, and *Phytomyza* and three species of all major genera resolved between *Agromyzidae* and *Phytomyza*. Specifically, we chose three species of *Agromyzidae* ( *Aulagromyza*, *Calycomyza*, *Cerodontha*, *Ophiomyia*, and *Phytoliriomyza arctica*). In addition, we used CAD sequence data for *Phytomyza ilicis* and *Phytomyza crassisseta* from Winkler et al. (2009). In all cases where sequences from our own RT-PCR or transcriptome analyses were available as well as sequences from Scheffer et al. (2007) or Winkler et al. (2009), the sequences were identical or the differences negligible (few silent substitutions in the protein-coding genes).

Sequence alignment for COI (1,404 bp) and CAD (704 bp) was straightforward and immediately congruent with the alignment by Scheffer et al. (2007) archived in TreeBASE (www.treebase.org, accession no. SN3150). In contrast, 28S proved difficult to align so that we used the alignment of Scheffer and coauthors and manually added our sequences to this alignment in MEGA7 (Kumar et al. 2016). This resulted in a total length of 823 bp and the exclusion of 168 bases distributed in small groups along the gene that Scheffer and colleagues had judged impossible to align unambiguously. The aligned concatenated sequences were subjected to maximum likelihood analysis with RAxML (Stamatakis 2014) fitting a GTR+GAMMA model of sequence evolution for each of the three gene partitions separately. Na,K-ATPase sequences of the focal species were aligned with MEGA7 and subjected to maximum likelihood analysis assuming a GTR+GAMMA model of sequence evolution. For both data sets, bootstrap support values were generated in 500 replicated analyses. Alignments are archived in TreeBASE (www.treebase.org, accession no. 20548).

**In Vitro Assays of Na,K-ATPase**

Resistance of the Na,K-ATPase can be monitored in vitro by measuring the enzyme’s activity through spectrophotometric quantification of inorganic phosphate released from adenosine triphosphate (ATP) hydrolysis under increasing concentrations of cardiac glycosides. We assessed in vitro resistance of Na,K-ATPase to ouabain (a water-soluble standard cardenolide) in all specialist agromyzid species feeding on cardiac glycoside–containing plants except for *N. scrophulariae* and *L. sp.* (not enough material available). For comparison, we included the nonadapted agromyzid species *P. ilicis* and *D. melanogaster* as an outgroup. For Na,K-ATPase in vitro assays, flies were decapitated, and pooled heads (between 17–45) were homogenized in an all-glass grinder (Wheaton) in 200–500 μL deionized water. Homogenates were frozen at −80°C and freeze-dried. Lyophilisates were stored at −80°C until use. Na,K-ATPase assays were carried out as described previously (Petschenka et al. 2013a). Briefly, head extractions reconstituted with 300–400 μL deionized water were incubated at a concentration of 10⁻⁸–10⁻³ M ouabain, and Na,K-ATPase activity was measured as the amount of inorganic phosphate generated from ATP hydrolysis compared to a noninhibited control. We used OriginPro 2016 (OriginLab, Northampton, MA) for nonlinear curve fitting. Top and bottom asymptotes were set to 100 and 0, respectively. For graphic display, replicate data (2–4 biological replicates per fly species) were fitted using the “concatenate fit” option. Pairwise statistical comparisons of dose–response curves were carried out using the “compare data sets” option in OriginPro. For this function of the program, curves fitted through the means of the data have to be used. We evaluated *P* values of pairwise statistical tests after Holm–Bonferroni corrections. Due to the unusual shape of the curve obtained for *P. hellebori*, curve fitting using the means was not possible, and thus we excluded *P. hellebori* from statistical analysis. Raw data are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.20191 (Petschenka et al. 2017).

**Results**

**Phylogenetic Analyses**

Analysis of a comprehensive sample of agromyzid flies including our focal species corroborated the suitability of the chosen phylogenetic comparisons. The three molecular markers yielded a concatenated data set of 2,931 bp of 28S, COI, and CAD sequences and resulted in a well-supported tree that is in good agreement with the previous analysis of Scheffer et al. (2007; fig. 2). All genera as represented here are resolved as monophyletic groups, with the exception of the placement of *Chromatomyia horticola* within the genus *Phy-
tomyza. The lack of clear differentiation between Chromatomyia and Phytomyza species has been extensively demonstrated and discussed in Scheffer et al. (2007) and Winkler et al. (2009). More importantly, for the present study, our phylogenetic analysis corroborated the five independent colonizations of different plant lineages with cardiac glycosides, including the cardenolide type (in Asclepias, Digitalis, and Erysimum) and the bufadienolide type (in Helleborus; fig. 2). In particular, the chosen species pairs, Phytomyza digitalis versus Phytomyza crassigaster, Phytomyza hellebori versus Phytomyza fallaciosa, Napomyza scrophulariae versus Napomyza lateralis, Liriomyza asclepiadis/Liriomyza sp. Mexico versus Liriomyza eupatori, and C. horticola versus Phytomyza ilicis were confirmed as appropriate phylogenetic comparisons that contrast species adapted to cardiac glycosides to non-adapted relatives.

Sequence Analysis of Na,K-ATPase Genes

Our molecular screening for target site insensitivity mediated by mutations in the cardiac glycoside–binding pocket of the Na,K-ATPase gene revealed multiple convergently acquired amino acid substitutions. No such substitutions were detected in four out of five species that are not exposed to dietary cardiac glycosides, rather these species maintained the ancestral amino acid residues present in Drosophila melanogaster at all positions known to be part of the cardiac glycoside–binding pocket (fig. A1, available online). In contrast, multiple substitutions of amino acid residues were observed at these positions in five out of six species feeding on cardiac glycosides plants. Specifically, two residues at the border of the first extracellular loop, 111 and 122, which are well known to be critical for cardiac glycoside resistance in other insects, were re-
peatedly exchanged (fig. 3). An exchange of glutamine 111 for leucine (Q111L), which we previously reported for the unidentified *Liriomyza* sp. collected from *Asclepias angustifolia* in Mexico (Dobler et al. 2012) was observed in two further species, *P. digitalis* and *N. scrophulariae*. The same exchange is also present in the polyphagous *N. lateralis* and must have occurred in the common ancestor of both species, *N. lateralis* and *N. scrophulariae*. The substitution of histidine at position 122 for the conserved asparagine detected earlier by direct RT-PCR amplification in *L. asclepiadis* (Dobler et al. 2012) was confirmed here by our transcriptome analysis. The same substitution was once more discovered in *P. hellebori* and coincides with an additional exchange of glutamine 111 for histidine, a substitution that has not been previously reported in insects or vertebrates.

A duplication of the Na,K-ATPase gene could be detected in only one of the agromyzid flies analyzed here. In *P. hellebori*, our transcriptome analyses and RT-PCR revealed evidence for a second gene copy. These two copies have a P distance of 0.009, with seven nonsynonymous substitutions of 21 total. In addition, all five individuals analyzed by PCR were polymorphic at the same sites. One of the two gene copies features the conserved glutamine at position 111 and the conserved asparagine at position 111 and the conserved asparagine at position 122, while the second gene copy has a histidine at both positions (figs. 3, A1). An analysis of the read frequencies in the transcriptome showed that the double-substituted gene copy is roughly fourfold higher expressed than the one with the ancestral condition. Other than these differences, the transcriptomes provided evidence for only splice variants at the N-terminus of the enzyme and for mutually exclusive exons (exon 6) immediately downstream of the region analyzed here. These splice variants correspond to the well-described situation in *D. melanogaster* (Palladino et al. 2003).

**Figure 3:** Maximum likelihood analysis of Na,K-ATPase sequences of Agromyzidae (GTRGAMMA model). The amino acid residues at positions 111 and 122 are indicated along the branches and at the terminal nodes (numbering according to the mammalian Na,K-ATPase); the derived conditions are indicated in boldface and italic. Species on cardiac glycoside plants are boxed. The insert shows the relevant amino acids in the enlarged cardiac glycoside–binding cavity of the Na,K-ATPase. H = histidine, L = leucine, N = asparagine, Q = glutamine. Sequences are from transcriptomes (asterisk) or reverse transcription polymerase chain reaction (double asterisk). Branch lengths are drawn according to scale; numbers along branches indicate bootstrap values in 500 replicated analyses.
Na,K-ATPase Assays

Our in vitro assays of tissue-extracted fly Na,K-ATPases revealed different degrees of cardenolide resistance across the fly species tested (fig. 4). Na,K-ATPase of the non-agromyzid outgroup species D. melanogaster was strongly inhibited by ouabain (half maximal inhibitory concentration \([IC_{50}]\) = 4.6 × 10⁻⁷ M). The Na,K-ATPase of the agromyzid species P. ilicis, which is not adapted to cardiac glycosides, showed a similar response to ouabain (\(IC_{50} = 4.86 × 10^{-7} M\)), and statistical comparison revealed no difference between inhibition of the P. ilicis and D. melanogaster Na,K-ATPase (\(F_{2,8} = 3.802, P = .069\)). In contrast, Na,K-ATPases of P. digitalis and L. asclepiadis displayed 3 and 30 times increased cardenolide resistance, respectively (P. digitalis: \(IC_{50} = 1.47 × 10^{-6} M\); L. asclepiadis: 1.43 × 10⁻⁸ M). Comparison of both dose-response curves to the curve obtained from P. ilicis and comparisons among each other revealed significant differences (P. digitalis vs. P. ilicis: \(F_{2,8} = 100.664, P < .001\); L. asclepiadis vs. P. ilicis: \(F_{2,8} = 732.122, P < .001\); L. asclepiadis vs. P. digitalis: \(F_{2,8} = 504.26, P < .001\)). Phytomyza hellebori was a special case as ouabain inhibition of the P. hellebori Na,K-ATPase deviated from the monophasic sigmoidal shape of the other dose-response curves. The differing form of this curve resembled a two-component curve caused by two coexisting forms of Na,KATPase with different sensitivities to ouabain (Blanco and Mercer 1998). For fly-head preparations from P. hellebori and P. digitalis, we verified that ouabain is not metabolized under the conditions of our assay, demonstrating that the observed resistance is based on the biochemical properties of Na,K-ATPase and is not due to degradation of the inhibitor. Moreover, ouabain is known not to be metabolized by the Malpighian tubules of D. melanogaster (Torrie et al. 2004).

Discussion

The strategies of herbivorous insects to cope with toxic cardiac glycosides represent a striking case of convergent molecular evolution (Dobler et al. 2012, 2015; Zhen et al. 2012) that allows insights into the origin and underlying rules of adaptive amino acid substitutions (Storz 2016). The most striking observation is the high frequency of occurrence of one specific amino acid substitution in the Na,K-ATPase, N122H, that has evolved independently in six orders of phytophagous insects exposed to cardiac glycosides: in the Caelifera (S. Dobler, G. Petschenka, and V. Wagschal, unpublished data), Coleoptera (at least twice), Diptera, Hemiptera, Hymenoptera, and Lepidoptera (Holzinger et al. 1992; Dobler et al. 2012, 2015; Zhen et al. 2012). The present study increases the number of independent origins of the N122H exchange, as we show that this substitution also evolved twice independently.

Figure 4: Sensitivity of dipteran Na,K-ATPases to ouabain in vitro. Data points represent the mean of 2–4 biological replicates (± SE); the number of actual replicates per species is given in parentheses after species names. Solid lines indicate species that are exposed to cardiac glycosides in their diets; dashed lines represent species that are not naturally exposed to the toxins. Dark blue = Phytomyza ilicis (4), green = Drosophila melanogaster (4), purple = Phytomyza digitalis (4), light blue = Phytomyza hellebori (4), orange = Liriomyza asclepiadis (2). Note the deviating characteristics of P. hellebori.

This content downloaded from 132.236.027.111 on July 22, 2017 05:14:35 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).
in the Agromyzidae (fig. 5). As we will discuss below, further potentially resistance-conferring amino acid substitutions have repeatedly evolved in the cardiac glycoside–binding pocket of the Na,K-ATPase in insects as well as in vertebrate predators exposed to cardiac glycosides in their diet.

As far as we can tell, a remarkable difference from our insects to other cases of convergent molecular evolution consists of the lack of intraspecific variability in resistance to dietary cardiac glycosides. Neither sequence analyses of the Na,K-ATPase α gene (Aardema et al. 2012; Pierce et al. 2016; S. Dobler and V. Wagschal, unpublished data) nor in vitro enzyme assays testing for the resistance of the corresponding protein to cardiac glycosides (fig. 4; Petschenka et al. 2013a; Bramer et al. 2015) so far revealed evidence for noticeable intraspecific variability. This contrasts with other well-investigated systems where the molecular foundation of adaptive traits has been unraveled, such as the blanched coloration displayed by lizards at White Sands.

Figure 5: Compilation of available data on Na,K-ATPase sequences in insects adapted and nonadapted to cardiac glycosides. The overall tree and node ages correspond to the best-supported insect phylogeny to date (Misof et al. 2014) that has been complemented by phylogenetic hypotheses presented in the Tree of Life Web Project (Maddison and Schulz 2007) and our own data. Species in black are not exposed to cardiac glycosides, species in orange are known to sequester cardiac glycosides from their host plants, and species in blue feed on cardiac glycosides plants but either are not known or not able to sequester cardiac glycosides. To the right are amino acid residues known to form the cardiac glycoside–binding pocket of the Na,K-ATPase (see fig. 1), with potential resistance-conferring exchanges highlighted by gray boxes. The exchange of E312D (glutamic acid for aspartic acid) between vertebrates and all insects is considered negligible. Stars indicate gene duplications. In *Lygaeus kalmii* and *Oncopeltus fasciatus*, both indicated copies are present plus a third less heavily substituted one featuring N122H and F786S exchanges. In *Rhyssomatus lineaticollis*, a second copy has the ancestral Q111 but a N122Y substitution. In *Chrysochus auratus*, the second copy is ancestral except for a Q111L exchange (Zhen et al. 2012). In *Phytomyza hellebori*, these two copies have the ancestral state. For GenBank accession numbers, see table 1.
(Rosenblum et al. 2010; this issue) and beach mice (Hoekstra et al. 2006; Mullen and Hoekstra 2008) or insect resistance to pyrethroids (Rinkevich et al. 2013). In these cases, heritable variability persists since the selective pressures vary between habitats. In the case of adaptations to dietary cardiac glycosides, adoption of these feeding habits lies in the distant past and the favorable mutations have apparently been driven to fixation (Aardema et al. 2012). A screen for intraspecific variability might be most promising in generalist insect species if populations vary in host use, feeding locally on plants with or without cardiac glycosides.

While we detected a striking amount of molecular convergence in resistance-conferring substitutions in the Na-K-ATPase, this strategy is not the only possible adaptation to avoid the toxic effect of dietary cardiac glycosides. Impermeable guts that prevent the toxins from entering the body, or fast and efficient elimination from the hemolymph via the Malpighian tubules, represent possible alternatives and have been detected in nonadapted as well as cardiac glycoside-adapted insects (Scudder and Meredith 1982; Torrie et al. 2004; Petschenka and Dobler 2009; Petschenka and Agrawal 2015). In addition, a protection of the sensitive neural tissue expressing high levels of Na-K-ATPase by active carriers in the insects’ blood-brain barrier plays an important role and is possibly sufficient to cope with small concentrations of hemolymph cardiac glycosides (Petschenka and Dobler 2009; Petschenka et al. 2013b). Thus, alternative evolutionary trajectories are possible to cope with dietary cardiac glycosides.

Remarkably, convergent molecular evolution of cardiac glycoside resistance by amino acid substitutions has also evolved in several vertebrates feeding on bufadienolidel-producing toads (Jaisser et al. 1992; Ujvari et al. 2013, 2015; Mohammadi et al. 2016). As in insects, amino acid substitutions in the first extracellular loop of the Na-K-ATPase mediate resistance to these toxins in their diet. It is, however, noteworthy that in vertebrates adapted to cardiac glycosides, the substitution of N122H observed so often in insects has never been described. Rather, other amino acid replacements are present at position 122 (Price and Lingrel 1988; Jaisser et al. 1992; Ujvari et al. 2013, 2015). Two reasons could account for these differences: First, the genetic background of the remaining enzyme may differ between insects and vertebrates, and thereby, the effect of specific substitutions may also vary (e.g., Poelwijk et al. 2007; Storz 2016). Mutagenesis experiments that introduced the N122H exchange into the human and the Drosophila Na,K-ATPase in cell culture corroborate this idea (S. Dobler, S. Dalla, and F. List, unpublished data). Second, the toxins the two groups are exposed to may require different adaptations. While the investigated vertebrates have evolved resistance to toad bufadienolides, most insects analyzed so far are exposed to plant cardenolides. Recent crystallography, however, suggests that both types of toxins occupy the binding cavity in a similar fashion (Laursen et al. 2015). Our data further support that the optimal adaptation does not depend on the type of cardiac glycoside (or other aspects of host plant variation, as a histidine at position 122 was found in Phytomyza hellebori exposed to bufadienolides (and the same applies to the savwly Monophadnus latus; Dobler et al. 2015), while Liriomyza asclepiadis, like the other species featuring the N122H exchange, is exposed to cardenolides in its Asclepias host plants (fig. 5).

In many cases, in insects and vertebrates alike, substitutions at position 122 are accompanied by exchanges at position 111. The two residues Q\(^{111}\) and N\(^{122}\) are situated at the entrance of the cardiac glycoside–binding cavity and have long been known to be of special importance for cardiac glycoside–binding affinity (Price and Lingrel 1988; Price et al. 1990; Croyte et al. 1997). As the crystal structures show, binding of cardiac glycosides leads to a conformational change bringing the αM1–M2 loop closer to the ligand, and these two residues get close enough for hydrogen bonding (Yatime et al. 2011; Laursen et al. 2015). The replacement of Q\(^{111}\) and N\(^{122}\) by residues not capable of hydrogen bonding or having bulkier and charged side chains like histidine apparently prohibit high-affinity cardiac glycoside binding. Experimental evidence suggests that substitutions at positions 111 and 122 have synergistic functional effects on resistance (Price et al. 1990; Dobler et al. 2012; Dalla et al. 2013; Dalla and Dobler 2016).

Compared to position 122, where we observed only exchanges of N122H, position 111 is more variable as to which amino acid substitutions cause resistance to cardiac glycosides. At this position, leucine, valine, threonine, and glutamic acid have been repeatedly observed in cardiac glycoside–adapted insects (Labeyrie and Dobler 2004; Dobler et al. 2012, 2015; Zhen et al. 2012), and here for the first time an exchange of Q\(^{111}\) for histidine is reported. Position 111 is of importance in vertebrates adapted to cardiac glycosides as well, and the observed exchanges partially match those present in insects. In snakes and lizards known to feed on toads, the Q111L exchange has evolved four times convergently, while toads, rodents, the European hedgehog, and the frog Leptodactylus latrans display a substitution of arginine for Q\(^{111}\) (Price and Lingrel 1988; Jaisser et al. 1992; Ujvari et al. 2013; Ujvari et al. 2015). Though in vertebrates the exchanges of Q\(^{111}\) are also coupled with substitutions at the C-terminal end of the first extracellular loop of the Na-K-ATPase, these exchanges (at positions 119, 120, or 122) differ from those observed in insects. Overall, this demonstrates an astonishingly high degree of molecular convergence across animal phyla, but the use of different (presumably optimal) substitutions at the same time provides evidence for constraints (or phylogenetic bias) that may be due to either the differences in the genetic background of the vertebrate versus insect Na-K-ATPase or different physiological conditions.
Although Q111L seems to be associated with resistance in vertebrates, the functional effect of this substitution has not been tested in isolation, that is, without the accompanying substitutions at the positions mentioned above. While the observed resistance in *Phytomyza digitalis* coincides with Q111L, there is no evidence that this substitution actually increases resistance of Na,K-ATPase in several lepidopteran insects, including nymphalid butterflies, arctiid moths, and the oleander hawk moth (Petschenka et al. 2012, 2013a; Dobler et al. 2015). Compared to butterflies and moths, the situation in leaf-mining flies might be different, and the effect of Q111L needs to be investigated in more detail. Along the same lines, it is currently unclear why the Na,K-ATPase of the generalist *Napomyza lateralis* carries this substitution, although the species might not be exposed to cardiac glycosides in its diet.

In insects as well as in vertebrates, the available data suggest that amino acid substitutions in and at the border of the first extracellular loop (M1–M2) are much more likely to evolve than exchanges in the remainder of the cardiac glycoside–binding cavity. According to the Na,K-ATPase crystal structures, bulky and hydrophobic side chains of the transmembrane helices M4–M6 form a general docking platform for the steroid core of all cardiac glycosides, while their lactone rings fit in a hydrophobic funnel of these transmembrane helices that is leading to the cation binding sites of the enzyme (Laursen et al. 2015). The proximity to the cation binding sites seems to create functional constraints that may explain why amino acid substitutions in these parts of the enzyme are so rare. So far they have been observed only in lygaeid bugs that possess three copies of the Na,K-ATPase α1 gene and in aphids where retrotransposition seems to have created a second gene copy with altered properties (Zhen et al. 2012). Functional data support the notion that the amino acid exchanges observed in transmembrane domains M5 and M6 of *Oncopeltus fasciatus* seriously compromise the enzyme’s activity as an ion transporter (Dalla and Dobler 2016).

Testing the functional effects of the observed amino acid substitutions is critical for understanding the value of single replacements versus combinations as well as potential pleiotropic constraints. Two approaches can be followed: enzyme assays, as used here, to test for the sensitivity of the native enzyme that is expressed at high levels in the nervous tissue of animals (Vaughan and Jungreis 1977; Moore and Scudder 1986; Al-Robai 1993; Petschenka et al. 2012) or genetic modification of Na,K-ATPase genes followed by evaluation of the physiological effects of the introduced mutations in enzymes expressed in cell culture (Price et al. 1990; Holzinger et al. 1992; Qiu et al. 2005; Dalla et al. 2013; Ujvari et al. 2013; Dalla and Dobler 2016). Both lines of evidence demonstrate that the exchange of N122H causes a high level of resistance to cardiac glycosides (Holzinger et al. 1992; Dalla et al. 2013; Petschenka et al. 2013a; Bramer et al. 2015). In accordance with these results obtained from genetically engineered *Drosophila* Na,K-ATPase, the inhibition curve obtained for *L. asclepiadis* shows that the single exchange of N122H in an enzyme that is otherwise highly similar to the sensitive *Drosophila* Na,K-ATPase is sufficient to increase resistance by 31-fold.

The heterogeneity of Na,K-ATPase sequences across the fly species adapted to cardiac glycosides is mirrored by the different sensitivities of their Na,K-ATPases to ouabain. *Drosophila melanogaster* and *P. ilicis* display similar sensitivities to ouabain, suggesting that the Na,K-ATPase of *P. ilicis* likely represents the ancestral condition in the Agromyzidae. The dose-response curve of *P. digitalis* Na,K-ATPase in comparison displays a threefold-higher resistance to ouabain, thus supporting adaptation to its host plant’s cardenolides. The highest in vitro resistance in our data set was observed for *L. asclepiadis*. However, *P. hellebori* may have an even more resistant form of the Na,K-ATPase, but its interpretation is confounded by the coexistence of two different forms of Na,K-ATPase. In *P. hellebori*, the dose-response curve can be interpreted as displaying the behavior of two forms with differing sensitivity: while inhibition of the sensitive copy (with the ancestral Q111 and N122 amino acids) leads to a rapid activity decline at relatively low concentrations of ouabain, the resistant copy (with H111 and H122) maintains a level of activity at millimolar ouabain concentrations that is higher than for *L. asclepiadis*. This resembles the situation described for rat testis, where both the α1 and α4 isoforms, with strongly differing ouabain sensitivities, are expressed together (Blanco and Mercer 1998).

Gene duplications as observed here for *P. hellebori* could buffer the detrimental effects of substitutions by counteracting purifying selection on individual proteins and could have a direct advantage due to expression of higher gene doses (Soskine and Tawfik 2010; Kondrashov 2012). Other models, in contrast, suggest that gene duplication is a neutral event that may not be under selective pressure (Soskine and Tawfik 2010). For insects adapted to cardiac glycosides, previous analyses suggested that resistance-conferring mutations may go along with duplications of the Na,K-ATPase α1 gene (Zhen et al. 2012; indicated by stars in fig. 5). This is especially prominent in lygaeid bugs where two duplications gave rise to three gene copies that differ in the number of substitutions, including unusual exchanges toward the C-terminus of the enzyme (Zhen et al. 2012). The Agromyzidae *P. hellebori* provides a further example for a duplication of the Na,K-ATPase gene that resulted in an unaltered, ancestral form and a fourfold-higher expressed form with two resistance-conferring mutations in the M1–M2 domain (Q111H, N122H). As for the debate whether substitutions precede duplications or only follow later when the duplicate is freed from purifying selection (Soskine and Tawfik 2010), this content downloaded from 132.236.027.111 on July 22, 2017 05:14:35 AM All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).
fik 2010), *P. hellebori* provides an example for an accumulation of mutations post duplication, as the ancestral copy that is shared by the closely related *P. fallaciosa* is still present. The duplication may have been driven by the advantage of increased gene dosage in the face of toxins compromising gene function and subfunctionalization of the dupli-
cates ensued later (Kondrashov 2012).

Duplications of the Na,K-ATPase gene can alleviate negative pleiotropy imposed by the dual role of the Na,K-ATPase for ion transport and morphogenesis (Xie and Askari 2002; Palladino et al. 2003; Horisberger 2004; Paul et al. 2007) and allow for resistance-conferring substitutions that compromise ion transport (Dalla and Dobler 2016). Nevertheless, in our analysis of Agromyzidae on cardiac glycoside plants, we detected four incidents of substitutions in the cardiac glycoside-binding pocket but only a single gene duplication event, de-
tpite the fact that all five independent lineages started from the same ancestral state. This suggests that in the case of the Na,K-ATPase, constraints on gene duplications seem to be stronger than constraints on amino acid substitutions, at least as much as substitutions occur only at the favored resi-
dues of the first extracellular loop and the bordering trans-
membrane domains (M1–M2). We still know too little about additional roles of insect Na,K-ATPases genes beyond ion transport (Paul et al. 2007), but the evidence accumulating for vertebrate Na,K-ATPases suggests that pleiotropic con-
straints by Na,K-ATPase-initiated signaling cascades may be immense (Xie and Askari 2002; Schoner and Scheiner-
Bobis 2007; Li and Xie 2009). The functional constraints of this enzyme imposed by its dual role thus seem to enforce a strongly canalized evolutionary path.

Acknowledgments
We thank A. Agrawal for the opportunity to contribute to this symposium, for helpful discussions, and for his critical reading of the manuscript. He, A. Hastings, and T. Züst helped with the collection of *Liriomyza asclepiadis*. M. Falkenberg and R. Trusch supported the rediscovery of *Phyl-
tomyza digitalis* in Germany. L. M. Arcila Hernandez, A. Baudach, R. Halitschke, M. Schott, and S. Stiehler helped with collecting leaf-mining flies and data. A. Vilcinskas gener-
ously provided materials for some parts of this study to G.P. The project was funded by a grant from the Templeton Foundation and Deutsche Forschungsgemeinschaft (DFG; Do 527/5-3) to S.D. G.P. was financially supported by the DFG (PE 2059/1, PE 2059/2).

Literature Cited


Al-Robai, A. A. 1993. Different ouabain sensitivities of Na⁺/K⁺-
ATPase from *Poekilotherus bufonius* tissues and a possible physi-
ological cost. *Comparative Biochemistry and Physiology B: Bio-


Bramer, C., S. Dobler, J. Deckert, M. Stemmer, and G. Petschenka. 2015. Na⁺/K⁺-ATPase resistance and cardenolide sequestration: basal adaptations to host plant toxins in the milkweed bugs (*He-


Dalla, S., and S. Dobler. 2016. Gene duplications circumvent trade-


Dobler, S., S. Dalla, V. Wagchal, and A. A. Agrawal. 2012. Community-


Dobler, S., G. Petschenka, V. Wagchal, and L. Flacht. 2015. Conver-


This content downloaded from 132.236.027.111 on July 22, 2017 05:14:35 AM. All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).


Symposium Editor: Anurag A. Agrawal
Symposium

Convergent Phenotypic Evolution despite Contrasting Demographic Histories in the Fauna of White Sands*

Erica Bree Rosenblum,1,2,† Christine E. Parent,1,2,§ Eveline T. Diepeveen,1,2,∥
Clay Noss,1 and Ke Bi1,3

1. Department of Environmental Science, Policy, and Management, University of California, Berkeley, California 94720; 2. Museum of Vertebrate Zoology, University of California, Berkeley, California 94720; 3. Computational Genomics Resource Laboratory, University of California, Berkeley, California 94720

Online enhancements: appendices.

ABSTRACT: When are evolutionary outcomes predictable? Cases of convergent evolution can shed light on when, why, and how different species exhibit shared evolutionary trajectories. In particular, studying diverse species in a common environment can illuminate how different factors facilitate or constrain adaptive evolution. Here we integrate studies of pattern and process in the fauna at White Sands (New Mexico) to understand the determinants of convergent evolution. Numerous animal species at White Sands exhibit phenotypic convergence in response to a novel—and shared—environmental condition: geologically young gypsum dunes. We synthesize 15 years of research on White Sands lizards to assess the contribution of natural selection. Overall, we find dramatic phenotypic convergence across diverse species at White Sands. Although the direction of phenotypic response is parallel, the magnitude of phenotypic response varies among species. We also find that species exhibit strikingly different demographic patterns across the ecotone. The species with the most genetic structure between White Sands and dark-soil populations generally exhibit the least phenotypic divergence, suggesting population demography as a key modulator of adaptation. Comparative studies are particularly important for understanding the determinants of convergence in natural systems.

Keywords: White Sands, convergent evolution, jumping spiders, sand-treader crickets, restriction site–associated DNA sequencing.

Understanding the factors that predict evolutionary response to natural selection is a central goal of evolutionary biology. Instances of convergent evolution—the independent evolution of phenotypic similarity in different lineages—provide unparalleled opportunities to study the mechanisms of adaptive trait evolution. Ultimately, whether different species exhibit a similar evolutionary response depends critically on three interacting determinants (reviewed in Rosenblum et al. 2014): (1) natural selection (e.g., whether the mode, strength, and dimensionality of selection is similar across species), (2) genetic architecture of adaptive traits (e.g., whether heritability, mutational effect sizes, and patterns of epistasis are parallel across species), and (3) population demography (e.g., whether population size, patterns of population structure, and dynamics of gene flow are similar across species).

A key challenge for understanding evolutionary predictability—even for cases of convergent evolution—is disentangling the factors that can promote or hinder adaptive trait evolution. When different lineages in different environments are compared, patterns of phenotypic evolution can seem idiosyncratic, and it can be difficult to distinguish the contributions of selection, demography, and genomic architecture to observed patterns. In contrast, cases of repeated evolution within a single community provide exciting opportunities to differentiate among the determinants of adaptive evolution. When distantly related species exhibit similar phenotypic responses in a shared environment, we can address more nuanced hypotheses about evolutionary predictability.

Here, we integrate studies of pattern and process in the White Sands system to understand the factors that promote and hinder convergent evolution. We first synthesize what we have learned about the patterns of convergent evolution—and the mechanisms influencing that pattern—from 15 years of work on White Sands lizards. We then present new data on two invertebrate species as a next step toward a community-scale synthesis. Finally, we highlight outstanding questions about convergent evolution at White Sands,
with particular attention to the interplay among natural selection, population demography, and genomic architecture. Throughout, we use a pattern-based definition of convergent evolution: the independent evolution of phenotypic similarity in different lineages. Defining convergence as a pattern—observed at the phenotypic level—allows us to be more explicit as we endeavor to link process to pattern, mechanism to outcome, and cause to consequence (reviewed in Rosenblum et al. 2014 and Stayton 2015).

White Sands as a Stage for Convergent Evolution

White Sands is a striking and geologically recent formation in south-central New Mexico. The expansive gypsum dune system formed after the Last Glacial Maximum, with the bulk of the sedimentation deposition in the past 2,000–7,000 years (Langford 2003; Kocurek et al. 2007). At ~650 km², White Sands is the largest gypsum dune field in the world. The gypsum sands contrast dramatically with the brown soils of the surrounding Chihuahuan Desert, creating a divergent selective environment. Many diverse animal species have colonized White Sands and have convergently evolved blanched coloration in the dune habitat (fig. 1).

The best-studied example of phenotypic convergence at White Sands is in the lizard fauna. The three lizard species that inhabit the heart of the dunes (Sceloporus cowlesi [Eastern fence lizard], Holbrookia maculata [lesser earless lizard], and Aspidoscelis inornata [little striped whiptail]) all exhibit dramatically blanched dorsal coloration. In contrast, populations of all three species exhibit darker dorsal coloration in the rest of their ranges, typically similar to local substrate colors (Degenhardt et al. 1996). The White Sands lizards exhibit similar general patterns of color evolution. In all three species, the change in coloration is in the same direction and is explained primarily by the brightness aspect of color (i.e., changes in brightness explain >80% of the interpopulation variation in color; Rosenblum 2006). Evolution of dorsal brightness is not only in the same direction across species but also of roughly similar magnitude across species (Rosenblum 2006; Rosenblum and Harmon 2011). While explaining less of total variation in dorsal color, changes in the hue and chroma aspects of coloration also trend in the same direction for all three species (Rosenblum 2006; Robertson and Rosenblum 2009).

In addition to dorsal color variation, sexual-signaling patches are also significantly different in color between dark and light lizards in all three species (Robertson and Rosenblum 2009). These differences are not consistent across species, which is unsurprising, given that the location, color, and use of sexual-signaling color patches vary across species (e.g., blue ventral color in male S. cowlesi, seasonal orange throat color in female H. maculata, and blue head color of male A. inornata; Degenhardt et al. 1996). Differences in sexual-signaling color patches between dark and light color morphs could contribute to observed preference for local mates in White Sands S. cowlesi and H. maculata (Rosenblum 2008; Hardwick et al. 2013), but the potential for incipient speciation in White Sands lizards is outside the scope of our focus here on convergent evolution.

In addition to convergent evolution of coloration in White Sands lizards, the three species exhibit shifts in a variety of other phenotypes. Although detailed studies of heritability for these traits have not been conducted, we observe striking differences between dark-soil and White Sands populations in a myriad of traits, including morphology (e.g., body size, limb length; Rosenblum and Harmon 2011), performance (e.g., sprint speed, bite force; Des Roches et al. 2013, 2016), resource use (e.g., diet, microhabitat; Des Roches et al. 2011, 2015, 2016), and behavior (e.g., predator wariness, mate choice; Rosenblum 2008; Robertson et al. 2011; Hardwick et al. 2013). For example, in all three species, White Sands lizards tend to have relatively larger heads size than dark-soil lizards (Rosenblum and Harmon 2011; Des Roches et al. 2016). Larger head size, in turn, correlates with stronger bite force, and bite force is associated with dietary differences between light and dark populations (Des Roches et al. 2016). Specifically, for all three species, White Sands lizards consume a more varied diet than dark-soil lizards (i.e., with stronger bite and wider gape, larger and harder-bodied prey can be included in the diet; e.g., Herrel et al. 2001; Des Roches et al. 2015, 2016).

Factors Affecting Convergence in White Sands Lizards

Despite strong evidence for convergent evolution of coloration and evidence for similar patterns of change for many other phenotypes in the three White Sands lizard species, patterns are never completely identical across species. We often find that the direction of phenotypic change is concordant but the magnitude of change is different. For example, the lesser earless lizard (Holbrookia maculata) often exhibits the most pronounced differences between White Sands and dark-soil habitats, while Sceloporus cowlesi often exhibits the least (e.g., Rosenblum and Harmon 2011 for dorsal coloration; Des Roches et al. 2015 for diet; Des Roches et al. 2013 for sprint speed).

What explains similarities and differences among species in the magnitude of divergence across the White Sands ecotone? Here we evaluate the contribution of three key determinants that can influence the direction and magnitude of phenotypic response across species: (1) natural selection, (2) genetic architecture of adaptive traits, and (3) demographic context.
Figure 1: Convergent evolution of bleached coloration in White Sands fauna. The top and bottom rows show representative phenotypes from White Sands and dark-soil habitats, respectively. Focal species for this study shown (from left to right): Sceloporus cowlesi, Aspidoscelis inornata, Holbrookia maculata, Habronattus ustulatus, and Ammobates arenicolus.
Natural Selection

The White Sands environment clearly provides a dramatic backdrop for natural selection. Classic observational work on reptile color variation has long posited the adaptive value of substrate matching, especially for small diurnal lizards (e.g., Cott 1940; Norris and Lowe 1964). Experimental studies have also demonstrated the functional importance of substrate matching to avoid predation in a variety of taxa (e.g., Dice 1947; Reed and Janzen 1999). Several experimental studies on lizards (e.g., Luke 1989) and small mammals (e.g., Kaufman 1973) specifically used avian predators that are known to prey on lizards at White Sands, such as the loggerhead shrike and greater roadrunner (E.B.R., personal observation). Our early common-garden rearing experiments also demonstrated that color variation in the White Sands lizards was not explained by ontogenetic or physiological plasticity (Rosenblum 2005). Moreover, there is evidence that all three species experienced ecological release associated with colonizing the gypsum dunes (Des Roches et al. 2011). Specifically, fewer predators and competitors are found in White Sands, relative to dark-soil habitats, and all three focal lizard species correspondingly exhibit higher abundances in the gypsum habitat (i.e., density compensation; Des Roches et al. 2011). Thus, the context for selection at White Sands appears simple: heritable phenotypes, clear optimality criteria, an obvious agent of selection, and decreased predation and interspecific competition.

However, the dynamics of selection might be different for different species, even if they inhabit a common macroenvironment. For example, differences in microhabitat use or behavior can modulate exposure to predation and ultimately the strength of natural selection. To date, we know little about differences among species in the dynamics of selection at White Sands. However, differences among the three lizard species in foraging mode, microhabitat use, and other traits could affect dynamics of natural selection (e.g., Dixon 1967). For example, *Aspidoscelis inornata* is an active forager, while *H. maculata* and *S. cowlesi* are sit-and-wait predators (e.g., Degenhardt et al. 1996). Further, *H. maculata* typically uses open, unvegetated microhabitat, while *S. cowlesi* is more commonly found in vegetated microhabitat (e.g., Hager 2001).

Thus, it is possible, for example, that selection has been stronger for optimal substrate matching in *H. maculata*, which would be consistent with the stronger phenotypic response in this species.

Explicit tests are required to understand whether dynamics of selection differ among species at White Sands. Our initial efforts to quantify selection at White Sands involved conducting large-scale enclosure experiments. We constructed replicated 100-m² enclosures in the natural gypsum habitat (Hardwick et al. 2015). We then painted dozens of White Sands *H. maculata* to match the average dorsal color of the White Sands population and the dark-soil population. We released painted lizards into the enclosures and scored survival of substrate-matched versus that of substrate-unmatched lizards. These experiments helped confirm the activity of avian predation at White Sands but also revealed differences in lizard survival rates across relatively fine spatial and temporal scales and also different survival patterns for males and females. An alternative promising approach is the use of mark-recapture studies to assess differences in strength of selection across species. We conducted a multiyear effort to mark lizards with elastomer tags and assess traits correlated with survival in *S. cowlesi* (Des Roches et al. 2017), and comparison with a similar *H. maculata* data set (S. Des Roches and E. B. Rosenblum, unpublished data) promises to shed further light on differences in dynamics of selection among species. Ultimately, the observational and experimental approaches will complement each other in the effort to understand whether White Sands species experience different selection pressures despite sharing a common environment.

Genetic Architecture

Difference among species in the genetic architecture of functionally relevant traits can influence species’ response to selection. Many aspects of genetic architecture can influence the probability of convergent evolution at the phenotypic level and the probability that similar genetic mechanisms underlie phenotypic convergence (reviewed in Rosenblum et al. 2014).

For White Sands lizards, convergent evolution of blanched coloration is the most obvious and tractable trait for genetic dissection. Our early candidate gene studies revealed mutations associated with blanched color in all three lizard species in the melanocortin-1 receptor (*Mc1r*) gene (Rosenblum et al. 2004). The protein produced by *Mc1r* is a key player in the vertebrate melanin synthesis pathway, and mutations in *Mc1r* are known to be associated with color variation in a variety of other systems (e.g., Barsh 1996; Manceau et al. 2010). Our early work showed a strong statistical association between blanched coloration and a single *Mc1r* mutation in each of the three White Sands lizards. All three mutations lead to amino acid substitutions, and all three substitutions occur in a transmembrane region of the protein, which is important for ligand binding, signal transduction, and structural integrity of the receptor. Our subsequent functional assays confirmed that the *Mc1r* mutations in two of the three species (*S. cowlesi* and *A. inornata*) have important functional effects leading to decreased melanin production (i.e., by reducing receptor integration in *S. cowlesi* and reducing receptor signaling in *A. inornata*; Rosenblum et al. 2010). Although data are less conclusive for the third species (*H. maculata*), we have not ruled out a role for *Mc1r* in color variation in this species.
The molecular signature of selection at and around the Mc1r gene is also striking—and similar—for S. cowlesi and A. inornata. We recently obtained population-level sequence capture data for a ~40-kb window of the chromosomal region around the Mc1r gene (the gene itself is a single exon <1 kb in length). We found no evidence for selection around Mc1r in dark-soil populations of S. cowlesi and A. inornata, but we found strong evidence for selection around Mc1r in White Sands populations of these species (Laurent et al. 2016). Moreover, estimates of the age of the blanched allele at White Sands were remarkably young and similar across species (i.e., 1,200 and 900 years for S. cowlesi and A. inornata, respectively; Laurent et al. 2016). Thus, there are noteworthy mechanistic similarities underlying convergent color evolution in at least two of the lizard species. Mutations at the same gene appear to have swept through the populations during a similar time period, consistent with the age of the White Sands formation itself.

In addition to similarities in genetic architecture of coloration in White Sands lizards, we find several important differences across species. One intriguing difference is the dominance of the blanched Mc1r allele in S. cowlesi but not in A. inornata. Allele frequencies in natural populations, patterns of association between color phenotype and different genotypic classes, and functional studies all suggest that the blanched allele is dominant in S. cowlesi but recessive in A. inornata (Rosenblum et al. 2010). Although both dominant and recessive alleles contribute to adaptation in natural systems, allelic dominance can affect the visibility of adaptive alleles to selection, the likelihood of maladaptive gene swamping via gene flow, and ultimately the probability of fixation (e.g., Orr and Betancourt 2001; Nuismer et al. 2012). The consequences of differences in allelic dominance on the adaptive trajectories of S. cowlesi and A. inornata require further work, as does understanding other genes that contribute to convergent phenotypic evolution in the novel gypsum habitat. Variation at Mc1r cannot explain all observed color variation in the White Sands system, suggesting that other genes are also involved in color variation. Moreover, we are still working to understand the molecular basis of convergent traits other than color. Ultimately, identifying the genes and gene interactions that underlie convergence at White Sands will allow us to understand how the genetic architecture of adaptive traits influences the direction and magnitude of evolutionary change across species.

Population Demography

Differences in underlying population demography can also contribute to different evolutionary outcomes across species. A number of demographic factors—such as population size, time since colonization, and rates of gene flow—have likely influenced species response to natural selection at White Sands. Moreover, differences among species in population demography may help explain differences among species in magnitude of evolutionary change across the ecotone.

Even our earliest molecular data suggested dramatic differences in underlying population demography among the three White Sands lizard species. Early mitochondrial data for White Sands and dark-soil populations demonstrated that patterns of population structure were different across species (Rosenblum 2006). For example, an analysis of molecular variance showed the three lizard species on a spectrum from nearly complete genetic separation between dark and light populations of H. maculata (ΦST = 0.82) to nearly complete panmixis in A. inornata (ΦST = 0.09), with intermediate structure observed for S. cowlesi (ΦST = 0.54). In the past decade, we sampled additional populations and additional regions of the genome. Our first analyses with multilocus nuclear data (~200 single-nucleotide polymorphisms [SNPs] for S. cowlesi and ~50 amplified fragment length polymorphism bands for H. maculata and A. inornata) largely supported the mitochondrial patterns (Rosenblum et al. 2007; Rosenblum and Harmon 2011), again suggesting that blanched coloration could evolve under very different demographic scenarios. Our most recent data sets that rely on broader genomic sampling (i.e., >20,000 SNPs for S. cowlesi and >13,000 SNPs for A. inornata) show that adding data improves our ability to discriminate genetically among populations but that differences in underlying population demography remain among species (Laurent et al. 2016). Specifically, the inferred split time between light and dark populations was younger for A. inornata than for S. cowlesi, and there was stronger evidence for ongoing migration after colonization for S. cowlesi (Laurent et al. 2016).

Our data point to the possibility that gene flow may constrain local adaptation at White Sands. For example, we have the most consistent and conclusive evidence for ongoing gene flow across the White Sands ecotone in S. cowlesi (Rosenblum et al. 2007; Rosenblum and Harmon 2011; Laurent et al. 2016). Sceloporus cowlesi is also the species that tends to exhibit the least phenotypic divergence across the ecotone (e.g., for color, body size, performance, and diet; Rosenblum 2006; Rosenblum and Harmon 2011; Des Roches et al. 2013, 2015, 2016). Ultimately, understanding the relationship between gene flow and local adaptation will be facilitated by continued sampling of dark-soil populations to refine our understanding of the ancestry, time of colonization, and dynamics of ongoing gene flow for the White Sands populations.

New Insights from New Data: White Sands Arthropods

Thus far we have focused on convergent evolution in the White Sands lizard fauna, but several additional—and more distantly related—species also exhibit blanched phenotypes...
in the gypsum habitat. Here we present new data on two other dramatic examples of phenotypic convergence associated with the colonization of White Sands in the terrestrial invertebrate fauna: the sand-treader cricket *Ammobaenetes arenicolus* (Strohecker 1947) and the jumping spider *Habronattus ustulatus* (Griswold 1979). Both species are characterized by pale forms at White Sands and darker forms in the surrounding Chihuahuan Desert, convergent with the pattern observed for the lizards (fig. 1). By increasing the phylogenetic breadth of our work, we can add generality to understanding the factors that influence the direction and magnitude of evolutionary response across a shared ecotone.

We compared sand-treader crickets and jumping spiders from White Sands to those from nearby dark-soil sites. To understand phenotypic response across the ecotone in these species, we quantified coloration of the dorsal body surface, using spectrophotometry and photographic analysis. To understand the demographic history for these populations, we collected genetic data, including thousands of loci obtained with a double-digest restriction site–associated DNA sequencing (RADseq) approach and the mitochondrial *cytochrome c oxidase subunit I* (COI) gene. Sampling and methodological details are presented in appendix A, and detailed results are presented in appendix B (apps. A, B available online). Here we highlight key findings that relate to and extend our previous work on White Sands lizards.

**Arthropod Phenotypic Convergence**

Consistent with patterns we have previously described for White Sands lizards (Rosenblum 2006; Rosenblum and Harmon 2011), we found strong evidence for phenotypic convergence in the two lineages of terrestrial invertebrates. Populations of the sand-treader cricket *A. arenicolus* and the jumping spider *H. ustulatus* at White Sands exhibit significantly lighter dorsal body coloration than populations in the surrounding Chihuahuan Desert (fig. 2). Although both invertebrate species exhibit the same direction of phenotypic change, the magnitude of color divergence between White Sands and dark-soil individuals differs. Specifically, the spiders exhibit less divergence in dorsal coloration than the crickets in comparisons of White Sands to dark-soil populations (fig. 2). The spiders also exhibit sexual dichromatism, while the crickets do not. Specifically, both males and females are lighter in coloration at White Sands for both species, but male spiders are significantly darker than female spiders in both habitats (fig. 2).

Similar to the lizards in this system, natural selection for substrate matching is the most likely explanation for blanched coloration in White Sands invertebrates. Both *A. arenicolus* and *H. ustulatus* spend time exposed on the dunes. Jumping spiders are diurnal and active predators of other terrestrial invertebrates (Foelix 1982; Griswold 1987), and they are most commonly found in low vegetation or on sandy substrate near vegetative cover at White Sands. Sand-treader crickets burrow in the daytime and emerge to forage at night (Weissman 1997), and they are commonly found near their burrows at the base of the sand dunes, typically in areas with sparse vegetation cover (E.B.R. and C.E.P., personal observation). Despite differences in activity period, both species are likely targeted by a range of visually oriented predators. Therefore, dorsal color variation could be linked to selection for crypsis in *A. arenicolus* and *H. ustulatus*, as has been found in numerous other animal species (e.g., Vignieri et al. 2010). Alternatively, dorsal color variation can serve roles in thermoregulation, mate choice, aggression, and immunity (e.g., Horth 2003; Fedorka et al. 2013; Roulin 2016). However, most of these alternative hypotheses have little intuitive support (e.g., because ambient and substrate temperatures are typically lower at White Sands than in the surrounding Chihuahuan Desert [Hager 2000], a thermoregulatory hypothesis would predict the opposite patterns, where darker body color would be favored in the gypsum habitat). Although natural selection for crypsis likely explains color differences across populations, sexual selection may play a role within populations, particularly for the spiders, which—like the White Sands lizards—exhibit some sexual dichromatism.

If dynamics of selection differ between the two invertebrate species, there could be an adaptive explanation for differences in phenotypic patterns. However, additional observational and experimental studies will be required to assess whether differences between the species (or between the sexes) in habitat use, life history, and exposure to predation may have contributed to differences in phenotypic patterns across the ecotone. Experimental studies will also be particularly useful for assessing the potential for phenotypic plasticity to contribute to color variation in *A. arenicolus* and *H. ustulatus* and for identifying genes contributing to color variation in these species. Although both species are relatively difficult to breed in the lab, anecdotally we found that baby spiders born in the lab from White Sands mothers were blanched in color.

**Arthropod Demographic History**

Another possible explanation for differences in phenotypic response across the White Sands ecotone is differences in underlying demography (e.g., population size, colonization history, contemporary patterns of gene flow). Convergence in dorsal coloration is accompanied by conspicuously different demographic histories in *A. arenicolus* and *H. ustulatus*.

The sand-treader cricket *A. arenicolus* exhibits strong population structure, while the jumping spider *H. ustulatus* exhibits little genetic differentiation across the White Sands ecotone. The cricket mitochondrial gene tree shows a well-supported monophyletic White Sands clade (fig. 3), and the
Figure 2: Dorsal color variation for *Ammobaenetus arenicolus* (A) and *Habronattus ustulatus* (B): mean dorsal color (and standard error of the mean) for White Sands (open symbols) and dark-soil (closed symbols) samples. Circles, squares, and triangles represent female, male, and substrate samples, respectively. Note that Y-axes differ because color was quantified differently for the two species (see “Color Quantification” in app. A, available online). PC1 = principal component 1.
RADseq data show strong population structure, with White Sands and dark-soil individuals clearly in separate genetic clusters (fig. 4). In contrast, the spider mitochondrial gene tree shows no clear clustering by habitat (fig. 3). Although there is support for multiple clades in the spider tree, these clades contain both White Sands and dark-soil individuals (with no substructure based on dark-soil collecting locality). Consistent with the mitochondrial data, the spider RADseq data show largely overlapping clusters of White Sands and dark-soil individuals, with little structuring of genetic variation by habitat (fig. 4).

Metrics of population differentiation also show strong differences in genetic patterns between the spider and cricket data sets. For the mitochondrial data set, $F_{ST}$ is more than an order of magnitude higher for the crickets ($F_{ST} = 0.45$) than for the spiders ($F_{ST} = 0.03$; table 1). Moreover, the mitochondrial data suggest dramatically reduced diversity in the White Sands cricket population but not in the spider population. For the crickets, both nucleotide diversity ($\pi$) and Watterson’s $\theta$ are more than an order of magnitude smaller for the White Sands population than for the dark-soil population (table 1), whereas for the spiders, $\pi$ and $\theta$ are high and comparable for the White Sands and dark-soil populations (table 1). Mitochondrial nucleotide diversity is more than two orders of magnitude higher in the White Sands crickets than in the White Sands spiders, and $F_{ST}$ is substantially lower for the mitochondrial data set than for the RADseq data set, but both data sets indicate much stronger population structure for the crickets than for the spiders.

Our molecular results are consistent with different demographic scenarios in the two focal invertebrate species. For the crickets, our results are consistent with a simple colonization history (likely with a reduced population size), limited ongoing gene flow, and strong genetic divergence.
Figure 4: Contrasting patterns of population structure based on restriction site–associated DNA sequencing (RADseq) data for *Ammobaenetes arenicolus* (A) and *Habronattus ustulatus* (B). A, For *A. arenicolus*, the population structure plot (top) for $K = 2$ shows strong population structure. Each individual is represented by a vertical bar and labeled as dark soil (black horizontal bar) or White Sands (white horizontal bar). The genetic principal-component analysis (PCA) plot (bottom left) shows no overlap between dark-soil (filled circles) and White Sands (open circles) individuals. Ellipses represent 95% confidence intervals for the dark-soil and White Sands populations. The joint site frequency spectrum (bottom right) shows only loosely correlated allele frequencies. B, For *H. ustulatus*, the population structure plot (top) shows less structure and more admixture between the two populations for $K = 2$. There is substantial overlap between the dark-soil and White Sands individuals in the genetic PCA plot (bottom left), and the joint site frequency spectrum shows strongly correlated allele frequencies (bottom right).
Convergent Evolution at White Sands

Table 1: Estimates of the nucleotide diversity (π), Watterson’s theta (θ), and F_{ST} for the mitochondrial and nuclear data sets

<table>
<thead>
<tr>
<th>Data set, measure</th>
<th>Ammobaenetes arenicolus</th>
<th>Habronattus ustulatus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dark soil</td>
<td>White Sands</td>
</tr>
<tr>
<td>COI:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>π</td>
<td>.0030</td>
<td>.0002</td>
</tr>
<tr>
<td>θ</td>
<td>.0048</td>
<td>.0003</td>
</tr>
<tr>
<td>F_{ST}</td>
<td>.4541</td>
<td></td>
</tr>
<tr>
<td>RAD sites:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>π</td>
<td>.0213</td>
<td>.0198</td>
</tr>
<tr>
<td>θ</td>
<td>.0316</td>
<td>.0215</td>
</tr>
<tr>
<td>F_{ST}</td>
<td>.0312</td>
<td>.0140</td>
</tr>
</tbody>
</table>

Note: COI = cytochrome c oxidase subunit I gene; RAD = restriction site-associated DNA sequencing.

across habitats. For the spiders, our results suggest admixture across the White Sands ecotone, and different scenarios could underlie the lack of population structure, including multiple colonization events, a large recent founding population with substantial genetic variation, and high levels of ongoing gene flow. Our results are consistent with patterns observed in other groups of jumping spiders. Hybridization, introgression, interspecific gene flow, and incomplete lineage sorting are not uncommon in the Habronattus clade, and gene trees often fail to resolve geographic or phenotypic groups (e.g., Maddison and McMahon 2000; Masta 2000; Hedin and Lowder 2009). To refine our understanding of the relationship between initial colonization history and ongoing gene flow in these species, we will need more thorough sampling throughout the Tularosa Basin (i.e., within species replication) and more explicit modeling of alternative demographic scenarios.

Convergence across the Community at White Sands

Cases of convergent evolution have long been used as evidence that species can exhibit similar evolutionary responses when exposed to comparable selection pressures (e.g., Arendt and Reznick 2008; Losos 2011; McGhee 2011; Wake et al. 2011; Conte et al. 2012). However, myriad factors influence whether species will adapt along parallel trajectories. Studying evolutionary outcomes across diverse lineages provides replicated variation in factors that can influence adaptive convergence (e.g., trait heritability, population size, gene flow, behavior). Further, comparisons across species in a single environment (e.g., where colonization time is geologically constrained and abiotic context is shared) can reduce complexity and analytical noise. Thus, we can better understand whether adaptation is constrained in similar ways across diverse species and which key factors modulate evolutionary response.

Our work thus far on White Sands lizards and arthropods shows both shared and unique patterns across species exposed to a common ecotone. While the direction of color evolution is typically convergent in the White Sands fauna, the magnitude of phenotypic change varies across species. Moreover, differences in phenotypic response appear to correlate with variation in underlying demographic patterns. Specifically, species with more genetic differentiation across the White Sands ecotone (e.g., Holbrookia maculata and Ammobaenetes arenicolus) typically exhibit more phenotypic differentiation. Thus, population demography appears to be an important modulator of phenotypic evolution in this system. Our research also suggests that subtle variation in the dynamics of natural selection and differences in genomic architecture across species can influence patterns of phenotypic convergence.

Moving forward, the White Sands system can be used to address important outstanding questions about the interplay among natural selection, genetic architecture, and population demography. For example, what is the role of phenotypic plasticity in promoting or hindering phenotypic divergence? How do different species experience the same environment? How similar are underlying molecular and functional mechanisms of adaptation in closely versus distantly related species? How do levels of gene flow and strength of selection interact to determine phenotypic outcome? What null expectations are most appropriate to use when studying convergent evolution?

Developing a general understanding of the factors that facilitate and constrain convergent evolution at White Sands will be promoted by studying additional species. Other animal species exhibit blanched forms on the dunes and can be integrated into a broader community-scale study. For example, the moth Euxoa misturata and the pocket mouse Perognathus flavescens also have blanched forms in the gypsum habitat that contrast with nearby dark forms. In addition to these dramatic examples, White Sands is home to multiple endemic species of Lepidoptera that are pale compared to close relatives (Metzler 2014).
Although White Sands is a relatively depauperate ecosystem, the species that have survived on the dunes typically exhibit thriving populations. For example in the herpetofauna, 28 species of reptiles and amphibians are found in the immediate vicinity (Prival and Goode 2005), but only three of these species are found commonly in the heart of the dunes. These three species exhibit dramatic density compensation, with much higher local population densities at White Sands than in nearby desert environments (Des Roches et al. 2011). Relatively high densities allow for a diversity of research approaches to be applied at White Sands. Integrating natural history studies, manipulative lab and field experiments, detailed phenotypic analysis, genomic sequencing, and demographic modeling across the entire White Sands community will ultimately shed light on the factors governing the probability—and the degree—of convergent evolution in natural systems.

Acknowledgments

We thank White Sands National Monument, White Sands Missile Range, and Jornada Long-Term Ecological Research Station for access to field sites and research facilitation over the past 15 years. In particular we are grateful to J. Anderson, D. Burkett, D. Bustos, B. Conrod, and P. Cutler. We thank M. Girard for help in the field and K. Mathews for help with care of spiders in the lab. We also thank D. Elias, M. Girard, D. Lightfoot, E. Metzler, and the Rosenblum lab for productive conversations about White Sands invertebrates. We thank H. Hoeckstra for sharing RADseq adapters, and we thank A. Agrawal, L. Mahler, K. Zamudio, and members of the Rosenblum lab for helpful comments on the manuscript. This work was supported by a National Science Foundation CAREER grant to E.B.R. (DEB-1054062) and a research grant from the Basler Stiftung für Biologische Forschung zu E.T.D. Sequencing was performed by the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, which is supported by National Institutes of Health S10 Instrumentation Grants S10RR029668 and S10RR027303. Animal work was approved by the Animal Care and Use Committee at the University of California, Berkeley (protocol 2014-11-6857), or the University of Idaho (protocol 2009-37).

Literature Cited


Dice, L. R. 1947. Effectiveness of selection by owls of deer mice (Peromyscus maniculatus) which contrast in color with their background. Contributions from the Laboratory of Vertebrate Biology of the University of Michigan 34:1–20.


References Cited Only in the Online Appendixes


Symposium Editor: Anurag A. Agrawal

“The farmer should know his true insect friends as well as his insect foes. We introduce to our readers a large family of ground-beetles (Carabidæ, from Carabus, the name of the typical genus) which prey on those insects largely injurious to crops.” From “Entomological Calendar” (The American Naturalist, 1868, 2:110–111).
Convergence and Divergence in a Long-Term Experiment with Bacteria*

Richard E. Lenski†

Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824

ABSTRACT: Suitably designed experiments offer the possibility of quantifying evolutionary convergence because the fraction of replicate populations that converge is known. Here I review an experiment with *Escherichia coli*, in which 12 populations were founded from the same ancestral strain and have evolved for almost 30 years and more than 65,000 generations under the same conditions. The tension between divergence and convergence has been a major focus of this experiment. I summarize analyses of competitive fitness, correlated responses to different environments, cell morphology, the capacity to use a previously untapped resource, mutation rates, genomic changes, and within-population polymorphisms. These analyses reveal convergence, divergence, and often a complicated mix thereof. Complications include concordance in the direction of evolutionary change with sustained quantitative variation among populations, and the potential for a given trait to exhibit divergence on one timescale and convergence on another. Despite these complications, which also occur in nature, experiments provide a powerful way to study evolutionary convergence based on analyzing replicate lineages that experience the same environment.

Keywords: adaptation, *Escherichia coli*, experimental evolution, fitness, mutation rate, parallel evolution.

Introduction

Against the backdrop of the diversity of life, nature is replete with examples of parallel and convergent phenotypic evolution. Beyond compiling examples of convergence, however, it is difficult to devise a framework for quantifying the extent of convergence or predicting when it is likely, or unlikely, to arise. One complication is the possibility of reporting bias, or what might be called the "denominator problem." For example, camera-like eyes have evolved independently in many lineages, including not only vertebrates and cephalopod mollusks but also some gastropod mollusks, marine annelids, spiders, and cubozoan jellyfish—the last of which strikingly lack a recognizable brain (Conway Morris 2003, pp. 151–158). But in how many lineages did camera-like eyes fail to evolve? Even if we had that number, how should we account for differences in, say, the sizes and ages of lineages when trying to formulate and test hypotheses about ecological or genetic factors that might have predisposed a lineage to evolve camera-like eyes? And in those cases where such features did not evolve, despite seemingly comparable opportunities, did some other difference in circumstances promote divergence instead of convergence?

Although lacking the grandeur of nature, laboratory experiments offer the possibility of quantifying convergence because the denominator—the number of replicate populations that do or do not evolve the same way—is known. Moreover, the sizes and ages of lineages are known in most experiments. Therefore, describing the repeatability of evolution seems simple in principle, although in practice complications may arise. For example, divergent outcomes over the short run could become convergent over the long run, or vice versa, depending on the rates of phenotypic transition and their sensitivity to prior events that promote or impede the transitions. Also, one might well see a mixture of convergent and divergent responses if one examines many different traits.

In recent years, evolution experiments have been used to address a wide range of questions (Elena and Lenski 2003; Garland and Rose 2009; Kawecki et al. 2012; Barrick and Lenski 2013), and the issue of evolutionary repeatability has been a focus of many of these studies. I will not attempt to review the entire field of experimental evolution here, or even those studies focused on repeatability, but I will begin by briefly mentioning three studies that call attention to important issues. First, Burke et al. (2010) found evidence of repeatable genomic evolution across replicate *Drosophila melanogaster* populations that had been selected for accelerated development. The replicates all derived from a single ancestral population that harbored substantial variation. Thus, the observed repeatability of the genomic changes probably stemmed from standing genetic variation—that is, collateral evolution in the sense of Stern (2013)—and not parallel evolution based on independent mutational
events. Second, in a study of the bacteriophage φX174, Wichman et al. (1999) found that almost half of the independent mutations that reached high frequency in two replicate populations were identical, indicating extensive parallelism at the level of the DNA sequence itself. In other studies, however, repeated phenotypic evolution usually results from different mutations in the same gene or different genes that affect the same trait. Third, using a high-resolution random–bar code approach, Levy et al. (2015) monitored thousands of lineages defined by new beneficial mutations within a single population of Saccharomyces cerevisiae. Some of these mutations undoubtedly affected the same gene, and perhaps even the same nucleotide, even though they were independent events, as indicated by different bar codes. Thus, parallel evolution can occur not only across replicate populations but also across multiple lineages within a single population.

Instead of reviewing the entire field, I will focus on one experiment that I have directed for nearly three decades and where the repeatability of evolution—the tension between divergence and convergence—has been at the forefront since it began. In fact, one reason I called this an “evolution experiment,” as opposed to the then-more-familiar “selection experiment,” was to emphasize the fact that any changes that occurred during this experiment encompassed the origin, as well as the fate, of genetic variation—in essence, highlighting the distinction between parallel and collateral evolution that Stern (2013) also emphasized. In addition to presenting what I think are interesting examples of evolution in action, I hope to make clear three facets of convergence and divergence that emerge from this experiment. First, we have observed striking examples of phenotypic and genomic convergence across the replicate populations. Second, we have also seen striking cases of divergence, despite the fact that the populations started from the same genotype and have been evolving in identical physical environments. Third, this mixture of convergence and divergence, when played out over time and across many traits, defies easy summation and instead requires a more nuanced description. As a consequence, assessing the repeatability of evolution is both more interesting and more difficult than I imagined when I began the experiment. And if things become complicated in a set of small, closed, identical flasks, how much more challenging it must be to study the repeatability of evolution in the natural world.

In the next section, I briefly describe the basic structure and methods of the long-term evolution experiment (LTEE) with Escherichia coli. I then summarize some of the phenotypic and genetic analyses that reveal convergence, divergence, and the complicated mix thereof. I conclude by discussing some of the challenges that arise in interpreting the data that result from even such a simple experiment as the LTEE, and by suggesting a few questions that might be addressed by evolution experiments to better understand the fascinating tension between chance and necessity in evolving systems.

Before moving to the experiment, I should briefly discuss a semantic issue that I hope will not cause confusion. In previous presentations on the LTEE, I have usually referred to repeated evolutionary changes as examples of parallelism, rather than convergence. I have done so because the replicate populations started with the same ancestral state, and so the changes are, geometrically speaking, parallel and not convergent. This distinction is clear when the relevant states are discrete (e.g., the identity of a particular nucleotide in a DNA sequence) and the history of changes is known precisely. But for traits that might change repeatedly, and especially for quantitative traits (e.g., size or fitness), we typically measure the net effect of many changes, some of which might be parallel, others divergent, and yet others convergent following prior divergence. In general, our knowledge of the steps along the way becomes less certain as the time elapsed since the identical ancestral states gets longer. Nonetheless, I think it is fair to say that many of us are most interested in the push and pull of evolutionary forces that promote divergence versus convergence. Therefore, in this review, I often refer broadly to divergence and convergence, with the understanding that convergence may, depending on context, include some changes that are, in a strict sense, parallel.

Experimental Overview

The LTEE has 12 populations, each founded from the same ancestral strain of Escherichia coli B except for a genetic marker that distinguishes two sets of six populations each (Lenski et al. 1991). The marker serves two purposes. First, it provides an indication of possible cross contamination, which would compromise the independence of the replicate populations. Second, it allows one to compete a sample from an evolved population against the ancestral strain that possesses the alternative marker, or to compete samples from two reciprocally marked populations against one another. Each population started from a single haploid cell, which was achieved by plating colonies (with each one the outgrowth of a single cell) and then using a different colony to found each population. As a result, all of the genetic variation available for evolution was generated by new mutations, and no variation was identical by descent across replicate populations, thus excluding collateral evolution (Stern 2013) as a potential source of convergence. Each population exists in a small Erlemeyer flask that is held in a shaking incubator at 37°C, where it grows in 10 mL of a medium with glucose as the limiting resource. Every day, someone removes 1% of the volume from each flask and transfers it to 10 mL of fresh medium. The 100-fold
dilution and regrowth allow \( \log_{10} 100 \approx 6.64 \) doublings, or generations, per day, which we round to 100 generations for every 15 transfers. Each day, the bacteria experience a lag phase followed by exponential growth, depletion of the limiting glucose, and a stationary phase where they sit until the next transfer. Even as the bacteria have evolved to grow faster than their progenitors, the number of generations each day is set by the dilution and regrowth. The population size initially fluctuated between \( \sim 5 \times 10^5 \) and \( \sim 5 \times 10^6 \) cells; those numbers have declined somewhat as the bacteria evolved larger individual cells.

The LTEE started in 1988, and it has now been running for over 65,000 generations. There have been occasional interruptions along the way, such as moving the experiment across the country and various accidents. Even more important than their rapid growth, which allows an experiment to encompass so many generations, bacteria can be frozen and then revived. Every 100 generations in the early history of the LTEE, and then every 500 generations since, samples have been stored at \( -80^\circ C \), providing a frozen fossil record. These samples serve two critical functions. First, in the event of an interruption or accident, the LTEE can be restarted without going back to the beginning. Second, one can directly compete or compare samples that lived at different points in time—in essence, time travel. The long duration of the LTEE not only provides an impressive number of generations but also means that new technologies, including the ability to sequence entire genomes, have emerged since it began.

**Convergence and Divergence**

Fitness in the LTEE Environment

All of the populations have become more fit, according to assays in which the evolved bacteria compete against the reciprocally marked ancestral strain in the same environment as the LTEE has run (Lenski and Travisano 1994; Wiser et al. 2013). Before every fitness assay, both competitors have been frozen, revived, and acclimated to that environment. After 50,000 generations, a typical population has a fitness of \( \sim 1.7 \) relative to the ancestor, meaning that the evolved bacteria grow about 70% faster during head-to-head competition (Wiser et al. 2013). Moreover, the fitness trajectories are similar in their overall form, with much greater improvements in the early generations than in later ones (fig. 1).

On the other hand, there is significant among-population variation in fitness—that is, heterogeneity greater than one would expect from measurement error (Lenski et al. 2015). Moreover, from 40,000 to 60,000 generations the relative-fitness ranks of the populations were strongly conserved, even as they all continued to experience significant fitness gains. The among-population variance in fitness shows no sign of declining with time. The square root of that variance—in essence, the genetic standard deviation—has hovered around a few percent, which is substantially less than one-tenth of the average fitness increase relative to the ancestor. Some, but not all, of the persistent variation is associated with evolved differences in mutation rates (Lenski et al. 2015), as described below. Also, one unlucky population, called Ara+1, appears to have gotten stuck in a less productive region of the adaptive landscape than all of the others, having both the lowest fitness and the slowest rate of improvement from 40,000 to 60,000 generations (Lenski et al. 2015).

**Correlated Responses to Other Environments**

One can also run competition assays between the evolved populations and their ancestor in environments that differ from the LTEE environment. In contrast to the high and relatively homogeneous fitness values measured in the LTEE environment, fitness levels tend to be lower and more variable in other environments (Travisano et al. 1995b). For example, when clones sampled at generation 2,000 competed against the ancestor in an environment where maltose replaced glucose, on average they were no more fit than the ancestor. However, the evolved bacteria were extremely heterogeneous, with some almost as fit in maltose as in glucose and others much less fit than the ancestor when competing for maltose. With lactose replacing glucose, the evolved lines were on average as fit as they were in glucose, but the genetic variation in fitness was even greater than it was in maltose. In short, the correlated responses to alternative environments have tended to be more variable than the direct response to selection.
But this tendency toward greater divergence in novel environments has some wrinkles. For example, all 12 populations completely lost their ability to grow on another sugar, ribose, over the first 2,000 generations of the LTEE (Cooper et al. 2001). In this case, the parallel losses resulted from deletion mutations that were demonstrably beneficial in the glucose-based medium and, moreover, occurred spontaneously at an exceptionally high rate. The high mutation rate was caused by a transposable element adjacent to the ribose operon that transposed into the operon and then underwent homologous recombination, leading to a deletion. These losses were so consistent across the replicate populations as to generate convergent, not divergent, outcomes.

There is also a sort of gray area that exists between divergence and convergence. Over time, many of the LTEE populations have lost their ability to grow on maltose; after 20,000 generations, clones from over half of them could not use that sugar (Pelosi et al. 2006). So one could say that the populations underwent divergent responses to maltose early in the LTEE but are now undergoing convergent evolution toward the maltose-negative state. Thus, both the metric we use—quantitative variation in performance versus discrete losses of function—and the timescale over which evolution takes place can affect whether we describe certain patterns as divergent or convergent.

One of the unexpected responses to an alternative environment is that many of the populations have gained resistance to infection by the phage Lambda (Meyer et al. 2010). That resistance is surprising because the bacteria in the LTEE environment are never exposed to Lambda or any other virus. Conventional wisdom about trade-offs is that resistance has a fitness cost, which would lead one to expect that evolution in the absence of parasites should lead to increased susceptibility, not resistance. In fact, the evolved resistance to Lambda has the same genetic basis as the altered performance on maltose. It is no secret that resistance mutations in a gene called malT result in reduced expression of a gene that encodes a porin protein through which maltose crosses the outer membrane of the _Escherichia coli_ cell. That same protein, called LamB, is also the receptor to which Lambda binds to initiate an infection. Mutations in _malT_ provide a demonstrable fitness advantage in the glucose-based LTEE environment (Pelosi et al. 2006), probably because they reduce the costs of unnecessary expression of LamB and other proteins regulated by _malT_, and the resistance to phage Lambda is simply a pleiotropic effect of these cost-saving mutations.

**Cell Size and Yield**

Had I thought in advance about how the size of the bacterial cells would change during the LTEE, I would have predicted they should become smaller. All else being equal, smaller cells have more surface area relative to volume, which seems advantageous in a resource-limited environment. However, that is not what has happened. Instead, all 12 populations produce much larger individual cells than the ancestral strain (Lenski and Travisano 1994; Vasi et al. 1994; Lenski and Mongold 2000). The adaptive significance, if any, of the larger cells remains unclear. One potential explanation is that faster-growing cells tend to be larger, with each cell having more chromosomal copies, ribosomes, and so on. The evolved cells grow faster than the ancestor, so perhaps they are larger simply because they are growing faster. We tested this hypothesis by forcing evolved and ancestral clones to grow at the same rate in separate chemostats, where we could control the growth rate by setting the dilution rate (Mongold and Lenski 1996). We confirmed that faster-growing cells are larger for both the ancestral and the evolved bacteria. However, the evolved bacteria exhibit a steeper increase in cell size with growth rate, such that they produce larger cells than the ancestral strain at all growth rates. Thus, this hypothesis correctly predicts the direction of the change in cell size, but it does not account for the magnitude of the change. Another possibility—one that remains hypothetical—is that larger cells have an advantage in the “feast and famine” transfer regime of the LTEE because they are, in effect, larger sponges and thus able to acquire glucose faster than smaller ones. The larger cells might thereby sequester that limiting resource and pass it (or its by-products) along as they divide. In any case, all of the populations produce much larger cells than their common ancestor, so that they have evolved in the same direction, although their resulting cell volumes are quite variable (Lenski and Travisano 1994).

There has also been substantial divergence in cell shape, with most lineages clearly rod shaped but others somewhat more spherical (Lenski and Mongold 2000). Also, as the cells have evolved to become larger, their numerical yield at the end of the daily growth cycle has decreased (Vasi et al. 1994; Lenski and Mongold 2000). That is, there are fewer, but larger, cells than there were at the outset of the LTEE. Nonetheless, the product of cell number and average cell volume has increased in all 12 populations, which means that the biovolume yield increased in parallel as well. Moreover, there was a near-perfect inverse relationship between the numerical yield and the average cell size of the populations at 10,000 generations (Lenski and Mongold 2000). The constancy of that product indicates that the changes in yield were not only qualitatively but also quantitatively convergent. Once again, however, different analyses of cell morphology—the particular traits measured as well as the direction versus the magnitude of evolved changes—may affect whether one is more impressed by the convergence or by the divergence of the replicate populations.
A New Metabolic Capability

The LTEE culture medium contains glucose as the limiting resource, but there is another large pool of carbon and energy in the medium that the cells cannot consume. In particular, the medium contains citrate as a chelating agent, and its concentration is much higher than that of glucose. However, a near-universal feature of *E. coli* as a species is that it cannot grow on citrate in oxygen-containing environments, because the cells are unable to take up citrate from the environment. So every day, after the bacteria consumed all of the available glucose, another source of carbon and energy sat there untapped. That situation continued for over 30,000 generations, but then one population, called Ara−3, began to use the citrate (Blount et al. 2008), resulting in a culture that was much more turbid than the others (fig. 2A). At first, I thought that a citrate-using contaminant of another species was responsible, and so the population was restarted from its most recent frozen sample. When the citrate users reemerged, we analyzed genetic markers that confirmed that they were indeed *E. coli* cells descended from the strain used to start the LTEE.

At first, we struggled to understand what had changed genetically that enabled the cells to grow on citrate. However, we ran what we call “replay” experiments to distinguish between two hypotheses. According to one, the new Cit^+ function resulted from some rare mutation—for example, an inversion that had to occur at two exact points in the genome—that could have happened at any time in that population’s history and would have yielded the same ability to grow on citrate. Alternatively, the origin of this function was contingent on one or more earlier “potentiating” mutations that, by themselves, conferred no capacity to use citrate but, with the right subsequent mutation, enabled its consumption. Several large replay experiments supported the hypothesis of historical contingency: neither the ancestor nor any clone tested from population Ara−3 before 20,000 generations produced any Cit^+ mutants, but later clones generated many such mutants (Blount et al. 2008).

Soon it became affordable to sequence whole genomes and identify the mutations responsible for the new Cit^+ function (Blount et al. 2012). This analysis revealed that the mutation that produced the first Cit^+ cell in population Ara−3 was a tandem duplication of a segment that includes...

Figure 2: Evolution of a new function in population Ara−3. A, Increase in optical density (i.e., turbidity) when some cells evolved the capacity to use citrate. Redrawn from Blount et al. (2008). B, Tandem duplication leading to two copies of the citT gene, with the second copy coming under the control of one of the promoter elements, shown as arrows, downstream of the ancestral copy. Redrawn from Blount et al. (2012).
the citT gene, which encodes a protein used to transport citrate and some other carboxylic acids under anoxic conditions. However, the effect of the duplication is not a doubling of the basal expression level; in fact, there was no detectable expression of the ancestral copy of citT (Blount et al. 2012). Instead, the duplication brought together the new copy of that gene with a promoter region that was downstream of the original copy (fig. 2B), which led to the expression of the citT-encoded transporter during aerobic growth just as the glucose was depleted. All of the Cit+ mutants found in the replay experiments also had rearrangements of their DNA that appeared to bring together the citT gene and either the same or other promoters, although in none of the replays was the rearrangement identical to the one in the original Ara−3 population. The genomic data also showed that the new module with the combined regulatory domain and structural gene underwent further duplications that improved growth on citrate. By cloning the module into a plasmid that could be moved into other genetic backgrounds, we found strong epistasis between those backgrounds and the module. The new module conferred some capacity to grow on citrate even in the ancestral strain; however, growth was much weaker in that background and several others than it was in the background of a late-generation clone closely related to the lineage in which the new function evolved (Blount et al. 2012). These results thus confirmed the importance of potentiating mutations in setting the stage for the subsequent evolution of the Cit+ phenotype. Later genetic analyses identified a specific mutation in another gene that played an important role in that potentiation (Quandt et al. 2015).

After the bacteria began to use the citrate, it caused important changes to the ecology of the system. Because of the high concentration of citrate in the medium, the population density increased several-fold. Also, the Cit+ bacteria did not drive the Cit− bacteria to extinction; instead, the two lineages coexisted for over 10,000 generations (Blount et al. 2012; Turner et al. 2015a). The Cit+ cells could persist because they had an advantage in competition for the glucose; in particular, they transitioned from stationary phase to renewed growth after transfer into fresh medium faster than did the Cit− cells (Blount et al. 2008). This difference probably occurred because the Cit− cells enter stationary phase after depleting the citrate and then switch over to the more valuable glucose in the fresh medium, whereas the Cit− cells used glucose both before they entered stationary phase and when they recommenced growth. Eventually, however, the Cit− lineage became extinct. The cause of that extinction is unclear, but it appears to have been caused by some transient perturbation to the experiment (Turner et al. 2015a).

In fact, the Ara−3 ecosystem became even more complicated than one might expect in going from one resource (glucose) to two (glucose and citrate) as a consequence of how the new function works. In particular, the citT-encoded protein is an antiporter: in the course of importing one molecule into the cell, it exports another molecule into the environment. Thus, as the citrate is consumed, other, less valuable but still usable sources of carbon and energy—succinate, fumarate, and malate—accumulate in the medium, leading to a system with not one, not two, but several distinct resources. Both the Cit+ and Cit− lineages further evolved in ways that improved their growth on these by-products of citrate transport (Turner et al. 2015b).

As of this writing, at 65,000 generations, no other LTEE population has evolved the ability to grow on the citrate. The evolution of this function is therefore a conspicuous example of divergence under identical conditions from the same initial genetic state. It will be interesting to see whether any of the other 11 populations make this transition in the years, decades, and centuries ahead. If some or all of them do, then what was once seen as divergence might eventually be called convergence. Having replicate populations evolve under identical conditions solves the denominator problem with respect to the number of lineages at risk, but it does not address the open-ended nature of the time at risk for some particular change to occur.

**Mutation Rates**

Six of the 12 LTEE populations evolved point-mutation rates roughly 100-fold higher than that of their ancestor (Sniegowski et al. 1997; Tenaillon et al. 2016). In four cases the hypermutability resulted from mutations that affect DNA mismatch repair, and in two cases it was caused by mutations affecting enzymes that remove oxidized nucleotides. Hence, the set of populations as a whole have diverged in their mutation rates. However, one could also say that six populations converged on higher mutation rates while diverging in terms of the biochemical basis for their hypermutability. In any case, these changes have had important consequences. For example, the hypermutable lineages have accumulated many more mutations, and a higher proportion of synonymous substitutions, than those that retained the low ancestral mutation rate (Tenaillon et al. 2016). Also, the populations that evolved hypermutability early in the LTEE have reached higher fitness levels than the other populations (Wiser et al. 2013; Lenski et al. 2015), although the fitness differences are far smaller than the differences in mutation rate owing to clonal interference (i.e., competition among lineages with different beneficial mutations) and to the increased load of deleterious mutations. Moreover, as a consequence of the increased mutational load and the declining opportunity for further adaptation, five of the hypermutable populations subsequently evolved lower mutation rates (Wielgoss et al. 2013; Tenaillon et al. 2016). Thus, we
see evidence of convergence not only in the mutation-rate phenotype but also in its temporal trajectory.

At this point, one might be wondering whether the lineage in population Ara−3 that gained the ability to grow on citrate had previously evolved hypermutability. In fact, it had not. However, it evolved a mutator phenotype not long after it acquired that function. Moreover, Ara−3 is the only population that became hypermutable but, so far, has not at least partially compensated for that change. As a consequence of its new ability to use citrate, this population has additional opportunities for adaptation, apparently overriding the benefit that would accrue from a reduced load of deleterious mutations.

In addition to the six populations that evolved higher rates of point mutation, another population—Ara+1, the laggard with respect to its fitness trajectory—evolved a different type of hypermutability (Lenski et al. 2015; Tenaillon et al. 2016). In this case, one of the several transposable elements present in the ancestral genome evolved heightened activity that caused a substantial increase in the number of insertion mutations.

Genomic Changes

When the LTEE began in 1988, not a single bacterial genome had been fully sequenced. In one recent paper, 264 genomes sampled from the LTEE were fully sequenced and analyzed, including two representatives from all 12 populations at 11 time points through generation 50,000 (Tenaillon et al. 2016). As discussed above, some of the LTEE populations evolved hypermutability, and those changes obviously affect genomic evolution. If we exclude the populations that became hypermutable (including the one with increased insertions), the genomes sampled from the other populations averaged ∼76 total mutations after 50,000 generations (fig. 3). Of that total, ∼56% were point mutations; various types of insertions and deletions account for most of the rest.

Most mutations that occurred during the LTEE were presumably neutral or deleterious, but those that gave some competitive advantage were more likely to reach high frequency and thus be present in the genomes that were sequenced. Even so, neutral and deleterious mutations can hitchhike with beneficial driver mutations; they may also be present in a sampled genome as the result of a recent mutation or random drift. However, several lines of evidence indicate that most of the mutations observed in the nonmutator populations, and in the others before they became hypermutable, were beneficial (Tenaillon et al. 2016). First, the trajectory for the number of mutations over time was better fit by a model with both neutral and beneficial mutations, which are expected to increase linearly and at a declining rate, respectively, than by models without both terms. Second, nonsynonymous and intergenic point mutations were found in substantial excess relative to the expectation based on synonymous substitutions, after adjustment for the number of sites at risk for each class. Third, insertions and deletions, as well as nonsynonymous and intergenic point mutations, were seen more often than expected in comparison to a mutation-accumulation experiment, in which passage through single-cell bottlenecks largely eliminated the effects of natural selection.

Moreover, there was strong convergent evolution at the level of genes. Fifty-seven protein-coding genes (out of >4,000) had two or more independent nonsynonymous mutations across all of the nonmutator and premutator lineages combined (Tenaillon et al. 2016). These 57 genes comprise only 2% of the total length of protein-coding genes, but they harbored fully 50% of the nonsynonymous mutations in those genomes. That degree of convergence is extremely unlikely ($P < 10^{-143}$), according to simulations in which mutations were randomly distributed among the protein-coding genes according to length (Tenaillon et al. 2016). Some of the genes showing the strongest convergence encode proteins with core regulatory and metabolic functions. Two such genes, *spoT* and *malT*, were first discovered from changes in gene-expression profiles, through transcriptomics and proteomics, respectively, and then tracing back to the causal mutations (Cooper et al. 2003; Pelosi et al. 2006). Another gene, *topA*, was found after changes in DNA superhelicity were observed (Crozat et al. 2010). Several other genes showing convergent evolution encode proteins involved in peptidoglycan synthesis, and mutations there contribute to the evolved changes in cell size and shape (Philippe et al. 2009). There was also a significant, though weaker, signal of convergent evolution in the lineages that evolved hypermutability (Tenaillon et al. 2016). That signal was weaker because the beneficial mutations in those
lineages were diluted in a sea of neutral or weakly deleterious hitchhikers and drifters.

Deletion and insertion mutations also exhibited substantial convergence in the LTEE (Tenaillon et al. 2016). However, it is more difficult to interpret these events, for two reasons. First, many of the parallel indels potentially affect multiple genes. Second, some of them seem to reflect hyper-mutability caused by transposable elements and homologous recombination involving repeated sequences, rather than convergence caused by natural selection. In any case, many of the gene regions most often affected by indels in the LTEE appear to have been acquired by horizontal transfer in the distant past. The disruption or deletion of such genes may be effectively neutral if their products are unused in the LTEE environment, or those mutations might be beneficial if the genes and their products impose some metabolic burden on the cells.

Trajectory of Genetic Parallelism

While the statistical signal for genetic convergence is extremely strong, what can be said about its magnitude and how it changes over time? In a review of published studies, Conte et al. (2012) sought to quantify the probability that the same genes were reused in cases of parallel and convergent evolution of various phenotypes in nature. Depending on the methods used in the underlying studies, they found that the same genes were involved in about one-third to one-half of the cases they examined. However, this probability varied, depending on how long ago the lineages had diverged—there was a greater chance that the same genes were responsible for parallel changes in the most recently diverged lineages.

As noted in the introduction, one of the advantages of an experiment like the LTEE is that we know the denominator (i.e., the number of populations evolving under a given set of conditions) when it comes to analyzing convergence. We also know precisely how long populations have been separated since their common ancestor. Therefore, it is feasible to address the same issues examined by Conte et al. (2012), albeit in a very different system. However, there are empirical challenges: the evolution of hypermutable lineages dilutes and thus obscures the driver mutations, and some mutational events affect multiple genes. Moreover, fitness is an extremely broad and integrative phenotype relative to, say, floral pigmentation or resistance to some toxin. Nonetheless, I sought to address the issue as follows. I focused on nonsynonymous mutations in protein-coding genes, using the data in Tenaillon et al. (2016). I used all 12 populations through 2,000 generations, because none had evolved hypermutability to that point; I used only the six populations that retained the ancestral point-mutation rate through 50,000 generations for the later samples. Using a Web-based tool to screen and sort mutations (http://barricklab.org/shiny/LTEE-Ecoli/), I asked, for each population at each of the 11 time points, how many genes had one or more nonsynonymous substitutions in at least one of the two sequenced clones. That number rose from ~2, on average, at 500 generations to ~32 at 50,000 generations. Then, for every pair of populations, I asked how many genes had nonsynonymous mutations in both of the samples from the same generation. That number increased steadily, from ~0.3 to ~5.8, on average, after 500 and 50,000 generations, respectively. I then calculated the proportional overlap in the affected genes for each population pair at each time point as $2n_{ab}/(n_a + n_b)$, where $n_a$ and $n_b$ are the total number of affected genes for each population and $n_{ab}$ is the number they share. This index of parallelism was ~15% after 500 generations, it rose to ~33% at 2,000 and 5,000 generations, and it then declined gradually to ~18% by 50,000 generations.

Conte et al. (2012) suggest two possible explanations for the declining probability of reusing the same genes at longer divergence times. First, more recently diverged lineages are more likely to share genetic variants that could fuel parallel evolution (i.e., collateral evolution, in the sense of Stern 2013). This issue is not relevant to the LTEE, however, because there was no shared variation at the start. Second, the genetic backgrounds in which new alleles must function diverge over time, introducing the potential for epistatic interactions to alter the spectrum of mutations that produce a particular beneficial effect and thereby generate historical contingencies. This explanation is certainly plausible for the LTEE as well, and indeed it is implicated in the evolution of citrate utilization (Blount et al. 2008).

Two other explanations are also relevant to the LTEE, especially for such an integrative trait as fitness. First, there are probably more genes that can yield mutations that confer small fitness gains than there are genes that can produce mutations that confer large gains. Because of the dynamics of selection, including clonal interference between asexual lineages that possess different beneficial mutations, there is a tendency for beneficial mutations of large effect to fix before those that confer smaller benefits (Gerrish and Lenski 1998; Wiser et al. 2013). As a consequence, adaptation may follow more similar paths—involving those few genes with the largest benefits—early in the LTEE, before fanning out along more diverse paths in later generations. Second, the fraction of accumulated mutations that are beneficial declines over time, with a corresponding increase in the proportion of neutral or weakly deleterious mutations (Tenaillon et al. 2016). While most mutations that affect the same gene in a pair of populations are beneficial, those parallel changes become a smaller fraction of the total mutations over time. This effect might well account for much of the apparent increase in the diversity of genetic pathways over time. The accumulation of neutral or weakly deleterious mutations could also
generate historical contingencies if some of them provide "stepping stones" that enable adaptations that would otherwise not be accessible by a strictly hill-climbing process (Covert et al. 2013).

**Stable Polymorphisms**

Bacteria reproduce asexually, and those in the LTEE lack any means for horizontal gene transfer. Unlike some bacterial species, *E. coli* is not naturally transformable, and the strain used to found the LTEE does not carry any plasmids or functional phages that could move genes between cells by conjugation or transduction, respectively. Bacteria are also haploid, and so polymorphisms maintained by heterozygote advantage are not possible. Moreover, the LTEE environment is simple, thereby limiting—but not eliminating—the opportunity for distinct ecotypes to evolve and stably coexist. In short, one might expect evolution in the LTEE to occur by sequential selective sweeps of beneficial mutations, either singly or as cohorts (Lang et al. 2013; Maddamsetti et al. 2015), such that any polymorphisms are transient, not persistent.

I already mentioned one exception to this expectation, namely, the coexistence between the Cit+ and Cit− lineages in population Ara−3 after the former gained the new ability to use the exogenously supplied citrate. Although the evolution of that ability was surprising, the fact that two lineages could stably coexist on two resources—glucose and citrate—is not. However, another population, Ara−2, produced a polymorphism that arose much earlier and has persisted far longer (Rozen and Lenski 2000; Le Gac et al. 2012; Tenaillon et al. 2016). The polymorphism was first detected through heritable differences in colony morphology, with two morphotypes dubbed "L" and "S" for their large and small colonies, respectively. Experiments in which the two types from the same generation were mixed at different initial frequencies demonstrate a negative frequency-dependent interaction that supports coexistence (Rozen and Lenski 2000; Le Gac et al. 2012). Other experiments have shown that the L type grows faster on glucose, although both types grow substantially faster than the ancestral strain. However, the S type can also grow in spent medium where L had previously grown but can grow no further (Rozen and Lenski 2000). Recent work indicates that acetate, which is excreted by both types during growth on glucose, is used more effectively by and supports the S population (Großkopf et al. 2016).

Although the two ecotypes converge on a stable equilibrium over the course of several weeks, over longer periods their relative abundance has shifted dramatically (Rozen and Lenski 2000), with their ratio varying by almost two orders of magnitude over the first 20,000 generations (fig. 4). These shifts reflect the ongoing evolution of the two lineages—one lineage gets a beneficial mutation that drives the equilibrium ratio in its favor, and then the other does, and so on (Le Gac et al. 2012). However, neither lineage has acquired a beneficial mutation that is sufficient to overcome the advantage when rare that keeps the other lineage from becoming extinct. That is, a sustained polymorphism depends not only on the effect sizes of the beneficial mutations that arise in the contending lineages but also on the strength of the frequency-dependent interaction that gives each ecotype its advantage when rare (Maddamsetti et al. 2015).

In population Ara−1, two genetically distinct lineages coexisted from about 7,000 to 15,000 generations, and competitions between representatives of the two lineages indicate a negative frequency-dependent interaction (Maddamsetti et al. 2015). There were also dramatic back-and-forth fluctuations in the relative abundance of the two lineages, similar to those seen in Ara−2. However, in Ara−1, one lineage eventually accumulated beneficial mutations that delivered a knockout blow, driving the other to extinction (Maddamsetti et al. 2015). Genome sequencing has also uncovered distinct lineages that coexisted in other populations for over 10,000 generations (Blount et al. 2012; Tenaillon et al. 2016). These deeply diverged lineages suggest that negative frequency-dependent interactions, whether transient or sustained, have evolved in at least several of the LTEE populations.

**Overview and Outlook**

I began this review by suggesting that evolution experiments offer the potential to quantify rigorously the extent of convergence and divergence, because the number of lineages
that might or might not change in a particular way is known and because the conditions can be made essentially identical with respect to both the environment and the initial genetic state. Using the *Escherichia coli* LTEE as an example, I have shown how this approach has allowed my team to quantify the among-population variation in fitness, test for parallel mutations in genes, and characterize the patterns of change in many other traits. When I look across the sweep of these analyses—from genetics and genomics through physiology and morphology to performance and ecology—I am fascinated by both the convergences and the divergences of the replicate populations. While we have seen many striking examples of convergence, the more traits we study and the more deeply we examine them, we see that each population is unique and follows its own evolutionary path. For example, all populations evolved to produce much larger cells than the common ancestor, and yet the extent of the size changes and the resulting shapes of the cells show considerable variation.

Another complication arises if we examine the extent of convergence and divergence over time. For example, six of the LTEE populations evolved hypermutator phenotypes, and six did not. But perhaps all of them will eventually become mutators. Or perhaps the mutator lineages will re-evolve lower mutation rates, as some of them have already done. Similarly, only one population has evolved the ability to grow on citrate even after 60,000 generations, but perhaps all of them will have gained this ability after 600,000 or 6,000,000 generations. Or maybe some of them will evolve other new abilities, such as horizontal gene transfer or predation on other cells. In short, we know that the populations started out identical, and we know that they now differ from one another despite having undergone many similar changes, but it is unclear whether the divergences will persist indefinitely or might yet prove to be transient.

My hope is that the LTEE will continue for many scientific generations (Fox and Lenski 2015), providing further insights into the core tension between chance and necessity in evolution. Of course, these *E. coli* populations are different in important respects from the systems that most evolutionary biologists study. Their asexual reproduction leads to clonal interference, which might promote convergence because the relatively few mutations with the largest beneficial effects are even more likely to fix first than they are in recombining populations. On the other hand, asexual reproduction may promote divergence if it allows synergistically beneficial combinations of mutations to fix more readily. Also, analyses of the LTEE populations have often focused on extremely broad and integrative phenotypic traits, including cell size and competitive fitness, where the number of genes with relevant mutations is likely to be much larger than that for specific traits, such as pigmentation and toxin resistance, that are typically studied by evolutionary biologists working with plants and animals. The mere handful of genes that underlie similarly specific traits that have changed during the LTEE, such as DNA superhelicity (Crozat et al. 2010) and hypermutability (Sniegowski et al. 1997; Wielgoss et al. 2013; Tenaillon et al. 2016), might provide more relevant comparisons with typical studies of evolutionary convergence.

In any case, other evolution experiments will be necessary to illuminate important factors that could influence the likelihood of convergent versus divergent responses. Does recombination accelerate adaptive evolution, and does it promote convergence above and beyond the effect of any shared variants present at the start of an experiment (Azevedo et al. 2006; Cooper 2007; McDonald et al. 2016)? Does environmental complexity, such as that produced by spatial structure, multiple resources, or coevolutionary interactions, promote not only diversity within populations but also divergence across replicate lineages (Corona et al. 1994; Rainey and Travisano 1998; Cooper and Lenski 2010; Paterson et al. 2010; Nahum et al. 2015)? How do population size and initial genotypes, whether sampled from nature or from other experiments, affect the propensity for subsequent divergence (Travisano et al. 1995a; Burch and Chao 1999; Moore and Woods 2006; Weinreich et al. 2006; Perfeito et al. 2007; Woods et al. 2011; Lindsey et al. 2013; Kryazhimskiy et al. 2014)? In short, what genetic and environmental factors determine whether fitness landscapes are smooth or rugged, and how do they interact to promote divergent or convergent solutions to a given set of challenges (De Visser and Krug 2014; Wang et al. 2016)? With so many questions to answer, the tension between evolutionary convergence and divergence—between the repeatable and the unique—should be a source of fascination and discovery far into the future.

**Acknowledgments**

I thank everyone who has worked on the LTEE over the course of nearly 30 years, including especially N. Hajela, who has been the lab manager and technician for almost 20 years; D. Schneider, who has worked to understand the genetic changes in the LTEE lines for nearly two decades; and Z. Blount, who has studied the causes and consequences of citrate utilization for over a decade. I thank A. Agrawal for inviting me to write this review; D. Schluter for a suggestion that led to the section on the trajectory of genetic parallelism; J. Barrick for developing the tool that facilitated the analyses in that section; and Z. Blount, O. Tenaillon, and M. Wiser for preparing figures. The LTEE is currently supported by the National Science Foundation (DEB-1451740), the BEACON Center for the Study of Evolution in Action (DBI-0939454), and Michigan State University.
A Long-Term Experiment with Bacteria

Literature Cited


Grade and content: 5/5

"The Bill-fish (Scomberesox Storerii), . . . which but fifteen years since I saw stranded on the shore by the thousands, driven in by its devouring pursuers, has gradually decreased, till at the present time it has nearly, if not quite, been driven away, and I think that during the past year there was not one specimen seen at Provincetown." From "The Habits and Migrations of Some of the Marine Fishes of Massachusetts" by James H. Blake (The American Naturalist, 1870, 4:513–521).
Evolutionary Scenarios and Primate Natural History*

Harry W. Greene†

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853

ABSTRACT: Scenarios summarize evolutionary patterns and processes by interpreting organismal traits and their natural history correlates in a phylogenetic context. They are constructed by (1) describing phenotypes (including physiology and behavior), ideally with attention to formative roles of development, experience, and culture; (2) inferring homologies, homoplasies, ancestral character states, and their transformations with phylogenetic analyses; and (3) integrating those components with ecological and other ancillary data. At their best, evolutionary scenarios are factually dense narratives that entail no known falsehoods; their empirical and methodological shortcomings are transparent, they might be rejected based on new discoveries, and their potential ideological pitfalls are flagged for scrutiny. They are exemplified here by homoplastic foraging with percussive tools by humans, chimpanzees, capuchins, and macaques; homoplastic hunting with spears by humans and chimpanzees; and private experiences (e.g., sense of fairness, grief) among diverse animals, the homologous or homoplastic status of which often remains unexplored. Although scenarios are problematic when used to bolster political agendas, if constructed carefully and regarded skeptically, they can synthesize knowledge, inspire research, engender public understanding of evolution, enrich ethical debates, and provide a deeper historical context for conservation, including nature appreciation.

Keywords: homology, homoplasy, scenarios, tools, weapons, private experiences.

Introduction

This paper stems from the 2016 American Society of Naturalists Vice Presidential Symposium, for which speakers were asked to consider homology and homoplasy in the context of big questions in biology. Taking that charge literally, and inspired by recent discoveries in anthropology and primatology, I first chose: Who are we, and over the long haul, how and why have we come to be this way? These are timeless big questions in biology. Taking that charge literally, and inspired by recent discoveries in anthropology and primatology, I first chose: Who are we, and over the long haul, how and why have we come to be this way? These are timeless big questions in biology.

As cognitive and cultural organisms, who we are is substantially contingent on what we do and with whom we associate, rather than simply what we look like, inside and out. How humans came to be as we are is a matter of memberships in families, societies, and more ancient cultural and genealogical lineages; why we have certain characteristics thus reflects, at least partly, causal relationships among those complex intrinsic and extrinsic influences. All that said, for evolutionary and organismal biologists, understanding the who, how, and whys of other species, including of our closest kin among other primates, involves comparing similar attributes among related taxa—assessing homologies and homoplasies for morphological, physiological, and behavioral traits in the context of ecological correlates (for a pioneering synthesis, see Brooks and McLennan 1991; among many taxon-focused examples: Kappeler and Heymann 1996, for primates; Losos 2009, for lizards, Cavender-Bares 2016, for oaks; for broader perspectives, see Autumn et al. 2002; Losos 2011). The resulting summaries are narrative in structure and thereby vulnerable to abridgement, selectivity, simplification, and other potential distortions (e.g., Landau 1984; O’Hara 1992; Cartmill 2002); consequently, because primate evolution involves our own stories, we should not be surprised that evidentiary and ideological controversies plague them (e.g., Caplan 1978; Seigerstråle 2000; Pinker 2002; Borofsky 2004; Prindle 2009; Marks 2015).

Accordingly, this paper explores the content, legitimacy, and relevance of what Eldredge (1979) termed phylogenetic scenarios, with special reference to primates (for their phylogeny and classification, see fig. 1). My treatment emphasizes comparative analyses for inferring historical changes in morphology, physiology, and behavior (e.g., Greene 1986, p. 3; Autumn et al. 2002; Ereshefsky 2007, pp. 668–670; Losos 2011), rather than denying a place for game-theoretical and other complementary approaches (e.g., Dugatkin and Reeve 2000; Borgerhoff Mulder and Beheim 2011). I first recount the conceptualization and critical early reception of what soon came to be called evolutionary scenarios; second, examine the status of behavioral, ecological, archaeological, and other data in their construction; third, exemplify some of their potential phylogenetic and evidentiary breadth with three case stu-
dies; and finally, delve into their implications for broader concerns.

**Evolutionary Scenarios Defined, Condemned, and Defended**

I use excerpts to reconstruct the origin and early reception of scenario as a term in evolutionary biology because decades-old references are likely obscure to many readers, some subsequent coverage is misleading, and the original authors’ wordings are instructive. As it happened, this innovation occurred amid the following contexts: (1) claims (Atz 1970; Klopfer 1973) and counterclaims (e.g., Greene and Burghardt 1978; Greene 1994, 1999) that behavior cannot be homologized and does not evolve (addressed in the next section); (2) Wilson’s (1975) book *Sociobiology: The New Synthesis*, in which the chapter on humans was only quasi-phylogenetic (see pp. 550–551) and spawned a highly public controversy (e.g., Caplan 1978; Segerstråle 2000; Prindle 2009); (3) claims (Lewontin 1978; Gould and Lewontin 1979; Cracraft 1981) and counterclaims (e.g., Mayr 1983; Wanntorp 1983; Baum and Larson 1991; Larson and Losos 1996; Alcock 1998) that adaptation is poorly substantiated, overemphasized as an evolutionary force, and at best difficult to study from macroevolutionary perspectives (also discussed in the next section); and (4) the rise of phylogenetic systematics (e.g., Eldredge and Cracraft 1980; Hull 1988), which beyond morphology soon encompassed physiology, behavior, and ecology (e.g., Wanntorp 1983; Brooks 1985; Donoghue 1989; Brooks and McLennan 1991)—and was itself a manifestation of the still-ongoing nineteenth-century historicization of biology (de Queiroz 1988; O’Hara 1988; Quinn 2016a).

These four threads were scarcely entwined, at least overtly, in the 1970s and early 1980s. However, Gould’s and Lewontin’s pejorative framing of adaptationism and the adaptationist program occurred in the context of their structuralist perspectives (see below) and disregard for ecology, strongly motivated as well by *Sociobiology’s* chapter on humans and

---

**Figure 1:** Phylogenetic relationships (*left*) and classification (*right*) of primates, based on Wilson and Reeder (2005), Cartmill and Smith (2011), and Fleagle (2013).

[A] phylogenetic scenario [is] . . . a [phylogenetic] tree with an overlay of adaptationist narrative. . . . [A] cladogram must precede the construction of a tree . . . a tree before a scenario. . . . [T]o be debated are (1) the procedures for cladogram, tree, and scenario construction and (2) whether or not the additional assumptions . . . [for the most] complex level of analysis are worth it. (Eldredge 1979, p. 168)

Scenarios are inductive narratives (the best are also seductive) concocted to explain how some particular configuration of events . . . took place. . . . [Their] hallmark is . . . the analysis of the adaptive significance of evolutionary changes in size, form, and structure. . . . [They] are engrossing and concern themselves with the application of what we think we know to the real world, as preserved in the fossil record . . . [but they are] mostly fairy tales . . . a maze of untestable propositions concerning selection, function, niche utilization, and community integration, and alas, do not generally represent good science. (Eldredge 1979, pp. 192–193)

[Scenarios] are easily capable of testing and refutation. . . . [T]he problem here is that we expect scenarios, even more than trees, to be wrong in detail . . . This ad hoc changing of hypothesis content rightly infuriates strict adherents of hypothetico-deductive methodology in science and is the feature of scenarios that most horrifies them. (Eldredge 1979, p. 194)

A year later, Eldredge’s coauthor of punctuated equilibrium theory (Eldredge and Gould 1972), after six pages of praise for a paleontologist’s fictionalized account of Ice Age humans, offered this stinging critique:

Kurtén’s [1980] novel is a more appropriate place than the professional literature itself for discussing many of the truly scientific issues that swirl about the Neandertal-Cro-Magnon debate. Evolutionary biology has been severely hampered by a speculative style of argument that records anatomy and ecology and then tries to construct historical or adaptive explanations for why this bone looks like that or this creature lived here. These speculations have been charitably called “scenarios”; they are often more contempluously, and rightly, labeled “stories” (or “just-so stories” if they rely on the fallacious assumption that everything exists for a purpose). Scientists know that these tales are stories; unfortunately, they are presented in the professional literature where they are taken too seriously and literally. Then they become “facts” and enter the popular literature, often in such socially dubious forms as the ancestral killer ape who absolves us from responsibility for our current nastiness, or as the “innate” male dominance that justifies cultural sexism as the mark of nature. (Gould 1980, pp. xv–xvi)

Lewontin, Gould’s antiadaptationist coauthor, was even more cynical about understanding, let alone portraying, the deep history of behavior, later opining, “It might be interesting to know how cognition (whatever that is) arose and spread and changed, but we cannot know. Tough luck” (Lewontin 1998, p. 130; also see Burghardt 2004). Criticisms of adaptationism, however, were soon countered with conceptual and methodological advances (e.g., Wanntorp et al. 1990; Baum and Larson 1991; Brooks and McLennan 1991; Brooks 1996; Larson and Losos 1996; Alcock 1998; Autumn et al. 2002; see below), while taxon-focused biologists elaborated the notion of scenarios with nonhuman examples, emphasizing explicit phylogenetic analyses, empirical support, and falsification of alternative versions (e.g., Greene 1983, 1992; Wake 1992; Desutter-Grandcolas 1995). The term scenario now is used routinely in diverse contexts (e.g., Sessions et al. 2016), including anthropology (e.g., Dominy 2015).

Nonetheless, almost four decades after Eldredge defined scenarios within the context of phylogenetic systematics, these quotations from a prominent anthropologist’s book, Tales of the Ex-Apes: How We Think about Human Evolution, illustrate contempt that I suspect is common in some circles:

We can study what a feature does, and we can study how it got there, but to ask what is it for is to decorate the scientific question with a lot of metaphysical accessories that it just doesn’t need . . . to assume that there is a reason for it—a deterministic, selective regime for the feature; a particular optimal solution to a problem. But actually, there may be no reasons for some things, just naturalistic causes and uses, and a lot of random noise; life may be more like clothes than like saws. (Marks 2015, p. 62)

[T]hat we can do something does not mean that we evolved to do it, a well-known fallacy known as adaptationism. (Marks 2015, p. 99)

Yet even for such a fundamental feature as bipedality, we know far more about how it evolved than why it evolved . . . what parts of the body changed, and how they changed . . . that’s boring, because it’s anatomy. What we can’t tell you is why it happened. That is his-
tory, that is origin myth, that is interesting. . . . The fact that apes sometimes stand erect when threatening each other . . . that humans can run longer distances than apes . . . that having your eyes higher off the ground allows you to see farther—and many others—have been suggested as scenarios for the evolution of bipedalism. That is to say, it must have been good for something. What these propositions all share is the property of futility. If there was an advantage to bipedalism, we can’t tell what it was, from our vantage point of 5 or 6 million years later. . . . We have to bracket the question “Why did we become bipedal?” and set it off from more empirically based scientific determinism: (Marks 2015, pp. 117–118)

Marks’s concerns are evidently imbedded in a more general disdain for approaches that to him, as with Gould and Lewontin (e.g., Segerstråle 2000; Allmon 2009; Prindle 2009), smack of dangerous biological determinism:

The “pop science” origin myth of human evolution observes the genetic intimacy of humans and apes, applies the cultural assumption that genetic relationships are the most important relationships, and concludes our identity can easily be established from our ancestors. But . . . the fact that your ancestors may have been peasants or slaves does not make you a peasant or slave. We find the cultural idea of reducing identity to ancestry to be morally repugnant. The reason is simple, that we are different from our ancestors, and our identity is established dialectically, recognizing that we are simultaneously both composed of their DNA and yet different from them. (Marks 2015, p. 110)

But, of course, clothes often do have design properties, their advantages contingent on particular environmental and cultural conditions—try exploring the Arctic in a sunga—and that people are not now slaves does not guarantee their ancestors were not, nor that knowing they once were would not prove useful for, say, socially progressive political advocacy. Moreover, although we might never know precisely why bipedalism evolved in our lineage, in terms of particular selective forces, data do constrain the plausibility of alternative evolutionary scenarios for hominins. Ardi (Ardipithecus ramidus) and Lucy (Australopithecus afarensis) are exciting fossils because (assuming the former is a hominin) they confirm that our lineage, after diverging from one leading to bonobos (Pan paniscus) and chimpanzees (Pan troglodytes), evolved bipedal skeletal attributes before we got big brains, not vice versa (White et al. 2009; Wood and Harrison 2011; Fleagle 2013). And as for the alleged futility of scenarios, we do have rich paleoecological and archaeological contexts for interpreting primate evolution, including the inference that our lineage diversified in more open habitats than those ancestrally occupied by hominins (e.g., Harrison 2011; O’Brien et al. 2013; Werdelin and Lewis 2013; Dominguez-Rodrigo 2014).

Even Eldredge, in his original formulation, offered a guarded reprieve and suggestions for progress:

I no longer oppose . . . scenarios (which are, after all, the most fun). . . . As long as we understand precisely what we are doing . . . [and have] an adequate grasp of the probability we are wrong and what assumptions we have added along the way, there no longer seems to me to be any reason for anyone to tell anyone else what not to do. (Eldredge 1979, p. 169)

[We can] improve scenarios: (1) by basing them more explicitly on [phylogenetic] trees and (2) by eliminating some of the more purely speculative elements. . . . Any explicit statement about mode or rate of selection would be idle and fatuous. . . . Adaptation and especially natural selection have no real value in the elaboration of most scenarios and are two examples of excess conceptual baggage. (Eldredge 1979, p. 193)

[Scenarios give] us ideas—lower level hypotheses we might very well be able to test (and which we might not have formulated . . .) —which force us to stretch our imaginations. . . . They are certainly more fun to construct than a mere tree or a dry cladogram . . . as long as we explicitly realize how we build scenarios and what their status is as scientific propositions, we should continue to build them, hoping to find them scientifically, as well as spiritually, uplifting and rewarding. (Eldredge 1979, pp. 194–195)

Gould’s critique likewise ended with a justification for scenarios—but apparently only when they are constructed by evolutionary biologists with opportunities to publish novels:

Yet these stories have a role in science. They probe the range of alternatives; they channel thought into the construction of testable hypotheses; they serve as tentative frameworks for the ordering of observations. But they are stories. So why not treat them as such [by presenting them in fiction], get all the benefits and pleasures, and avoid the wrinkles that arise from their usual, inappropriate placement? (Gould 1980, p. xvi)

I conclude that phylogenetic scenario arose under a cloud of skepticism, owing to concerns about evidence, scientific status, and political misuse—the former two including allegations that we never can know much about ancestral behaviors and ecological contexts, the latter with respect to modern human affairs. Nonetheless, building on earlier discussions (e.g., Greene 1983, 1992; Wake 1992), as well as Eldredge’s
Evolutionary scenarios summarize inferred patterns and processes by overlaying phenotypes and their ecological correlates on phylogenetic trees and then interpreting the results in narrative format. At their best, scenarios are factually dense, and unlike Kurten’s (1980) prehistorical fiction, they entail no unidentifiable speculation, let alone known falsehoods; their methodological and empirical shortcomings are obvious, they might be rejected based on new discoveries, and their ideological pitfalls are flagged for scrutiny. They are constructed by (1) describing phenotypes (including physiology and behavior), with attention, when possible and applicable, to formative roles of developmental genetics, experience, and culture; (2) distinguishing homology from homoplasy in the context of a phylogenetic tree, such that ancestral phenotypic states and subsequent transformations can be inferred; and (3) integrating the forgoing components with ancillary data (e.g., from paleoecology, archaeology) in a manner that is transparent in terms of facts, uncertainties, and ideological implications.

As such, evolutionary scenarios can range from explicitly speculative (e.g., canid-human relationships; Treves and Bonnarc 2016) to extensively corroborated, from a few paragraphs in length (e.g., alcohol consumption by primates; Dominy 2015) to monographic, at which point they may be labeled theories (e.g., play: Burghardt 2005; snakes and primates: Isbell 2009; fire and cooking: Wrangham 2009; motherhood: Hrdy 2011; aggression and hunting: Pickering 2013; weaponry and battle: Emlen 2014). I also reckon that carefully constructed and skeptically regarded, scenarios can play positive roles in anthropology, biology, and society more generally—whether or not they are regarded as scientific, testable, or falsifiable (e.g., Eldredge 1979; Cartmill 2002; Fitzhugh 2016) and despite risks of political pitfalls (e.g., Gould 1980; Klopf 1999; Burkhardt 2005; Marks 2015). As for careful construction and skeptical reception, important concerns include the nature of resemblances, windows on ancient behavior and ecology, and macroevolutionary perspectives on adaptation.

Elaborating on Three Core Issues

**Resemblance, Homology, and Homoplasy**

Homology, granting some important conceptual nuances, is resemblance due to common ancestry (e.g., Hall 1994; McCune and Schimenti 2012). In older literature, it was contrasted with convergence, the latter earlier called analogy, and often defined as independently evolved similarities due to similar functional or ecological roles. The rise of phylogenetic systematics included preference for homoplasy as an overarching term for independently evolved resemblances, whether due to convergence, parallelism, or reversals; analogy has since been widely abandoned as a term in evolutionary biology (e.g., Eldredge and Cracraft 1980; Sanderson and Hufford 1996; Wake et al. 2011; Quinn 2016b; but see Ereshefsky 2012). Moreover, some researchers prefer that homoplasy not be defined with respect to selection, such that from this structuralist (or formalist) standpoint, processes favoring or constraining independent evolution of similar phenotypes are a matter for separate investigations and inference, rather than unsupported assumptions (e.g., Cracraft 1981; Wake 1996; Gould 2002, pp. 1076–1081).

Evaluating resemblances as homology or homoplasy—elucidating phenotypic transformations during lineage diversification—initially entailed parsimonious congruence of character state data in the context of a cladogram (e.g., Eldredge and Cracraft 1980; Hall 1994; Sanderson and Hufford 1996); ancestral-state reconstruction now also relies on likelihood analyses and hypothesis-testing frameworks, statistical aspects of which will continue to evolve (e.g., Griffiths et al. 2015). Regardless of analytic techniques, a core task is delimitation of characters and, ideally, their generative mechanisms, such that when possible we analyze the phylogeny of ontogenies (de Queiroz 1985)—thus encompassing developmental genomics, individual experience, phenotypic plasticity, and, when applicable, culture (e.g., Greene 1994; Ereshefsky 2007; McCune and Schimenti 2012; Foster 2013; Hall 2013; for the ontogeny of prehensile-tail use in two platyrhines, see Beazons 2012; of locomotion and postures in wild chimpanzees, see Sarringhaus et al. 2014). Another core challenge is adequate phylogenetic sampling, such that component lineages are neither under- nor overrepresented; in particular, we should be cautious of incorporating only one member (e.g., chimpanzees) of a multispecies sister group as representative of character states for the ancestor of both lineages (e.g., Bate man 1996; for hominids, Cartmill and Smith 2011, p. 104; Wood and Harrison 2011; Pickering 2014).

An important topic of ongoing exploration, conceptually and analytically, is phylogenetic analyses of culture and its tangible indicators (material culture; e.g., Gray et al 2007; Témkin and Eldredge 2007; Collard et al. 2007, 2008, 2011; O’Brien et al. 2012), as well as placing that research within broader frameworks of gene-culture evolution (e.g., Richerson and Boyd 2005). Of course, individual and cultural experience shape much of primate behavior, complicating comparative assessments—toothbrushes are recent inventions, so, parsimony notwithstanding, observations of geographically disparate and ethnically diverse people cleaning their teeth with similar motor patterns need not imply homologous resemblance to the behavior of ancestral Homo sapiens >200,000 years ago. On the other hand, tooth picking with twigs and flossing with hair appear sporadically among Old and New World nonhuman anthropoids (McGrew and Tutin 1973; Watanabe et al. 2007; Haslam and Falótico 2015), thus ancient proclivities for dental tool use might underlie modern oral hygiene.
practices. And in any case, the emergence of nonhuman archaeology raises prospects that we might someday study transitional material cultures among primate lineages (Haslam et al. 2009; see also below).

**Behavior and Ecology as Phenotypic Attributes with Histories?**

Behavior is what animals do and do not do, visibly and otherwise; besides external manifestations (e.g., postures, movements, vocalizations, immobility), it can include surveillance, cognition, emotions, and other private experiences (Burghardt 1997; see below). As such, behavioral systems reflect neurosensory, neuromuscular, neurohormonal, integrative, and mental processes, themselves influenced—as are morphology and physiology—in complex ways by genetics, development, experience, and, in some species, culture (Burghardt and Bowers 2017). Although early ethologists treated motor patterns comparatively in the same way as anatomy (e.g., Lorenz 1941; Daanje 1950), Atz (1970) asserted that even those overt, often stereotyped behaviors are too difficult to describe, too variable, too subject to selection and convergence, and unsubstantiated by fossils; thus, he argued, behavioral homologies are at best difficult to detect and should not be expected above the level of genera. In fact, none of these criticisms can be generalized, unsupported claims to the contrary notwithstanding (e.g., Klopfer 1999, pp. 124–125), and well-corroborated behavioral homologies, judged by the same sorts of evidence used for morphology and physiology, characterize even some large, ancient clades (e.g., Greene and Burghardt 1978; Wenzel 1992; de Queiroz and Wimberger 2016; Scholes 2008). Important related issues include sources and methodologies of information for ancient and internal behavior, and the status of ecological data in phylogenetic analyses and scenario construction.

Trace fossils and morphological correlates based on living taxa also provide relatively common evidence for the behavior of extinct animals (Seilacher 2007). Classic examples of the former include Jurassic social dinosaur trackways (Ostrom 1972) and the 3.7-million-year-old Laetoli footprints, evidently made by two adult *Australopithecus afarensis* with a child in tow (Leakey 1981; White and Suwa 1987). Anatomical indicators of ancient behavior are exemplified by association of an enlarged gyrus in 10-million-year-old otter brain endocasts with that structure and tactile hand sensitivity in extant lutrines (Radinsky 1968); of bipedal skeletal features among hominins (Fleagle 2013, pp. 364–365); and of innovation, social learning, and tool use with brain size among primates (Reader and Laland 2002). So-called fossilized behavior deserves skeptical appraisal, however, because animals usually do not die during routine activities—granted, the posture of a Cretaceous *Oviraptor* dinosaur over eggs so closely resembles that of incubating modern birds as likely to reflect homology (Norell et al. 1995); on the other hand, a fossil early amniote with a congener in its throat, rather than indicative of cannibalism, probably choked attempting atypical tail-first ingestion (Greene 1994). Other records of ancient behavior come from prehistoric artist-ethologists (e.g., Guthrie 2005; Greene 2013, p. 236), evidence by >30,000-year-old representations of courtship in cave lions (*Panthera atrox*; Yamaguchi et al. 2004), and for hominines, archaeological sources (for an extended example, see Pickering 2013).

With respect to activities that are internal and not observable directly, recall that Tinbergen (1963) codified four questions for ethologists: How does behavior develop, how is it controlled, what is its function or ecological role, and what is its evolutionary history? In posing a fifth question—What are the private experiences of animals?—Burghardt (1997) called for investigations of cognition, emotion, intention, consciousness, awareness, and other internal behavior within Tinbergen’s framework. As such, private experiences might be studied across taxa by combining critical anthropomorphism (Burghardt 1991, 2016) with laboratory (e.g., Brosnan and de Waal 2003) and field (e.g., Holekamp et al. 2007) experiments, as well as with brain imaging, genomics, and other emerging technologies (Preuss 2012). Private experiences thereby can be subjected to phylogenetic analysis, and indeed de Waal (2016a) recently has summarized a career’s worth of primate research in that context. Core problems, once again, include adequate phylogenetic sampling and defining characters (e.g., for insight, see Shettleworth 2012; for psychological homologies, see Ereshefsky 2007)—the latter perhaps entailing trade-offs between, on the one hand, stringent criteria and “killjoy explanations” (Shettleworth 2010), and on the other, more open-minded skepticism and optimistic prospects for detecting trans-species mental homologies (e.g., Cartmill 2000; Rivas and Burghardt 2002).

Ecological correlates of behavior, whether biotic (e.g., predators, competitors) or abiotic (e.g., climate, soil type), are not themselves attributes of organisms, passed from genera-
Adaptationist "Stories" as Macroevolutionary Hypotheses

Even today some might argue, paraphrasing Brown (1982, p. 886), that most evolutionary biologists are interested in natural populations instead of fossils, in variation among individuals, rather than among species and higher taxa. I believe instead that many anthropologists and biologists, as well as members of the public, are fascinated by big questions like those that began this essay and, consequently, will be interested in evolutionary scenarios. A key issue, then, is just what can we infer about adaptations as historical phenomena, as opposed to reflecting natural selection within populations today? Three points bear emphasis.

First, behavioral character states can be characteristic of species and higher taxa, rather than variable within populations (e.g., Greene 1994, 1999; Proctor 1996); they thus can be corroborated as homologous or homoplastic at deeper levels in evolutionary history and used to construct scenarios (e.g., Greene 1986, 1992; Coddington 1988; for social behavior in primates, see Shultz et al. 2011). Moreover, as with other phenotypic attributes, hypotheses of homology and homoplasy are falsifiable if their constituent claims prove inconsistent with new evidence from paleontology, archaeology, and other ancillary sources (e.g., Greene 1986; Larson and Losos 1996; Wood and Harrison 2011).

Second, inferring natural selection from phylogenetic patterns depends on evidence for its general prevalence as a force in evolution and resulting assumptions about past population-level processes. Some evolutionary biologists believe selection must have caused most if not all phylogenetic correlations between phenotypes and ecological circumstances (e.g., Mayr 1983; Coddington 1988; Baum and Larson 1991; Losos 2011), while others are more minimalistic (e.g., Cracraft 1981; Greene 1986; Autumn et al. 2002)—in any case, selection evidently underlies numerous well-studied examples of homoplasy (e.g., Losos 2011; Langin et al. 2015).

Third, irrespective of whether we infer that selection shaped ancient adaptations, widespread correlations of homoplasy with shifts in functional and ecological contexts imply law-like evolutionary patterns (e.g., Brooks 1996; Losos 2011; Armstrong 2017)—the expensive tissue hypothesis, for example, first proposed for hominins, is supported by a correlation between large brains and small guts in amphibians (Liao et al. 2016). Moreover, phylogenetic analyses can reveal whether novel forms, functions, and ecological roles arose simultaneously or sequentially, thus distinguishing between adaptive and exaptive explanations (Gould and Vrba 1982; Greene 1986). Conversely, absence of homoplasy might reflect lineage-wide constraints (e.g., life history and socioecological traits of primates; Kappeler and Heymann 1996). And although Cartmill (2002) argued that prospects are slim for homoplasies revealing much about humans, some astonishing windows on behavioral evolution in primates have opened during the past decade, as illustrated below.

Evolutionary Scenarios and (Mainly) Primate Behavior

Percussive Tools and Extractive Foraging

Bonobos and chimpanzees compose our living sister group within hominine catarrhines, and their use of tools has been increasingly appreciated over the past half century (e.g., Goodall 1964; McGrew 1992). Recent discoveries, however, reveal that although bonobos rarely use stone hammers and anvils for extractive foraging (Roffman et al. 2015), chimps show complexity and cultural variation in that behavior (e.g., Haslam et al. 2009; Lycett et al. 2009). Moreover, percussive tool use also occurs in a cercopithecoid catarrhine, the Burmese island subspecies of long-tailed macaque (Macaca fascicularis aurea), which feeds on nuts and shellfish (e.g., Carpenter 1887; Gumert et al. 2009), and in a platyrrhine, the bearded capuchin (Sapajus libidinosus), which harvests fruits and nuts (fig. 2A; e.g., Frąszczyk et al. 2004; Falótico and Ottoni 2016; Luncz et al. 2016). Similar behavior thus has originated four times among primates, including within hominins; most intriguingly, gracile (Pan paniscus) hominin 3.3 million years ago (Harmand et al. 2015), chimps ~4.3 million years ago (Mercader et al. 2007), Burmese macaques >100 years ago (Haslam et al. 2016).
Figure 2: A. Bearded capuchin with stone hammer, poised to crack a nut. B. Nonhuman primate archaeological site in Brazil, with stone hammer in situ (arrow), used >600 years ago by a bearded capuchin. Photo credits: P. Oxford (A); M. Haslam (B).
(2016a), and bearded capuchins >600 years ago (fig. 2B; Haslam et al. 2016b). A deeper, denser archaeological record might reduce independent origins to three if the common ancestor of Pan and Homo used percussive tools (Hoevers 2015), to two if ancestral catarrhines were implicated, or even—as seems unlikely given the phylogenetic distance among those taxa—to one if percussive tools characterized the ancestor of catarrhines and platyrrhines. Moreover, developmental studies of that behavior might reveal homologous precursors across anthropoids or an even larger clade (e.g., stone handling; Huffman et al. 2010; object manipulation and/or cognitive abilities: Hoevers 2015), such that percussive tool use could have arisen as multiple parallelisms, perhaps each time encouraged by particular ecological conditions (e.g., terrestriality; Meulman et al. 2012).

In any case, when bearded capuchins pound stones together, perhaps for nutritional powder, they produce sharp-edged flakes that are indistinguishable from early hominin tools (Profitt et al. 2016). This discovery suggests that stones inadvertently flaked by nonhuman primates might be erroneously interpreted as tools produced by hominins, as well as that early hominins could have adopted similar incidental flakes as prototools, including projectiles, in a foraging context. Also, because large brains and problem solving are phylogenetically correlated among primates and carnivores (Reader and Laland 2002; Benson-Amram et al. 2016), similar underlying cognitive innovations might be more broadly homoplastic among mammals, including for percussive tool use (e.g., for sea otters; Fujii et al. 2015).

Lances and Javelins

Probing with sticks (e.g., for termites) and aiming projectiles (e.g., stones) are widespread practices among anthropoids, but fashioned weapons (modified to inflict bodily injury) were long believed restricted to our lineage, as documented by a rich archaeological and historical record (e.g., Pickering 2013; Emlen 2014). Spears are long-shafted, penetrating weapons, of which the most ancient known for hominins are ~500,000-year-old hafted, stone-tipped lances (thrusting spears; Wilkins et al. 2012); a 400,000-year-old wooden spear tip (delivery mode unknown; Schoch et al. 2015); 300,000-year-old wooden lances and javelins (light throwing spears; fig. 3A; Thieme 1997; Schoch et al. 2015); and 280,000-year-old hafted, stone-tipped throwing spears (Sahle et al. 2013). The ancient javelins are remarkable for having been made with stone tools from trunks of small spruce (Picea sp.) and pine (Pinus sylvestris) trees, with their shaft and tips carved offset from the softer central axis of the wood and then polished (Schoch et al. 2015). Nonetheless, savanna-woodland chimpanzees in Senegal break off selected saplings, roughly sharpen an end with their teeth, and use those lances to puncture their prey, lesser bushbabies (Galago senegalensis), in tree cavities (fig. 3B; Roach 2008; Pruetz et al. 2015). Also intriguingly, bearded capuchins break off and thin the tips of sticks, with which they probe for honey and small prey animals, as well as jab at dangerous snakes and their mimics (e.g., a false jarraca pit vipers [Xenodon merremi]; fig. 4; Falótico and Ottoni 2014).

Once again, character descriptions and nomenclature are central to phylogenetic analyses and interpretation, and according to Roach (2008) and Pickering (2013), some other researchers were critical of Pruetz and Bertolani’s (2007) calling the savanna-woodland chimps’ modified tools spears and their behavior hunting (perhaps because bushbabies often are killed by females, and prey are small compared to the colobus monkeys [Piliocolobus sp.] taken by forest-dwelling chimps). Like Roach (2008) and Pickering (2013), I regard both words, employed conventionally, as appropriate for the weapons made by early humans and by chimps, whereas the thinned sticks with which bearded capuchins poke (but evidently do not pierce) snakes can be termed protolances. Other plausible antecedents of defensive weapons in hominins, deserving of further study in that context (see Crabb and Elizaga 2008), include unfashioned stick clubs and projectile points used by free-living white-faced capuchins (Cebus capucinus) against a terciopelo pit vipers (Bothrops asper; Boinski 1988) and by semifree tufted capuchins (Sapajus sp.) against a leopard tortoise (Stigmochelys pardalis; Hamilton and Frágasy 2014).

Spear making has arisen repeatedly in primate history, yielding lances, javelins, and hafted projectiles within Homo; lances in savanna-woodland chimps; and, arguably, protolances in bearded capuchins. In all three instances, those weapons increase the distance between primates and prey or otherwise dangerous adversaries, diminishing the risk of fatal retaliation and fostering emotional distance (Pickering 2013). Like human hunters, savanna-woodland chimps appear intentionally composed while trying to skewer their victims, whereas forest chimps are highly agitated while hand-capturing monkeys—a difference used to argue that hunting and aggression were decoupled early in hominin evolution (Pickering 2013). Furthermore, also consistent with the savanna-woodland hypothesis of human ancestry (e.g., Pickering and Domínguez-Rodrigo 2010; Fleagle 2013; Pickering 2014), chimps in that habitat have converged with early Homo by taking shelter in caves, behaving as if familiar with fire, and at least to my eye, being more sparsely haired than forest-dwelling individuals of their species (Pruetz and LaDuke 2010; see images in Roach 2008).

Private Experiences

Internal activities are increasingly subject to phylogenetic perspectives, involving considerations of, among other behaviors, consciousness (e.g., Cartmill 2000), cognition (e.g.,
Figure 3: A. Wooden spear and horse bones from Germany, dated at ~300,000 years ago; note pointed tip (arrow, upper right) and thicker area (arrow, center) typical of javelins. B. Field sample of lances fashioned by chimpanzees for hunting bushbabies in Senegal. Photo credits: P. Pfarr, State Agency for the Heritage of Lower Saxony (A); J. Pruetz (B).
MacLean et al. 2012), emotion (e.g., Clark 2013), and musi-
cality (Honig and Ploeger 2012) in diverse animals; of social
intelligence among primates and carnivores (Reader and Laland
2002; Holekamp et al. 2007; Benson-Amram et al. 2016); of
fairness, empathy, and other putative precursors of morality
among anthropoids (Brosnan and de Waal 2003; de Waal
2016a, 2016b); and of psychological attributes more gener-
ally (Ereshefsky 2007). Although such studies typically ad-
dress private experiences with reference to homology and ho-
mplasy, central to understanding behavioral evolution, that
perspective is not yet true of popular writing on the subject.

Among four recent widely and favorably reviewed books,
only Are We Smart Enough to Know How Smart Animals
Are? (de Waal 2016a) mentions homology; that word and
homoplasy are lacking from the indices and, as far as I can
tell, otherwise unmentioned in Animal Wise: How We Know
Animals Think and Feel (Morrell 2013), How Animals Grieve
(King 2013), and Beyond Words: What Animals Think and
Feel (Safina 2015). Whether thinking, grieving, and feeling
might not be the same behaviors across species as phyloge-
netically disparate as whales, elephants, and primates—and
thus, perhaps, not have shared a common evolutionary der-
ivation—is variously explicit (e.g., King 2013, pp. 7–8) or not
in the other volumes. And although those authors are moti-
vated by conservation and animal welfare concerns (as am I),
they do not grapple in detail with how whether private expe-

Figure 4: Bearded capuchin jabbing with protolance (upper arrow) at plausibly dangerous pit viper mimic (lower arrow) in a rock crevice
(from video). Photo credit: T. Falótico.
xperiences in other species are homologous or homoplastic with our own might affect ethical debates on those topics (see, e.g., Cartmill 2000).

Beyond matters of definition and historical explanation, robust evolutionary scenarios for private experiences will require denser and broader sampling than currently available. In terms of denser representation among relatively basal primates, given logistical and ethical restrictions on captive research, field experiments might more likely yield new insights (e.g., for strepsirrhines: Huebner and Fichtel 2015; for tarsiers: Gursky 2005; for a cercopithecoid: Isbell and Etting 2017). As for phylogenetic breadth, many of the species discussed by King (2013), Morrell (2013), Safina (2015), and de Waal (2016a) are mammals, and with more distant taxa it becomes more difficult to imagine their inner worlds, let alone experimentally address them. However, if we could access the private experiences of eyelid-less, limbless creatures like snakes—"in which we detect no joy and no emotion" (Skutch 1980, p. 257)—that effort might inspire new ways to illuminate more familiar and accessible species (Burghardt 1977, 1991).

**Discussion: Now What?**

Almost four decades after Eldredge (1979) conceptualized evolutionary scenarios, I judge that their shortcomings have been substantially overstated. With respect to the deep history of primates, including of our own species, we have an ever-growing empirical basis for studying the evolution of morphology, physiology, and behavior in the context of ecology and archaeology—but then again, back in the 1970s and 1980s, who would have imagined material culture for nonhuman primates dating back more than four millennia or that savanna-woodland chimps fashion crude spears? Now, as a generalization and granting further refinements of analytic methods, scenario construction seems unproblematic—limited primarily by lack of phylogenies and natural history details for many taxa, and, for better or worse, by restrictions on invasive neurobiological studies of primates (Preuss 2012).

Nevertheless, concerns that evolutionary scenarios can be used to support ideological and political agendas sometimes are justified (e.g., Klopfer 1999; Marks 2015). As narratives, scenarios indeed are vulnerable to distortions and self-serving rhetoric, including by critics; they also are susceptible to interpretive excesses and lapses (for verbatim examples, see Prindle 2009, pp. 135–138), as well as faulty interpretation because of obscure terminology and concepts. Christina Cauterucci’s (2016) blog post asks, for example, “Why do we idolize chimps when we could be imitating female bonobos?”—but even setting aside why we might aspire to a close relative’s social system (e.g., Cartmill 2002), how many lay readers will understand that we are equally related to bonobos and chimps, such that as a generalization and all else equal, neither species is better or worse as a model for thinking about ourselves? (For a critical overview favoring chimps, see Pickering and Domínguez-Rodrigo 2010.) In fact, robustly inferring ancestral states for hominins could require observations of additional taxa, including extinct species, and for some attributes might never be possible (Wood and Harrison 2011). Minimizing political problems surely will follow, at least in part, from transparency in scenario presentation with regard to shaky underpinnings, as well as from admissions of bias by critics and proponents of diverse viewpoints (for extended discussions, see Borofsky 2004; Prindle 2009; Dreger 2011).

As even Eldredge (1979) and Gould (1980) acknowledged, and as exemplified by the above examples, scenarios can play positive roles in anthropology, biology, and society more generally. By highlighting homologies and homoplasies in the diversification of clades, we focus on the controversial extent to which adaptation, constraint, stasis, and repeated evolution of sameness have characterized life’s history (e.g., Brooks 1996; Kappeler and Heymann 1996; Wake 1996; Losos 2011; Wake et al. 2011; Edwards and Donoghue 2013; Bridgham 2016; Agrawal 2017). Moreover, by laying out detailed case studies of Darwin’s descent with modification, scenarios can inspire public understanding of evolution and our long-term roles in the current extinction crisis, as well as inform management strategies and appreciation for biodiversity (e.g., Brooks and McLennan 2010; Greene 2013, pp. 183–187; Werdelin 2013; Cavender-Bares 2016; Sullivan et al. 2017)—we might well regard island macaques, bearded capuchins, and savanna-woodland chimps as especially worthy of conservation, given their homoplastic resemblances to hominins, the possibility of deeper homologies shared among us, and thus the contributions of those other primates to illuminating our own past. Likewise, evolutionary scenarios for a sense of fairness, grief, and other private experiences, by clarifying the nature and origins of those phenomena, could enrich our deliberations about ethical treatment of other species.

With that last point in mind and a nod to de Waal’s (1999) notion of anthropodenial—"blindness to the human-like characteristics of animals, or the animal-like characteristics of ourselves" (p. 258)—I will close with two koan-like tensions. First, some anthropologists and biologists, myself included, strive to confirm private experiences similar to our own in other species, all the while yearning to justify treating them better (e.g., Greene et al. 2002, pp. 198–199; King 2013; Morrell 2013; Safina 2015; de Waal 2016a). Other scholars (and perhaps some of the same ones) insist on human exceptionalism and minimize evolution’s relevance for knowing ourselves (e.g., Lewontin 1998; Klopfer 1999; Marks 2015). It is as if we could be inspired by two metrics, yet ignore their fundamental irreconcilability—one whereby nonhuman animals resemble us ever more closely with each new discovery, the other with which, no matter those revelations, we always
must be profoundly unlike even bonobos and chimps. Second, having spent a career advocating for snakes, I long have regarded their persecution as ignorant, gratuitous, and shameful—especially by the likes of Albert Schweitzer (who famously “revered all life”) and Alexander Skutch (an acclaimed ornithologist), both of whom routinely killed those reptiles (Greene 2013, pp. 125, 181). Now, though, I ponder such negative biases, my own motivations, and prospects for conserving serpents in the light of some 75 million years of mortal conflict between our two lineages (Isbell 2009; Headland and Greene 2011). Well-corroborated scenarios of primate evolution, however uncomfortable to contemplate, will likely play important roles in clarifying these and other contemporary dilemmas.

Note Added in Proof

While this article was in press, I encountered Tattersall and Eldredge’s (1977) earlier formulation of scenario in a phylogenetic context; although not cited by Eldredge (1979), my quotes herein from the latter reference faithfully reflect views advanced in the former.

Acknowledgments

The American Naturalist was the journal I most eagerly read as a graduate student and most aspired to publish in as a sole author, and I am grateful to A. Agrawal for organizing the symposium that finally led to realizing that goal. I also thank G. M. Burghardt, my PhD advisor, for more than 40 years of mentorship; D. B. Wake and M. H. Wake for their friendship, scholarly examples, and support at Berkeley; W. D. Allmon, W. Bemis, M. Bezanson, C. J. Campbell, D. Cundall, K. de Queiroz, F. B. de Waal, H. Drummond, T. Falótico, T. N. Headland, K. R. Hill, D. M. Hillis, K. E. Holekamp, L. A. Isbell, R. E. Keen, B. J. King, J. B. Losos, K. C. MacKinnon, J. Marks, G. Mayer, A. R. McCune, G. B. Pauly, T. R. Pickering, J. Pruettz, A. Quinn, J. P. Sparks, P. Weldon, A. D. Yoder, and K. R. Zamudio for feedback and/or inspiration; E. Behrens, R. Bish, T. Falótico, M. Haslam, G. Legant, N. A. Mason, A. Meyer, R. C. Nali, P. Oxford, J. Pruettz, and T. Terberger for assistance with figures; an anonymous reviewer for cogent criticisms; and the National Science Foundation for support from Opportunities for Promoting Understanding through Synthesis grant 1354156.

Literature Cited


90 in C. A. Ristau, ed. Cognitive ethology: the minds of other animals. Lawrence Erlbaum, Hillsdale, NJ.


de Waal, F. 2016a. Are we smart enough to know how smart animals are? W. W. Norton, New York.


This content downloaded from 132.236.027.111 on July 22, 2017 05:14:50 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).


Isbell, L. A. 2009. The fruit, the tree, and the serpent: why we see so well. Harvard University Press, Cambridge, MA.


Mercador, J., H. Barton, J. Gillespie, J. Harris, S. Kuhn, R. Tyler, and C. Boesch. 2007. 4,300 year-old chimpanzee sites and the origins of percussive stone technology. Proceedings of the National Academy of Sciences of the USA 104:3043–3048.


Symposium Editor: Anurag A. Agrawal

"Then follows a general ‘popular’ account of the forms of apes, their geographical distribution, dwelling-places, food, motions, social life, language, reproduction, education, rearing of young, diseases, life in confinement, and of the apes figured on the Egyptian temples... [Figure:] the head of a five-year-old tschego in the Dresden Zoological Garden, which was brought from the Loango coast." From the review of Brehm's *Animal Life* (The American Naturalist, 1877, 11:557–559).
Convergence, Consilience, and the Evolution of Temperate Deciduous Forests*

Erika J. Edwards,1,‡ David S. Chatlet,1,‡ Bo-Chang Chen,2 Jin Yao Ong,3 Shuichiro Tagane,4 Hironobu Kanemitsu,5 Kazuki Tagawa,5 Kentaro Teramoto,5 Brian Park,6 Kuo-Fang Chung,2,7 Jer-Ming Hu,3 Tetsukazu Yahara,4 and Michael J. Donoghue6

1. Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912; 2. School of Forestry and Resource Conservation, National Taiwan University, Taipei 10617, Taiwan; 3. Institute of Ecology and Evolutionary Biology, National Taiwan University, Taipei 10617, Taiwan; 4. Department of Biology, Kyushu University, 744 Motooka, Fukuoka 819-0395, Japan; 5. Graduate School of Systems Life Sciences, Kyushu University, 744 Motooka, Fukuoka 819-0395, Japan; 6. Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut 06520; 7. Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan

Dryad data: http://dx.doi.org/10.5061/dryad.k505s.

ABSTRACT: The deciduous habit of northern temperate trees and shrubs provides one of the most obvious examples of convergent evolution, but how did it evolve? Hypotheses based on the fossil record posit that deciduousness evolved first in response to drought or darkness and preadapted certain lineages as cold climates spread. An alternative is that evergreens first established in freezing environments and later evolved the deciduous habit. We monitored phenological patterns of 20 species of Viburnum spanning tropical, lucidophyllous (subtropical montane and warm temperate), and cool temperate Asian forests. In lucidophyllous forests, all viburnums were evergreen plants that exhibited coordinated leaf flushes with the onset of the rainy season but varied greatly in the timing of leaf senescence. In contrast, deciduous species exhibited tight coordination of both flushing and senescence, and we found a perfect correlation between the deciduous habit and prolonged annual freezing. In contrast to previous stepwise hypotheses, a consilience of independent lines of evidence supports a lockstep model in which deciduousness evolved in situ, in parallel, and concurrent with a gradual cooling climate. A pervasive selective force combined with the elevated evolutionary accessibility of a particular response may explain the massive convergence of adaptive strategies that characterizes the world’s biomes.

Keywords: Viburnum, biome assembly, freezing tolerance, leaf habit, climate change, phylogeny.

Introduction

Biologists have long been attracted to convergent evolution, and for good reason, as it provides powerful evidence of natural selection on organismal performance. Many convergent traits reflect organismal interactions, both mutualistic and antagonistic (Fenster et al. 2004; Agrawal and Fishbein 2006; Wilson et al. 2007), but past climate change has undoubtedly also driven convergence on a global scale. For example, the latter half of the Cenozoic witnessed the formation and spread of deserts and grasslands, resulting in the repeated evolution of succulent life-forms and C4 photosynthesis (Edwards et al. 2010; Arakaki et al. 2011).

Convergence also provides insights into the relative evolutionary accessibility of certain phenotypes and how structural features of organisms may influence their evolutionary response (Sanderson and Hufford 1996; Donoghue and Ree 2000; Christin et al. 2013). The degree to which a particular character repeatedly emerges must be a function of both the pervasiveness of the selection pressure(s) and the relative ease of its evolution. As climate change is experienced by virtually all organisms in a region concurrently, it makes sense that it would be one of the most powerful agents of massive convergent evolution. However, whether such massive convergence is the outcome depends on whether a given adaptive response is also the most evolutionarily accessible (e.g., Weinreich 2006; Meyer et al. 2012; Agrawal 2017) in multiple lineages. The combination of a ubiquitous selective force with the evolutionary accessibility of a particular adaptation virtually guarantees rampant convergence.

On the surface of it, massive convergence would appear to lend itself perfectly to comparative phylogenetic analyses. It is only through a phylogenetic lens that convergence can be identified, after all, and a multitude of comparative methods have been developed to measure convergence and
the factors that are correlated with it, perhaps causally. In characters with moderate levels of convergence, phylogenetic approaches can work well, but for extremely labile traits, uncertainties in phylogenetic inference can become limiting (Schultz et al. 1996; Schluter et al. 1997). This situation is complicated further by the observation that convergence itself is often distinctly concentrated in particular regions of the tree of life (Edwards and Donoghue 2013). For example, C4 photosynthesis has evolved over 60 times in plants, but fully two-thirds of these origins occur in just two flowering plant lineages, the grasses and the Caryophyllales (Sage et al. 2011), and even within grasses, C4 origins are clustered yet again in a clade that includes roughly half of all grass species (Christin et al. 2013). Furthermore, if evolutionary responses are rapid enough, it becomes impossible to identify the evolutionary sequences and transitional conditions that might provide us with clues about cause and effect, especially about evolutionary events that happened long ago. Faced with these difficulties, convincing answers may not be achieved using phylogenies alone (Christin et al. 2010; Hancock and Edwards 2014). Progress, then, depends on the integration of phylogenetic studies with other lines of evidence (Weber and Agrawal 2012; Olson and Arroyo-Santos 2015) and, as we highlight, perhaps by further deconstruction of the traits and potential selective factors of interest.

Here we focus on the evolution of the deciduous leaf habit, one of the most obvious and highly convergent adaptations of woody plants to temperate environments that experience an annual period of prolonged freezing temperatures. Deciduous plants shed their leaves with the onset of the cold period and remain leafless until they flush a new set of leaves in the spring. This habit dominates northern hardwood forests and has evolved independently many times in distantly related plant groups (e.g., maples and oaks). It is important to note that deciduousness has also evolved many times in settings that we will not consider here, particularly in dry tropical forests where there is also a season unfavorable for growth—in this case, determined by drought, not cold (Murphy and Lugo 1986). We do not know of any attempts to infer the number of independent origins of the deciduous leaf habit in angiosperms, but considering the extremely broad phylogenetic distribution of deciduous plants and its demonstrated lability within smaller lineages (e.g., Schmerler et al. 2012), we guess that it has evolved more frequently than other well-known convergent plant traits (e.g., >62 origins of C4 photosynthesis [Sage et al. 2011]; >130 origins of bilaterally symmetrical flowers [Reyes et al. 2016]).

As a case study, we analyze the evolution of deciduousness in the woody plant clade *Viburnum*, combining standard phylogenetic analyses with a phenological field study along a latitudinal transect in eastern Asia. Independent of the traditional evergreen versus deciduous leaf habit categories, our fieldwork revealed a natural decomposition of phenological behavior into two distinct and potentially independently evolving elements: the flushing of new leaves and the senescing of old ones. It also exposed the inadequacy of the standard binary breakdown of biome or habitat type into tropical versus temperate forests. Specifically, we found it important to also recognize the existence of subtropical montane and warm temperate deciduous forests (Tang 2010; Tang et al. 2013). These Asian forests are characterized by the dominance of broad-leaved angiosperm trees (typically Lauraceae and Fagaceae) in a climate that is distinctly monsoon in comparison to tropical rainforests but where freezing is infrequent and winter temperatures are mild compared to cold temperate forests. The *Viburnum* species that we studied in these forests show patterns of leaf flushing and senescence that provide important clues to the transition between the archetypal evergreen and deciduous conditions. Considered in isolation, our phylogenetic and field studies each provide insight into the evolution of deciduousness but not overwhelming support for a particular evolutionary pathway. Considered together, however, the consilience of evidence leads us to propose a model in which the deciduous habit evolves quite rapidly, in lockstep with a gradual in situ transition to a routinely freezing climate. We hypothesize that the novel behavior that emerged with the deciduous habit was the tight coordination of leaf senescence, perhaps as a means of ensuring the resorption of nutrients prior to leaf death by freezing (Feild et al. 2001; Niinemets and Tamm 2005). To our surprise, this lockstep model appears to be novel, as prior theories have envisioned a distinctly stepwise process. As we review below, these entail either the evolution of the deciduous habit, first in response to a period of drought or darkness (with establishment in freezing climates emerging only later), or the establishment of evergreen plants in freezing climates, followed later by the evolution of deciduousness. We suggest that the stepwise perspective that has explicitly or implicitly oriented many phylogenetic studies of adaptation may be inappropriate or even misleading when studying past evolutionary responses to the gradual climate changes that have shaped the assembly of the world’s biomes.

**Background Information**

**Evolution of the Deciduous Habit in Temperate Forests**

In general, the deciduous leaf habit in woody plants is understood to be a response to long periods of time that are unfavorable for growth, which can result from episodes of freezing or drought (Chabot and Hicks 1982). By definition, annually deciduous plants must have leaf life spans of less than 12 months, with many cold temperate woody plants averaging no more than 6 or 7 months (Reich et al. 1992;
van Ommen Kloek et al. 2012). Deciduous plants also tend to fall on the fast or resource-acquisitive end of the leaf economic spectrum (Wright et al. 2004), and their leaves are typically thinner, with higher nitrogen contents and higher photosynthetic rates than evergreen species. Soil fertility has thus been identified as an important secondary variable: nutrient-poor, cold sites tend to have a higher representation of broad-leaved evergreens, presumably because leaf nitrogen contents are too low to maintain the higher photosynthetic rates required to balance investment costs in a short-lived leaf. These arguments, revolving around trade-offs in leaf carbon and nutrient economy, are well developed elsewhere (Chabot and Hicks 1982; Kikuzawa 1991; Givnish 2002), and our aim here is not to elaborate on the adaptive significance of the deciduous leaf habit. Instead, we simply assume that deciduousness is adaptive for all of the reasons above and consider a related question that has received far less attention, namely, how the deciduous habit evolved from an evergreen state. In particular, we analyze leaf flushing and leaf senescence as two potentially independent behaviors that became coordinated with the emergence of the deciduous habit.

Climate Change and the Origin of Temperate Deciduous Forests

In spite of its importance as a modern biome, there are relatively few hypotheses about the origin of broad-leaved deciduous forests. Inferences based on fossil leaves indicate that deciduous angiosperm species have existed for a long time, appearing as elements of fossil floras from the Late Cretaceous through the Cenozoic (Wolfe 1987). There is evidence for the dominance of the deciduous habit in the highest northern latitudes (polar deciduous forests) during the Late Cretaceous, when global temperatures were significantly warmer than today. These polar forests persisted through the Paleocene and Eocene periods, with broad-leaved evergreens occupying significant areas in mid-latitude and tropical regions (Axelrod 1966; Wolfe and Upchurch 1986; Wolfe 1987). Post-Eocene global cooling shifted many of these midlatitude zones toward the tropics, but it was not until the late Miocene that the modern distribution of cool temperate deciduous forests was established (Tiffney and Manchester 2001; Utescher and Mosbrugger 2007; Pound et al. 2011).

Axelrod (1966) proposed that deciduousness first evolved in a warm, subtropical climate zone that experienced mild winter drought. He hypothesized that such forests were widespread throughout the midlatitudes of the Northern Hemisphere in the Late Cretaceous and early Cenozoic but became restricted during the Miocene. As modern analogues, he specifically cited the lucidophyllous forests of Asia (particularly Taiwan) and the cloud forests of southern Mexico, which experience strong seasonality in both rainfall and temperature but rarely freeze. Axelrod hypothesized that the deciduous leaf habit first appeared in such forests and that this later enabled these plants to expand into the freezing zone.

Wolfe and Upchurch (1986) and Wolfe (1987) explained the origin of deciduous forests in another way. They proposed that deciduousness evolved in situ at high latitudes, when even areas at 80°N did not experience significant freezing. However, winters at such high latitudes were as dark as they are today, and they hypothesized that the deciduous habit evolved there as a response to many months with little or no sunlight. At the Cretaceous–Paleogene (K–Pg) boundary, the brief impact winter was credited with reorganizing plant communities at lower latitudes, and Wolfe and colleagues suggested that it was during this time that polar deciduous lineages migrated south and became established as elements of midlatitude floras. Only during the late Miocene did they outcompete warm-adapted evergreen lineages to establish a new deciduous biome.

It is noteworthy that both the Axelrod and Wolfe hypotheses rest on the idea that deciduousness was a preadaptation to freezing—that is, it evolved first as a response to some other stress but then allowed lineages to tolerate freezing climates when they were eventually exposed to them. Neither author explained why they preferred such an order of events as opposed to the perhaps simpler hypothesis that deciduousness evolved directly in response to freezing. More recently, Zanne et al. (2014), without reference to these older hypotheses, addressed the same question using a megaphylogeny for angiosperms and concluded that most lineages first became established in the freezing zone as evergreen plants and only later evolved the deciduous habit. Elsewhere we reanalyzed the data of Zanne et al. and raised doubts about their conclusions (Edwards et al. 2015), but we highlight this study here because it focused directly on freezing as a driver. Like Axelrod and Wolfe, Zanne et al. envisioned a stepwise evolutionary process but with a reversed order of events. In the terminology of Zanne et al. (2014), Axelrod and Wolfe favored a trait-first hypothesis, whereas they favored a climate-first hypothesis.

The Viburnum Study System

*Viburnum* (Adoxaceae, Dipsacales, Campanulidae) is an angiosperm clade of ∼165 species of shrubs and small trees that is broadly distributed (and widely cultivated) around the Northern Hemisphere. *Viburnum* species all occupy mesic forests, but they have adapted to a range of climatic conditions and forest environments (fig. 1). At one end of the spectrum, multiple (and distantly related) species in tropical forests are never subjected to freezing temperatures. At the other end, upwards of 10 *Viburnum* lineages appear to have
evolved independently into cool temperate forests that experience consecutive months of below freezing temperatures. In Asia, members of several clades occupy lucidophyllous forests, and members of the Neotropical Oreinotinus clade occupy montane cloud forests, with reduced temperature seasonality and only rare and intermittent freezing temperatures.

Evolutionary shifts between these different environments have been accompanied by transitions in leaf habit (fig. 1). In tropical, lucidophyllous, and cloud forests, Viburnum plants are most often evergreen, while, with very few exceptions, in cool temperate forests they are seasonally deciduous. Of the 165 species of Viburnum that we currently recognize, we consider 84 to be evergreen and 81 to be deciduous. Eight of the well-supported major clades within Viburnum include both evergreen and deciduous species, implying a minimum of eight evolutionary shifts. However, we infer no fewer than 20 transitions in leaf habit using any of our recently published phylogenetic trees (Clement et al. 2014; Spriggs et al. 2015; Eaton et al. 2017), including some clear-cut evolutionary shifts in both directions. For example, the deciduous species Viburnum plicatum appears to

**Figure 1**: Phylogeny, climatic envelope, forest habitat, and leaf habit of Viburnum. A 120-taxon phylogeny, pruned from Spriggs et al. (2015). Branch colors represent inferred ancestral minimum weekly temperatures (BIO6). Pie charts at internal nodes are ancestral state estimates of leaf habit, inferred using the threshold model (Revell 2012). Species at tips are coded by leaf habit in the left column (evergreen vs. deciduous) and habitat in the right column (tropical, lucidophyllous, cold temperate, or cloud forests). Species with names in boldface were included in the phenological monitoring study.
have been derived from evergreen ancestors within the Lutescentia clade. Likewise, the evergreen species Viburnum rhytidophyllum and Viburnum utile appear to have originated from deciduous ancestors within the Euviburnum clade, as did Viburnum sempervirens and its several evergreen relatives within Succotinus.

As we have shown previously (Schmerler et al. 2012), these repeated evolutionary shifts in forest environment and leafing behavior have been accompanied by specific changes in leaf form. Adaptation to cool temperate forests at more northerly latitudes has entailed coordinated shifts to rounder leaf shapes with more marginal teeth or lobes. Marked seasonal heteroblasty in the temperate species (the development of preformed leaves with more derived leaf shapes and neofomed leaves with more tropical shapes) suggested to us that the evolution of temperate leaf forms may relate to the development of leaf primordia within the resting buds of deciduous species (Edwards et al. 2016). In any case, Viburnum provides clear evidence, directly in line with global trends (Bailey and Sinnott 1916; Wolfe 1995; Royer and Wilf 2006; Peppe et al. 2011), of repeated evolutionary changes in both leaf habit and leaf form in connection with shifts between tropical and temperate forests.

Material and Methods

Phenology. To monitor Viburnum leafing behavior across a wide range of environments, we established a latitudinal transect spanning 27° latitude, from montane tropical forests in Kinabalu National Park, Sabah, Malaysia, at 6°N to temperate forests in the mountains surrounding Fukushima, Japan, on Kyushu Island at 33°N (fig. 2). Sites were chosen to represent different forest types (tropical rain forest in Borneo, both lucidophyllous and cool temperate forests in Taiwan and Japan) and also to maximize the number of co-occurring Viburnum species in each location. From May to June 2013, we selected a total of 19 populations representing 18 species in Taiwan and Japan, and in February 2014 we added two species in Borneo. For logistical reasons, not all populations were monitored for the same length of time. The Borneo populations were monitored for 13 consecutive months, the Japanese populations for 10–19 consecutive months, and most of the Taiwanese populations for 25 consecutive months. We achieved our best sampling in Taiwan, where we were also able to monitor two populations of Viburnum luzonicum that spanned its elevational range (120 m, 2,100 m).

In each population of each of our focal species, we tagged four branches per individual and made biweekly to monthly census trips to record the persistence of each leaf on these branches, the emergence of new leaves, and all events of leaf senescence using methods described in Edwards et al. (2014). At the start of the study, each branch usually had only a few apical meristems, but due to the growth architecture of most Viburnum species (stems typically terminate in an inflorescence, with growth recommencing from two subtending buds; Dornhue 1981; Edwards et al. 2014), by the end of the study each branch consisted of multiple independent leafy shoots, sometimes bearing over 100 leaves.

Viburnum-wide climate data, leaf habit, and forest type. We downloaded georeferenced locality data for Viburnum from the following databases: Global Biodiversity Information Facility (http://www.gbif.org/), Chinese Virtual Herbarium (http://www.cvh.org.cn/), Interactive Agricultural Ecological Atlas of Russia and Neighboring Countries (http://www.agroatlas.ru/), Plant DNA Bank in Korea (http://pdbk.korea.ac.kr/), and Virtual Viburnum (http://viburnum.peabody.yale.edu/). We also georeferenced occurrence records from Hará’s (1983) treatment of the Viburnum species of Japan. The raw data were filtered to remove spurious records. Specifically, we matched taxon names with those in Spriggs et al. (2015) and removed records with coordinates at 0°, 0° (and those from botanical gardens and herbaria) and records of the same species with identical coordinates. We also plotted the records for each species using the R package ggmap (Kahle and Wickham 2013) and removed points outside of the known geographic ranges of these species based on floristic treatments and herbarium specimens examined in preparation for our worldwide monograph of Viburnum.

We used the CliMond database (Kriticos et al. 2012) at a 10-min resolution (~18 km² at the equator) to extract estimates of climate for each locality in the filtered data set using the R package raster (Hijmans and van Etten 2012). This spatial grain best matches the georeferencing precision of our data set. The species locality data were then spatially rarefied so that there was no more than one occurrence per climate grid cell using SDMtoolbox (Brown 2014), which increased the evenness of coverage across the species’ entire range. Species with fewer than three grid cell occurrences were removed from the data set. We calculated the mean minimum temperature of the coldest week (BIO6) and seasonality of precipitation (BIO15) for each species. We note that although CliMond estimates BIO6 as a weekly measure, their variable is interpolated from monthly temperature data and is strongly correlated with the more commonly used BIO6 variable, the mean minimum temperature of the coldest month ($R^2 = 0.995$). In all, our data set consists of 7,718 records for 120 Viburnum species. We also used the CliMond database to extract mean monthly temperature and precipitation data for the sites of our monitored populations.

We scored each species in our data set as (1) belonging in the traditional evergreen or deciduous category and
(2) occurring in tropical, lucidophyllous, cool temperate, or cloud forests. These scores were based on monographic and floristic treatments of *Viburnum* (e.g., Rehder 1908; Killip and Smith 1931; Morton 1933; Kern 1951; Hara 1983; Yang and Malecot 2011). For the majority of these species we have first-hand knowledge based on our own field and herbarium studies, but for Japanese species we also relied on Hara (1983) and for Chinese species we followed Yang and Malecot (2011). We were uncertain about assignment to forest type for seven species and scored these as unknown.

**Analysis of phenological data.** We recorded leaf gain and loss using hand-drawn sketches of branches, and dates of emergence and disappearance of each leaf were transcribed into a spreadsheet. The total number of leaves gained or lost at each census was summed across branches within an individual and averaged across individuals within a population. Only a small subset of leaves were observed for their entire life span in the evergreen species in our sample; the majority of leaves were either present when we initiated the study and senesced partway through the monitoring period or emerged during the study and were still present at the end. Population-level patterns in flushing and senescence were visualized using the ribbon function in ggplots2 (Wickham 2016) in R. Most species appeared to exhibit strong seasonality in leaf flushing and sometimes also in senescence. To evaluate the statistical significance of these patterns, for each species we compared our observed data with a null distribution of phenological events. Because the timing of our population visits was not perfectly spaced throughout the year, we could not assume a flat null distri-
bution; just by chance we would expect to see more events in spring and fall seasons, when our visitation rates were slightly higher. For each species, we pooled all visitation dates (as Julian days) and generated 1,000 data sets each of new leaf and senesced leaf observations, sampled with replacement from the pool of Julian day visits, with the total number of events equal to that in the observed data. We then tested whether observed distributions were statistically different from the null distribution using a Mann-Whitney test. All data and scripts for analyses are publicly available in a GitHub repository (https://github.com/ejedwards/amnat_2017/phenology/null_analyses).

Analysis of Viburnum-wide climate data, leaf habit, and forest type. We conducted comparative analyses using the maximum-clade credibility phylogeny from Spriggs et al. (2015). We pruned the phylogeny to the 120 species (of the 165 species of Viburnum) represented in our data set using the drop.tip function in ape (Paradis et al. 2004). These species represent all of the named clades in the comprehensive phylogenetic classification of Viburnum developed by Clement et al. (2014). We conducted several analyses to test the relationship between leaf habit and climate. First, we estimated liability values to inferred ancestral estimates of the relationship between these traits using phylogenetic independent contrasts (Felsenstein 1985) as implemented for use with a discrete and continuous character in the function crunch in the R package caper (Orme 2013), treating leaf habit as the dependent variable and either BIO6 or BIO15 as the independent, continuous variable. Second, we inferred evolutionary shifts in leaf habit using the threshold model first proposed by Wright (1934) and developed as an explicitly phylogenetic method by Felsenstein (2012). The threshold model assumes that discrete character changes are governed by shifts in an underlying, unobserved continuous variable called the liability. As the value of the liability reaches a particular threshold, the binary character changes. This is an appealing approach to examine possible evolutionary relationships between observable continuous and discrete characters by evaluating the relationship between the continuous character and the estimated liability parameter. We estimated the liability values underlying shifts in leaf habit using the function threshBAYES in phyttools (Revell 2012), running Markov chain Monte Carlo for 10^6 generations, sampling every 1,000, and discarding the first 20% of steps as burn-in. We regressed estimated liability values to inferred ancestral estimates of BIO6 and BIO15. Third, we utilized standard methods for assessing the correlated evolution of two binary characters with likelihood ratio tests (Pagel 1994). In this case, we converted BIO6 into a binary character, binning species into whether their mean value of BIO6 was greater or less than 0°C. Because six species with mean values just above or just below 0°C were assigned to categories incongruent with our field knowledge, we created an additional binary variable where we adjusted these assignments. We ran Pagel tests using both the curated and the uncurated binary variable. All data and scripts for analyses are available in a GitHub repository (https://github.com/ejedwards/amnat_2017/analyses/phylogeny). All data used in all analyses are available in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.k505s (Edwards et al. 2017).

Results

Phenology. The two tropical Viburnum species that we followed in Borneo (Viburnum clemensiae, Viburnum vernicosum) are evergreen and exhibited little leaf turnover during the course of our study (fig. 3). This was especially true for leaf senescence, with some individuals losing no more than a single leaf over the course of 13 months. Viburnum clemensiae produced leaves from October to April and no leaves at all during the months of May–September. In contrast, V. vernicosum produced leaves throughout the year despite the fact that these two species occupy the same montane tropical forest, with individuals less than 1 km apart. The very low rates of leaf turnover in both species necessitate a much longer period of data collection to make robust observations of seasonality.

In both Taiwan and Japan, we included evergreen and deciduous species in our sample, though the ratios were inverted: in Taiwan we monitored three deciduous species and six evergreen, while in Japan we monitored six deciduous and three evergreen. In each location, one of the evergreen species is perhaps better described as a leaf exchanger; Viburnum luzonicum (Taiwan) and Viburnum erosum (Japan) plants shed almost all of their leaves as they produced a new flush.

Across all species and all sites in Taiwan and Japan, there was a striking coordination of the timing of new leaf flushes; all species, whether deciduous or evergreen, produced a significant flush of new leaves in the months of March, April, and May (fig. 3; table A1). In sharp contrast, senescence patterns were more variable across species and across the seasons. In deciduous species, as expected, leaf senescence was highly seasonal and varied little between species or locations; all deciduous species demonstrated coordinated leaf shedding as winter approached, typically in November and December. In evergreen species, senescence periods were typically much longer, occurring in the months preceding, during, and after spring leaf flushes (fig. 3; e.g., Viburnum propinquum, Viburnum taitoense). The tightest coordination of flushing and senescence appeared in V. luzonicum and V. erosum, but as noted, these do not fit neatly into the strictly evergreen or deciduous category. Indeed, they may have a short leafless period in one year but not the next (e.g., V. luzonicum was evergreen.
Figure 3: Flushing and senescence patterns of 20 *Viburnum* species living along a tropical-temperate gradient. Lines represent the mean number of leaves gained or lost per observational period across three to six individuals per species; shading represents ±1 SE. Green lines indicate new leaf flushes, and yellow lines represent leaf senescence. Species are grouped by geographical area (Borneo, Taiwan, Japan) and leaf habit (evergreen, deciduous). Asterisks indicate species that may be better described as leaf exchangers rather than strictly evergreen.
in 2014 but deciduous in 2015; V. erosum was deciduous in 2014 but evergreen in 2015.

The significance of these visual patterns was confirmed by comparison to our null distributions of flushing and senescence (table A1). All deciduous species exhibited highly significant clustering of both flushing and senescence. Flushing patterns were all significant in our evergreen species; in contrast, in most cases senescence was indistinguishable from our null distributions (table A1; see additional files and analyses at https://github.com/ejedwards/amnat_2017/final_analyses/phenology/null_analyses). The three exceptions were Viburnum japonicum in Japan and V. propinquum and Viburnum integrifolium in Taiwan. These species all exhibit a similar senescence pattern, with a period of senescence directly preceding and then directly following the spring leaf flush (fig. 3). We note that both V. japonicum and V. integrifolium belong to the Succotinus clade along with V. erosum and V. luzonicum and likely evolved an evergreen habit from a deciduous ancestral state.

Correlation of phenology with climate. We chose sites in both Taiwan and Japan that spanned freezing and nonfreezing climates, reflecting differences in elevation. In both countries, there is a pronounced monsoonal climate, with the heaviest rains falling during the warm summer months. As precipitation is indistinguishable across the freezing/nonfreezing boundary, we were able to compare phenology in contrasting temperature regimes while holding rainfall patterns constant (fig. 4).

As noted above, leaf flushes were coordinated across all sites and species in both freezing and nonfreezing zones. These flushes occurred in the spring as temperatures warmed and monsoonal rains returned. Due to the correlation between temperature and rainfall, it is not possible at this time to discern which climatic factor the flush is primarily responding to. Ancillary evidence that rainfall may be the primary driver is provided by our two populations of V. luzonicum (fig. 5). Our high-elevation site showed a strong monsoonal rainfall pattern and, correspondingly, a well-defined spring flush; our low-elevation site experienced pronounced temperature seasonality but with high rainfall year-round, and its leaf flushing patterns were considerably more erratic (fig. 5).

In contrast to flushing, we found evidence for a single environmental driver of coordinated senescence and deciduousness: in both Taiwan and Japan, our deciduous species all lived in areas experiencing routinely freezing temperatures during the coldest months of January and February (fig. 4). In contrast, all of our evergreen species lived in areas where minimum monthly mean temperatures remained above freezing year-round. It is important to note that this does not mean that they never experienced occasional frosts, only that the mean monthly minimum temperatures never fell below 0°C. Fortunately, our sampling allows us to separate the influence of temperature from the influence of drought, as our evergreen and deciduous species in these areas all experienced similar periods of relative drought from November to February (fig. 4).

The clear association between the deciduous leaf habit and regular exposure to freezing temperatures but not changes in precipitation seasonality was corroborated in our Viburnum-wide analyses. Figure 1 maps the mean minimum temperature of the coldest week (BIO6) on a Viburnum phylogeny derived from Spriggs et al. (2015), together with ancestral reconstruction of leaf habit inferred from the threshold analyses. This implies that the first viburnums lived in warm climates without significant freezing and that there were multiple independent transitions into colder (and then back again into warmer) environments. Scanning the distribution of the leaf habit and forest type characters gives the impression of a strong correlation: species living in warmer climates tend to be evergreen and to occupy tropical or lucidophyllous forests, while species living in colder climates are generally deciduous and occupy cool temperate forests. This is confirmed by our independent contrast analyses, which showed a significant relationship between leaf habit and BIO6 ($R^2 = 0.607, P = .00001$) but not BIO15 ($R^2 = 0.037, P = .3785$). Likewise, our thresholdYuES analysis yielded a strong positive relationship between leaf habit and BIO6 ($R^2 = 0.868, P = 2.2e^{-16}$) but not BIO15 ($R^2 = 1.6e^{-5}, P = .9647$). Similarly, our binary tests strongly favored a model of correlated evolution of leaf habit and habit (uncurated BIO6 binary: $\Delta AIC = 58.56, P = 1.22e^{-13}$; curatured BIO6 binary: $\Delta AIC = 73.2, P = 9.63e^{-15}$).

Discussion

Flushing, senescence, and emergence of the deciduous habit. The deciduous habit so characteristic of Northern Hemisphere temperate forests involves two distinct behaviors: in the spring, there is a rapid and dramatic flush of new leaves, and in the fall, there is an equally dramatic, coordinated, and often colorful display of leaf senescence. By monitoring the flushing and senescing of leaves as independent behaviors in both evergreen and deciduous Viburnum species, we can begin to understand how these two aspects of phenology relate to different climatic factors and how their coordination may have evolved along an evergreen-to-deciduous evolutionary trajectory.

The preliminary data on Viburnum phenology presented here clearly document several patterns. Unfortunately, the very low leaf turnover in the essentially aseasonal tropical site (Borneo) limits interpretation of that data set for now. We note only that our single year of data collection provides some evidence of seasonality to new leaf production (fig. 3) and that periods of leaf flushing are not uncommon in tropical
Figure 4: Seasonal drought, freezing, and the coordination of flushing and senescence. Box plots represent mean monthly precipitation (in mm), and line plots represent mean monthly maximum temperature (red) and minimum temperature (blue) in °C. Density curves below the climate plots were derived from pooling all new (green) and senesced (orange) observation events from all species within that category (e.g., evergreen Japan includes Viburnum japonicum, Viburnum awabuki, and Viburnum erosum). Climate data are from one representative field site per category. In cases where we had multiple locations of evergreen or deciduous populations within a country, their climate diagrams were virtually indistinguishable.
Figure 5: Variation in phenological behavior in *Viburnum luzonicum*. Lines on left-hand graphs represent the mean number of leaves gained or lost per observational period across four to six individuals per species; shading represents ±1 SE. Green lines represent new leaf flushes, and yellow lines represent leaf senescence. Climate diagrams on the right are similar to figure 4: A, *V. luzonicum* population in the Danfeng region just outside of Taipei, Taiwan, at an elevation of 120 m. B, *V. luzonicum* population in the Hehuanshan area of the Central Mountain of Taiwan at an elevation of 2,124 m.
plants more generally, perhaps in response to herbivore pressure (Coley and Barone 1996). We focus instead on our observations in Taiwan and Japan, where all of our sites are mesic but with strong rainfall seasonality. Here it is clear that a pronounced and coordinated leaf flush is present in all species. Senescence patterns, on the other hand, are quite varied, and leaf drop appears to be less concentrated in time. Although our evergreen species varied considerably, most exhibited patterns that were indistinguishable from a random null distribution (table A1). The senescence period could be quite extended in these species, and the intensity (i.e., the absolute number of leaves gained or lost per event) of senescence was generally reduced compared with leaf flushing. Our results are remarkably similar to an earlier study of phenology in lucidophyllous Japanese forests, where most species produced a single spring flush of leaves but presented at least four different senescence patterns (Nitta and Ohssawa 1997). In marked contrast, our deciduous species all had short and coordinated senescence periods that commonly matched their spring flush in intensity. Senescence also occurred concurrently, in the late fall, in all of these species. Owing to the high level of homoplasys, the directionality of character change in Viburnum is still difficult to infer in some areas, but our sample of nine deciduous species likely captures a minimum of five independent origins of the deciduous habit (i.e., within the Euviburnum, Pseudotinus, Urceolata, Solenotinus, and Succotinus clades; Clement et al. 2014; Spriggs et al. 2015). Therefore, this provides a clear case of the convergent evolution of coordinated leaf senescence.

Several species stand out as notable exceptions. Two, Viburnum luzonicum and Viburnum erosum, are what we describe as leaf exchangers, as their senescence events directly preceded or were concurrent with their flushes. The range of *V. luzonicum* extends to warm temperate forests in southern China and south to the Philippines. In contrast, we sampled *V. erosum* in the warmest part of its range in Japan; it extends northward into colder forests, where it is certainly deciduous (Hara 1983). Two other exceptions are Viburnum integrifolium and Viburnum japonicum, two evergreen species that exhibit senescence patterns that appear to be slightly concentrated in two bouts, one preceding and one following their annual leaf flush. Interestingly, all of these species are nested within the large eastern Asian Succotinus clade and have close relatives in cool temperate forests (Clement et al. 2014; Spriggs et al. 2015). We infer that these species have extended their ranges back into warmer climates and have shifted (at least in the southern parts of their geographic ranges) in the direction of an evergreen habit. Without a phylogenetic context, one might consider a leaf exchanging phenotype to be an intermediate phenotype in the transition from an evergreen to a deciduous habit. However, in the case of *Viburnum* it seems more likely to be an intermediate phenotype in the transition from a deciduous to an evergreen habit. This makes sense given that a coordinated senescence period had already evolved in deciduous species; in the absence of freezing, senescence may have become delayed until it effectively co-occurred with new leaf production. A single peak of senescence might have been disrupted in shifting to a more fully evergreen condition, as evidenced by *V. integrifolium* and *V. japonicum*.

A close comparison of the climates occupied by our evergreen versus deciduous *Viburnum* species highlights the climatic factor that most likely drove the initial evolution of coordinated senescence. All of our sites in Taiwan and Japan experience an annual dry season beginning in October–November, but in the evergreen species this is not associated with coordinated leaf loss; instead, coordinated senescence appears only when plants are subjected to prolonged freezing. All of our deciduous species and only our deciduous species occur in locations where minimum monthly temperatures routinely dip below freezing in winter (January–February; fig. 4). This perfect correlation in our 20-species data set was corroborated by our *Viburnum*-wide phylogenetic analyses, which strongly supported minimum temperature of the coldest week (BIO6) as a significant correlate of shifts in leaf habit across the clade and strongly rejected any influence of precipitation seasonality (BIO15). Taken together, our data indicate that periods of leaf flushing are universal in *Viburnum*, perhaps even in the tropical forest species (where more data are clearly needed). A more concentrated burst of leaf flushing may have evolved first in lucidophyllous forests that experience monsoonal rainfall patterns with a period of drought in the cool season. Specifically, we suggest that the initiation of flushing might be triggered by the onset of spring rains (generally in April–May), though given the obvious correlation, it is difficult to disentangle the relative effects of rainfall versus temperature or even rainfall versus lengthening photoperiod at that time of year. In the case of leaf senescence, on the other hand, the environmental trigger seems quite clear. Our data suggest that a regular and prolonged period of freezing temperatures was the primary driver of highly coordinated leaf senescence and consequently the emergence of a fully deciduous leaf habit.

How do our results square with expectations from the Wolfe, Axelrod, and Zanne et al. hypotheses outlined above? Although we cannot rule out Wolfe’s hypothesis that darkness at high latitudes favored the evolution of deciduousness in the Late Cretaceous in some plant lineages (Wolfe and Upchurch 1986; Wolfe 1987), we doubt that this explanation applies to *Viburnum*. Our phylogenetic analyses, the *Viburnum* fossil record, and the absolute dates that we have inferred (Spriggs et al. 2015) are not consistent with Cretaceous origins of deciduousness in *Viburnum*. Instead, most origins probably occurred in the late Miocene, when temperate vegetation was becoming more common at lower latitudes (Utescher and Mosbrugger 2007; Found et al. 2011).
Axelrod’s hypothesis that deciduousness evolved in warm but monsoonal forests with seasonal drought (Axelrod 1966) does not apply neatly to *Viburnum* either, considering both the phylogenetic and phenological analyses. Breaking deciduousness into two components—leaf flushing and leaf senescence—allows us to consider the possibility that these may have had separate evolutionary causes. Temporally restricted and coordinated leaf flushing might have evolved in response to the strong rainfall seasonality of monsoonal climates, with the onset of leafing possibly controlled by the onset of rains or increasing temperatures in the spring. Restricted and coordinated leaf senescence, on the other hand, might have been driven by exposure to freezing temperatures. Drought is potentially also a factor in senescence, as Axelrod supposed, but as we have shown, our evergreen species also experience similar periods of significant drought in the fall without dropping their leaves. Freezing, not drought, seems to be most directly related to coordinated senescence in our deciduous species.

We concur with a simple physiological explanation for a direct link between freezing and deciduousness: a coordinated period of senescence allows plants to remobilize key nutrients prior to the inevitable death of leaves by predictable freezing (Feild et al. 2001; Keskitalo et al. 2005; Niinemets and Tamm 2005). In all of the varied senescence patterns observed in our evergreen species, it is worth noting that no peaks in senescence coincide with the onset of the dry season; instead, they are scattered throughout the rainy season. It appears that viburnums can weather dry periods by either tolerating lower water potentials or maintaining low stomatal conductance. A modest drought tolerance combined with interannual variation in the strength of drought periods may render drought a weaker selection pressure for coordinated senescence, as many leaves may be able to persist and continue to function during the following rainy season. It is important to note, of course, that there are evergreen woody plants in the northern temperate zone (e.g., *Rhododendron*, *Ilex*) and that these species instead produce freezing-tolerant leaves with longer life spans. With very few exceptions (fig. 1), *Viburnum* has not evolved this alternative strategy, suggesting that deciduousness was the more evolutionarily accessible adaptation in this lineage. Investigating the very few instances of evergreen viburnums that experience freezing (e.g., *Viburnum rhytidophyllum*, which is nested within a cold temperate deciduous clade) might provide insights into the relative advantages of these two distinct ecological strategies to the cold temperate zone.

In proposing that deciduousness evolved directly as an adaptation to predictable and prolonged freezing, we might appear to be more aligned with the hypothesis of Zanne et al. (2014). These authors concluded that many lineages first became established in the freezing zone as evergreen plants and that deciduousness evolved later. Our data, however, do not support this order of events. One expectation of such a stepwise model is the existence of one or more ancestrally evergreen clades living in freezing climates, with one or more deciduous species nested within it. However, there is no unambiguous pattern of this type in *Viburnum* phylogeny. Furthermore, in the species that we monitored, evergreens never occur in forests that experience predictable freezing. On these grounds we judge the Zanne et al. (2014) explanation to be unlikely in this case, although the hypothesis could be salvaged by imagining the systematic extinction of evergreen species in cold climates.

An alternative model of evolution. As we have emphasized, the previously proposed hypotheses envision a sequence of steps ending in the association of cold climates with a deciduous habit. A slightly different view seems much simpler in the face of the evidence that we have presented for *Viburnum*. We might consider the possibility that the deciduous habit and in particular the coordinated senescence of leaves in the fall evolved in some plant lineages directly in response to freezing. By “directly” we mean that the senescence component of the deciduous habit evolved in situ as populations were subjected to a gradually cooling climate and the establishment of a prolonged and predictable period of freezing during the winter. We envision that this change in climate took place in fits and starts over the course of thousands to millions of years (Zachos 2001; Herbert et al. 2016), and this is a timescale during which the final steps in the evolution of a fully deciduous habit could certainly have evolved.

Under this hypothesis the deciduous habit is not a preadaptation that first originated as the solution to some other problem, nor did it evolve in lineages that got there first and adapted later. Instead, we are suggesting that climate change and evolutionary adaptation occurred together, effectively in lockstep with one another. This scenario is especially realistic for an adaptation related to leafing phenology, which we have shown to be quite sensitive to climate even within a single species across its range (fig. 5).

The distinction we are making between stepwise and lockstep may seem like a subtle one, but we think it is crucially important in considering biome assembly and the causes of concerted convergence. At the very least it warrants greater care in our use of phrases like “moved into” or “shifted into” in relation to biome transitions. Likewise, the Zanne et al. (2014) terms “climate first” and “trait first” quite explicitly describe a stepwise process—in other words, that one thing happened first and the other thing happened later. In the case of the emergence of the temperate deciduous biome, it appears that a widespread and floristically rich warm evergreen forest (the boreotropical flora sensu Wolfe 1975; Tiffney 1985a, 1985b) was subjected to late Miocene cooling, eventually resulting in a climate with annual periods of freezing temperatures. The many existing lineages of woody plants in these forests would have experienced this in situ di-
rectional trend. Some of them probably fared poorly, either going extinct or shifting their ranges accordingly, but many others simply kept up with the changing times and stayed put.

What factors determine the winners and losers as large geographical regions gradually become reconfigured and new biomes emerge? In our case, the deciduous habit might have been more accessible to species that already produced resting buds and that flushed their leaves annually in response to variation in precipitation or temperature. The key step in such lineages would be the consolidation of leaf senescence, and it is this, we argue, that likely was a direct response to freezing, not drought. Indeed, such predispositions likely explain the repeated evolution of the deciduous habit within particular lineages and in general the clustering of the origin of certain traits in particular regions of the tree of life (Edwards and Donoghue 2013; Donoghue and Edwards 2014). *Viburnum* provides an excellent example, with possibly as many as 10 separate origins of deciduousness, but we suspect that other familiar temperate lineages that were present in boreotropical forests (e.g., *Acer*, *Carya*, *Nyssa*, *Hamamelis*, *Sassafras*, *Cornsus*; Tiffeny 1985a), when analyzed carefully, will show similar patterns of in situ convergence.

It is certainly the case that biome assembly also reflects other processes, including preadaptation, habitat filtering (cf. Ackerly 2004), and even occasional long-distance dispersal events (Pennington and Dick 2004; Crisp et al. 2009). We also do not doubt that all of these pathways have contributed to the assembly of temperate deciduous forests. However, we believe that these other phenomena may have had a relatively minor influence as compared to the in situ adaptation of multiple lineages to gradual climate change. We suggest that such concurrent responses may largely underpin the assembly of new biomes and that this process may be responsible for the extremely high number of origins of certain traits.

Although we have concentrated here on direct shifts between mesic tropical and temperate forests, we fully appreciate that there have likely been other biome pathways and in particular that some members of cool temperate forests may have been derived within lineages that occupied seasonally dry tropical forests (e.g., possibly *Celtis*). Such a pathway would be far more consistent with a stepwise model, where deciduousness evolved first in response to one variable (drought) and served later as a preadaptation for another (cold). Nevertheless, in both pathways (i.e., with ancestors in either mesic forests or dry forests) the final steps in the process may have been similar. In both cases, senescence may have been relatively uncoordinated at first, as in the case of dry forests, where leaves tend to senesce at different times depending largely on individual plant water status (e.g., Reich and Borchert 1984). However, freezing events are experienced by all individuals simultaneously, and routine exposure to prolonged cold could have tightly coordinated the senescence period to a more condensed time period across the community.

Consilience and integration in phylogenetic biology. How do we justify our lockstep model? If we approached this with only our phylogenetic tools, we would not be entirely satisfied. The high lability of both the deciduous habit and climate makes it difficult to infer with any confidence many of the deeper and potentially most relevant transitions (fig. 1). We do know that deciduousness and freezing climate are very tightly correlated across our tree, so much so that changes in both of them very often co-occur along the same branches, making it difficult to establish cause and effect (Baum and Donoghue 2001). If there had in fact been a repeated order of events, history has largely erased its signal. We might also be frustrated by our inability to analyze both characters as continuous variables, which would have provided a more direct test of a simultaneous, lockstep evolutionary model. Transforming habitat into a continuous climate variable was possible by using the CliMon database, but assigning continuous values to species along an evergreen-deciduous continuum is logistically and conceptually more difficult and would require detailed phenological studies of every species, not just the 20 that we have presented here.

But what if we consider these conclusions together with the results from the phenological study? Here we also see evidence of a very tight association between freezing in particular and the deciduous habit (fig. 4) and of the extreme lability of phenological patterns, even among populations within a single species (fig. 5). These observations are completely consistent with the strong correlation across the *Viburnum* phylogeny, and together these lines of evidence suggest to us that the “simultaneous” changes that we see in phylogenetic analyses may often be real simultaneous changes rather than artifacts of extinction. If individual species can be evergreen in one part of their range and deciduous in another (as in *V. erosum* and *V. luzonicum*), gradual and concurrent emergence of the deciduous habit as a freezing climate becomes established begins to seem highly likely.

The study of phylogenetic comparative methods appears to be shifting into a more introspective phase, with renewed scrutiny of the limitations of our inferences (Edwards et al. 2015; Maddison and FitzJohn 2015; Pennell et al. 2015; Wright et al. 2015). Critiques typically end with a call for better models, and we agree completely that there is much room for improvement. However, at the end of the day, we need to remind ourselves of the extreme difficulty of convincingly inferring evolutionary events in the distant past with data usually only from extant organisms, especially when convergence is rampant and densely clustered in particular
clades. Under these circumstances, phylogenetic insights will be most powerful when interpreted alongside other, quite independent lines of evidence until (hopefully) consilience emerges (Weber and Agrawal 2012; Olson and Arroyo-Santos 2015).

From this perspective, we believe that research conducted at the scale of *Viburnum* will continue to be highly productive, precisely because the results of multiple independent analyses can most effectively be brought into alignment in such model lineages. For example, the data presented here are relevant to questions about the deciduous leaf habit, but they are also directly connected to our studies of leaf form and function (Chatelet et al. 2013; Scoffoni et al. 2016), growth architecture and leaf life span (Edwards et al. 2014), and even flowering phenology (L. Garrison, M. Donoghue, and E. Edwards, unpublished manuscript) and wood anatomy (D. Chatelet, M. Donoghue, and E. Edwards, unpublished manuscript). Consequently, we have the potential to establish connections between what might at first appear to be unrelated phenomena. It is this form of synthesis that we find most inspiring and most likely to provide new conceptual breakthroughs as we face the formidable challenge of inferring evolutionary history. Immersion in the messy details of particular lineages, far from being a distraction, may be the most fertile ground for achieving consilience—the moment when multiple, sometimes weakly supported findings all point to the same conclusion.

**Acknowledgments**

We thank A. Agrawal for his invitation to participate in the 2016 American Society of Naturalists Vice Presidential Symposium and for his patience in handling our manuscript. We appreciated insightful comments by D. Ackerly, T. Pennington, and two other reviewers, which greatly improved our manuscript. We are also grateful to J. Pereira and the staff of Kinabalu National Park for their help with phenological monitoring, E. Johnson for help with data cleansing and management, and J. de Vos, J. Gauthier, M. Olson, and members of the Edwards and Donoghue labs for many thoughtful discussions. This work was supported by National Science Foundation grants IOS-1257262 to E.J.E. and IOS-1256706 to M.J.D.

**APPENDIX**

**Supplementary Table**

**Table A1:** Significance of leaf flushing and senescence patterns in 20 species of *Viburnum* along a latitudinal gradient in Asia

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th>Leaf habit</th>
<th>Event</th>
<th>Total leaves</th>
<th>Mean Julian day (observed)</th>
<th>Mean Julian day (null)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>V. awabuki</em></td>
<td>Japan</td>
<td>Evergreen</td>
<td>Leaf flush</td>
<td>41</td>
<td>89</td>
<td>194</td>
<td>2.60E-12</td>
</tr>
<tr>
<td><em>V. awabuki</em></td>
<td>Japan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>48</td>
<td>176</td>
<td>195</td>
<td>.239</td>
</tr>
<tr>
<td><em>V. betulifolium</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf flush</td>
<td>211</td>
<td>118</td>
<td>187</td>
<td>3.20E-25</td>
</tr>
<tr>
<td><em>V. betulifolium</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>195</td>
<td>336</td>
<td>187</td>
<td>1.90E-88</td>
</tr>
<tr>
<td><em>V. hitchiana</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf flush</td>
<td>353</td>
<td>102</td>
<td>207</td>
<td>4.00E-102</td>
</tr>
<tr>
<td><em>V. hitchiana</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>523</td>
<td>305</td>
<td>207</td>
<td>2.60E-115</td>
</tr>
<tr>
<td><em>V. clemensiae</em></td>
<td>Borneo</td>
<td>Evergreen</td>
<td>Leaf flush</td>
<td>68</td>
<td>110</td>
<td>169</td>
<td>4.20E-06</td>
</tr>
<tr>
<td><em>V. clemensiae</em></td>
<td>Borneo</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>15</td>
<td>183</td>
<td>170</td>
<td>4.80E-01</td>
</tr>
<tr>
<td><em>V. dilatatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf flush</td>
<td>353</td>
<td>106</td>
<td>189</td>
<td>6.57E-39</td>
</tr>
<tr>
<td><em>V. dilatatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>352</td>
<td>311</td>
<td>189</td>
<td>1.39E-96</td>
</tr>
<tr>
<td><em>V. erosum</em></td>
<td>Japan</td>
<td>Leaf exchanger</td>
<td>Leaf flush</td>
<td>240</td>
<td>91</td>
<td>195</td>
<td>3.54E-65</td>
</tr>
<tr>
<td><em>V. erosum</em></td>
<td>Japan</td>
<td>Leaf exchanger</td>
<td>Leaf drop</td>
<td>205</td>
<td>276</td>
<td>195</td>
<td>2.67E-34</td>
</tr>
<tr>
<td><em>V. furcatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf flush</td>
<td>229</td>
<td>134</td>
<td>205</td>
<td>1.41E-32</td>
</tr>
<tr>
<td><em>V. furcatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>341</td>
<td>311</td>
<td>205</td>
<td>6.47E-106</td>
</tr>
<tr>
<td><em>V. integrifolium</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf flush</td>
<td>154</td>
<td>126</td>
<td>187</td>
<td>4.65E-12</td>
</tr>
<tr>
<td><em>V. integrifolium</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>109</td>
<td>170</td>
<td>188</td>
<td>2.00E-02</td>
</tr>
<tr>
<td><em>V. japonicum</em></td>
<td>Japan</td>
<td>Evergreen</td>
<td>Leaf flush</td>
<td>134</td>
<td>81</td>
<td>195</td>
<td>3.01E-47</td>
</tr>
<tr>
<td><em>V. japonicum</em></td>
<td>Japan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>95</td>
<td>156</td>
<td>195</td>
<td>3.28E-04</td>
</tr>
<tr>
<td><em>V. luzonicum</em></td>
<td>Taiwan</td>
<td>Leaf exchanger</td>
<td>Leaf flush</td>
<td>205</td>
<td>89</td>
<td>187</td>
<td>4.31E-56</td>
</tr>
<tr>
<td><em>V. luzonicum</em></td>
<td>Taiwan</td>
<td>Leaf exchanger</td>
<td>Leaf drop</td>
<td>133</td>
<td>154</td>
<td>187</td>
<td>3.37E-07</td>
</tr>
<tr>
<td><em>V. odoratissimum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf flush</td>
<td>220</td>
<td>151</td>
<td>187</td>
<td>1.35E-08</td>
</tr>
<tr>
<td><em>V. odoratissimum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>91</td>
<td>180</td>
<td>186</td>
<td>4.26E-01</td>
</tr>
</tbody>
</table>
Table 1 (Continued)

<table>
<thead>
<tr>
<th>Species</th>
<th>Location</th>
<th>Leaf habit</th>
<th>Event</th>
<th>Total leaves</th>
<th>Mean Julian day (observed)</th>
<th>Mean Julian day (null)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>V. parvifolium</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf flush</td>
<td>405</td>
<td>99</td>
<td>187</td>
<td>2.37E-86</td>
</tr>
<tr>
<td><em>V. parvifolium</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>396</td>
<td>332</td>
<td>187</td>
<td>1.95E-174</td>
</tr>
<tr>
<td><em>V. phlebotrichum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>198</td>
<td>97</td>
<td>202</td>
<td>4.50E-62</td>
</tr>
<tr>
<td><em>V. phlebotrichum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>196</td>
<td>306</td>
<td>202</td>
<td>1.14E-57</td>
</tr>
<tr>
<td><em>V. propinquum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>295</td>
<td>122</td>
<td>187</td>
<td>8.12E-33</td>
</tr>
<tr>
<td><em>V. propinquum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>125</td>
<td>165</td>
<td>187</td>
<td>4.10E-02</td>
</tr>
<tr>
<td><em>V. sieboldii</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>333</td>
<td>100</td>
<td>215</td>
<td>3.70E-105</td>
</tr>
<tr>
<td><em>V. sieboldii</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>332</td>
<td>277</td>
<td>215</td>
<td>5.33E-27</td>
</tr>
<tr>
<td><em>V. symmodiale</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>508</td>
<td>121</td>
<td>187</td>
<td>3.73E-53</td>
</tr>
<tr>
<td><em>V. symmodiale</em></td>
<td>Taiwan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>399</td>
<td>328</td>
<td>187</td>
<td>1.31E-156</td>
</tr>
<tr>
<td><em>V. taiuense</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>273</td>
<td>105</td>
<td>187</td>
<td>9.60E-47</td>
</tr>
<tr>
<td><em>V. taiuense</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>109</td>
<td>180</td>
<td>187</td>
<td>7.10E-01</td>
</tr>
<tr>
<td><em>V. taiwanianum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>326</td>
<td>145</td>
<td>187</td>
<td>1.77E-11</td>
</tr>
<tr>
<td><em>V. taiwanianum</em></td>
<td>Taiwan</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>245</td>
<td>208</td>
<td>187</td>
<td>4.00E-01</td>
</tr>
<tr>
<td><em>V. arcelatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>96</td>
<td>111</td>
<td>203</td>
<td>1.62E-19</td>
</tr>
<tr>
<td><em>V. arcelatum</em></td>
<td>Japan</td>
<td>Deciduous</td>
<td>Leaf drop</td>
<td>96</td>
<td>271</td>
<td>204</td>
<td>1.95E-09</td>
</tr>
<tr>
<td><em>V. vernicosum</em></td>
<td>Borneo</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>126</td>
<td>117</td>
<td>169</td>
<td>3.25E-07</td>
</tr>
<tr>
<td><em>V. vernicosum</em></td>
<td>Borneo</td>
<td>Evergreen</td>
<td>Leaf drop</td>
<td>17</td>
<td>199</td>
<td>170</td>
<td>3.60E-01</td>
</tr>
</tbody>
</table>

Note: Species in boldface exhibited leaf senescence patterns statistically indistinguishable from our generated null distributions. Evergreen species marked with an asterisk exhibited senescence patterns significantly different from the null at P < .05. All species exhibited highly significant leaf flushes, occurring earlier in the year than the null expectation.

Literature Cited


Sympoium

Geographical Variation in Community Divergence: Insights from Tropical Forest Monodominance by Ectomycorrhizal Trees*

Tadashi Fukami,1,† Mifuyu Nakajima,1 Claire Fortunel,2 Paul V. A. Fine,3 Christopher Baraloto,4 Sabrina E. Russo,5 and Kabir G. Peay1

1. Department of Biology, Stanford University, Stanford, California 94305; 2. Department of Ecology and Evolutionary Biology, University of California, Los Angeles, California 90095; 3. Department of Integrative Biology and University and Jepson Herbaria, University of California, Berkeley, California 94720; 4. L’Institut National de la Recherche Agronomique, Unité Mixte de Recherche Ecologie des Forêts de Guyane, Kourou, French Guiana; and Department of Biological Sciences and International Center for Tropical Botany, Florida International University, Miami, Florida 33199; 5. School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588

Online enhancements: supplementary material. Dryad data: http://dx.doi.org/10.5061/dryad.cf0k7.

ABSTRACT: Convergence occurs in both species traits and community structure, but how convergence at the two scales influences each other remains unclear. To address this question, we focus on tropical forest monodominance, in which a single, often ectomycorrhizal (EM) tree species occasionally dominates forest stands within a landscape otherwise characterized by diverse communities of arbuscular mycorrhizal (AM) trees. Such monodominance is a striking potential example of community divergence resulting in alternative stable states. However, it is observed only in some tropical regions. A diverse suite of AM and EM trees locally codominate forest stands elsewhere. We develop a hypothesis to explain this geographical difference using a simulation model of plant community assembly. Simulation results suggest that in a region with a few EM species (e.g., South America), EM trees experience strong selection for convergent traits that match the abiotic conditions of the environment. Consequently, EM species successfully compete against other species to form monodominant stands via positive plant-soil feedbacks. By contrast, in a region with many EM species (e.g., Southeast Asia), species maintain divergent traits because of complex plant-soil feedbacks, with no species having traits that enable monodominance. An analysis of plant trait data from Borneo and Peruvian Amazon was inconclusive. Overall, this work highlights the utility of geographical comparison in understanding the relationship between trait convergence and community convergence.

Keywords: community assembly, mycorrhizae, plant-soil feedback, plant traits, priority effects, species pools.

* This issue originated as the 2016 Vice Presidential Symposium presented at the annual meetings of the American Society of Naturalists.
† Corresponding author; e-mail: fukami@stanford.edu.

ORCID:s Fukami, http://orcid.org/0000-0001-5654-4785; Nakajima, http://orcid.org/0000-0001-7778-2029.

DOI: 10.1086/692439

Introduction

At the species level, convergence is defined by distantly related species sharing similar traits, but convergence is also possible at the community level, in which distantly located communities develop to have similar species abundance distribution, functional group composition, and other aspects of community structure (Fukami 2009). These forms of convergence are thought to result from the predictable response of independently evolved species (e.g., Reich et al. 1997; Conte et al. 2012) and separately assembled communities (e.g., Samuels and Drake 1997; Li et al. 2016) to similar environmental conditions. Identifying factors that promote or prevent convergence is therefore fundamental to the understanding of predictability at both the species and community levels of biological organization (Stern 2013). Despite this duality of convergence, the link between trait convergence and community convergence remains poorly investigated (Cavender-Bares et al. 2009). It may seem obvious that trait convergence automatically translates into community convergence (Melville et al. 2006). As we argue in this article, however, the link may not be so straightforward, particularly when multiple lineages of species converge in traits among themselves but diverge from members of other lineages, which can promote divergence—rather than convergence—of community structure.

How does community convergence and divergence take place? Insights on this question can be found as early as Clements, who developed the climax concept of plant succession (e.g., Clements 1936). According to this well-known concept, communities converge to a predictable species composition determined by the abiotic environment (but see Gleason 1927). More recently, increasing evidence suggests that community divergence—in the form of alternative sta-
ble states—may be more common than previously thought (Schröder et al. 2005), but understanding when alternative stable states emerge remains a major challenge for community ecologists (Petraitis 2013). Local communities are considered to be in alternative stable states when they diverge in species composition, even though the communities share the same environmental conditions and the same species pool (Fukami 2015). This divergence is driven by priority effects, in which the order or initial abundance in which species arrive influences the effects that species have on one another in local communities (Palmgren 1926; Sutherland 1974; Drake 1991). As factors determining when alternative stable states occur, more attention has been paid to the abiotic conditions of local habitats than to the trait values of potential colonists in the regional species pool (but see, e.g., Fox 1987; Fargione et al. 2003). Consequently, the connection between convergence in species traits and in community structure remains poorly understood. In this article, we explore this connection, using tropical trees and their association with mycorrhizal fungi as an illustrative example.

**Mycorrhizal Association and Forest Monodominance**

A classic case of convergent evolution in plants is their association with mycorrhizal fungi (Brundrett 2002). There are at least 19 evolutionarily independent origins of ectomycorrhizal (EM) symbiosis in plants (Koele et al. 2012) and more than 60 in fungi (Tedersoo et al. 2010). In addition, plants associated with EM fungi may potentially experience common selective pressure and, because of it, undergo further convergent evolution in other traits, particularly those related to resource economy (e.g., Read 1991; Cornelissen et al. 2001; Phillips et al. 2013; but see Koele et al. 2012).

As a result, EM host plant species may form a species alliance (sensu Van Nes and Scheffer 2004). That is, EM host plants may modify the local environment to make it more favorable to the members of their own alliance (EM plant species) than to those of other plants, which are often associated with arbuscular mycorrhizal (AM) fungi. Thus, if by chance a locality were initially dominated by EM plants, it might persist as such owing to positive feedbacks resulting from this alliance. Consequently, it is possible that local plant communities diverge as alternative stable states, each developing as either EM or AM dominated as a result of priority effects, as we explain in more detail below.

A particularly striking example of such alternative stable states is seen in the phenomenon of tree species monodominance in tropical forests, in which forests otherwise characterized by highly diverse plant communities are dotted by occasional stands in which one tree species is far more dominant than any other (Richards 1952; Janzen 1974; Connell and Lowman 1989; Hart et al. 1989; Torti et al. 2001). In tropical forests, such stands can range in area from one to several thousand hectares (Peh et al. 2011). The term “monodominance” has been variously defined in the tropical forest literature, but here we consider those cases in which one species contributes more than 60% of total basal area (Hart et al. 1989) or more than 25% of total stem density of a forest stand that is surrounded by mixed forest.

There are a number of hypothesized mechanisms by which monodominance can emerge (Torti et al. 2001; Frederickson et al. 2005; McGuire et al. 2008; Peh et al. 2011), but one factor that is frequently associated with monodominance is the mycorrhizal status of the plants (Hart et al. 1989). Tree species that host any type of mycorrhiza can form monodominant stands. For example, multiple species from primarily AM plant families (Brundrett 2009)—such as the Apocynaceae, Burseraceae, Euphorbiaceae, Fabaceae, Lauraceae, Moraceae (reviewed in Peh et al. 2011), and Rubiaceae (Frederickson et al. 2005)—form monodominance. However, monodominance-forming species are disproportionately from EM lineages within the Dipterocarpaceae, Fabaceae, Fagaceae, and Juglandaceae (reviewed in Peh et al. 2011; Smith et al. 2013; Corrales et al. 2016). One defining feature of monodominant stands of EM host tree species in Neotropical rainforests is that they are embedded in forest with a more even abundance distribution of primarily AM host trees. EM host monodominance is those forests is particularly dramatic, given that far fewer tree species form EM compared with AM associations in typical mixed forest stands and that EM host species are often found at low abundances outside of the monodominant patches (Henkel 2003; Corrales et al. 2016).

The mechanisms that cause mycorrhizal status to generate monodominance are not fully understood but likely involve positive plant-soil feedback as one form of priority effects. EM monodominant stands are thought to result in part from positive feedback caused by their EM association, in which local soil conditions are made more favorable to EM than to AM host trees (Dickie et al. 2014). These changes in local soil conditions may involve the species composition of mycorrhizal fungi and other soil microbes or plant-induced nutrient depletion (Corrales et al. 2016; Peay 2016). Priority effects driven by such self-enforcing habitat modification—or niche construction (sensu Odling-Smeee et al. 2003)—can lead to alternative stable states, in which EM trees may form monodominant stands when they establish earlier than other species after local disturbance but are otherwise excluded by AM trees.

Intriguingly, however, these potential instances of alternative stable states with either high-diversity AM assemblages or low-diversity EM monodominant stands are observed in only some tropical regions (e.g., South America and Africa) but not in others (e.g., Southeast Asia), where a diverse suite of AM and EM plant species codominate. To our knowledge, no study has directly addressed why this difference exists,
yet this geographical contrast presents itself as an opportunity to gain a better understanding of how and when monodominance may arise as an alternative stable state.

**Natural History of Mycorrhizal Association as a Motivation for This Article**

The purpose of this article is to develop a biogeographical hypothesis to explain why EM monodominance arises in some tropical forest regions and not in others. More generally, our goal is to use monodominance as a case in point to discuss how community convergence can be influenced by trait convergence and how the extent of this effect can be mediated by the properties of the species pool.

Our motivation for this goal comes from the idea that EM monodominance may be associated with convergent values of plant traits related to resource economy, such as leaf chemistry and decomposability (Torti et al. 2001; Peh et al. 2011). Traits of EM host plant species (such as specific leaf area, leaf C and N content, and C:N ratio) have been found to differ significantly from those of AM host plant species (Read 1991; Phillips et al. 2013), suggesting that several lineages of EM plant species may have converged and, in the process, as a functional group diverged from AM plant species in these traits. Even so, there can still be substantial trait variation across plant species within the EM group (Koele et al. 2012), which may affect the strength and direction of plant-soil feedbacks. For example, interspecific variation in leaf C:N ratios may increase the variation in the strength and direction of plant-soil feedbacks among EM species (Ke et al. 2015).

In this article, we seek to link the possibility that interspecific trait variation is associated with variation in plant-soil feedbacks with one aspect of forest communities that is highly variable among tropical regions: the size of the local pool of tree species that host EM fungi (Peay 2016). This size ranges from only a few species in South America (e.g., Baraloto et al. 2011) to hundreds in Southeast Asia (e.g., Brearley 2012). We focus on this difference among regions in species pool size and the associated potential variation in the strength of species interactions mediated by plant-soil feedbacks. We hypothesize that a diverse EM species pool and the resulting complexity in local interactions among EM species makes trait convergence in these species difficult and that this obstacle to trait convergence in turn prevents the strong divergence of local communities that is necessary for the emergence of monodominant patches.

To develop this hypothesis, we use results from computer simulation of plant community assembly. The simulation model is not intended to replicate actual community assembly to quantitatively predict community patterns, but instead to explore possible qualitative outcomes arising from a small set of assumptions that characterize species interactions mediated by plant-soil feedbacks. We also examine whether empirical leaf functional trait data are consistent with the hypothesis developed with our model. Variation in leaf functional traits reflects selection for ecological strategies of resource acquisition and conservation (Donovan et al. 2011; Reich 2014) and is correlated with mycorrhizal status (Read 1991; Phillips et al. 2013). The leaf data we use for this purpose are from two typical mixed forests in different geographic regions. One is from a Neotropical forest in the Peruvian Amazon, where EM plants are rare and the EM plant species pool is small. The other is from a Paleotropical forest in Borneo, where EM plants are common and the EM plant species pool is large.

**Simulation Methods**

**Overview**

The simulation model we used is a modified version of Fukami and Nakajima’s (2013) individual-based, spatially implicit model of plant community assembly through stochastic sequential immigration of species from an external species pool. In this model, which is built on traditional plant competition models (Chesson 1985; Pacala and Tilman 1994; Mouquet et al. 2002; Fukami and Nakajima 2011), plant individuals compete for local resources during the individual establishment stage. They also affect one another’s competitiveness via plant-soil feedback. We assume that plants can take on one of two mycorrhizal associations, EM or AM. There are other potential mycorrhizal states (e.g., non-mycorrhizal or ericoid mycorrhizal), but EM and AM associations represent most tree species and individuals in tropical forests (McGuire et al. 2008; Brundrett 2009). We also assume that plants that host both EM and AM fungi can be considered primarily EM associated, because the root systems of these plants tend to become dominated by EM fungi (Egerton-Warburton and Allen 2001).

We vary the size of the EM host species pool (the number of EM host species) while holding constant the number of AM species in the pool, and we examine how the size of the EM species pool influences the way local plant communities assemble, particularly with respect to the emergence of the alternative stable state of EM monodominance. By replicating community assembly in many local forest patches with the same species pool, we examine how species abundance distributions may vary among patches that have the same set of local abiotic conditions within them but differ in the history of stochastic species arrival. In particular, we assess the conditions under which the local abundance of EM hosts shows strong bimodality, in which a single EM host species becomes monodominant in some local patches but remains rare or absent in all other patches.

As in Fukami and Nakajima’s (2013) model, plant recruitment occurs in sites that are arranged in a patch. Plants
are recruited from both seeds produced by established individuals within the patch and immigration from an external species pool. For immigration, species of plants were randomly chosen for immigration each time step from the species pool. The chosen species immigrated as a small number of propagules to a local forest patch consisting of 1,000 recruitment sites. One thousand recruitment sites roughly correspond to the sapling abundance we may expect in a hectare, the scale at which small monodominant stands are observed. All recruitment sites were initially empty. Subsequently, only one plant individual could establish and produce propagules in each recruitment site, even when multiple propagules arrived from the species pool or from local dispersal within the patch. Propagules were distributed randomly to recruitment sites within the patch each time step. Of the propagules that arrived at a recruitment site, the one that belonged to the species that best fit the environmental condition of the recruitment site could establish and produce propagules (determination of environmental fit is described in more detail below). All plants within a patch produced propagules once each time step (e.g., each year or each mast event) until they died. All plants died with a fixed probability, and when they did, recruitment sites became empty and available for a new plant individual to establish. Real tree species vary in dispersal ability, fecundity, and mortality rates, but we kept these constant across species to focus on the mechanisms of monodominance in relation to our hypothesis. This process of immigration, arrival, establishment, reproduction, and death was repeated for 400 time steps. All simulations were carried out using Mathematica 8.0 (Wolfram Research, Champaign, IL). Code and data are available as supplementary material, available online.

**Species Pools**

Regional species pools each contained 100 AM plant species and one to 20 EM plant species. Each species was defined by a value $R_i$, which can be thought of as a multitrait phenotype that determines how well species perform in local abiotic conditions during the recruitment stage (see details below). Values of $R_i$ were chosen randomly between 0 and 1 from a beta distribution, where the probability density for value $x$ was proportional to $x^{a-1}(1-x)^{b-1}$ (Mouquet et al. 2002). We set $a = b$, which causes the probability density for $R_i$ to have a peak at 0.5, but the value of $a$ was varied in order to examine the effect that the amount of variation in $R_i$ among species had on community assembly. We refer to $1/a$ as interspecific phenotypic variability $v$. A larger $v$ means that $R_i$ values are more variable among species (fig. 1B). In addition to $R_i$, species were also characterized by a set of values ($S_{ij}$) that defined the strength of plant–soil feedback, in which the presence of a plant individual belonging to species $j$ at a recruitment site during a given time step changed the competitive ability of a plant belonging to species $i$ at that recruitment site during the following time step (see details below).

**Local Patches**

Local forest patches each consisted of 1,000 recruitment sites. Each recruitment site could accommodate only one adult tree and was characterized by a value that represents abiotic conditions ($H_k$), chosen randomly between 0 and 1 from a beta distribution. For the beta distribution for $H_k$, we set $a = b = 2$, which—as with the case of $R_i$—produces a

![Figure 1](image)

**Figure 1:** Model description.  
(A) Schematic summary of the strength of plant-soil feedback ($S_{ij}$) assumed in the simulation model. The values indicated were used for the simulation shown in figures 2–4 but changed as specified in figure 5 to evaluate the effect of feedback strength on community assembly. An $S_{ij}$ of 0 means no feedback. An $S_{ij}$ value that is $>1$ means positive feedback. All intraspecific feedbacks (i.e., $i = j$) were assumed to have $S_{ij}$ of 1.1, and $S_{ij}$ values for interspecific feedbacks (i.e., $i \neq j$) were assigned a random value between 1 and 1.1, so that they are all positive, but their strengths vary with species identity and are never greater than intraspecific feedbacks.  
(B) Probability distribution for abiotic condition of recruitment sites ($H_k$; dashed line) and for species phenotypic value ($R_i$) among ectomycorrhizal species in the species pool under different degrees of species variability ($v$; which is $1/\beta$ for the $\beta$ distribution for $R$).

1. Code that appears in the *American Naturalist* is provided as a convenience to the readers. It has not necessarily been tested as part of the peer review.
peak of $H_k$ values at 0.5 (fig. 1B). For all simulations in a run, the $H_k$ values remained the same.

Competitive Ability

Values of $H_i$, $R_i$, and $S_{ij}$ together determined competitive ability ($C_{ik}$) of species $i$ at recruitment site $k$ when the recruitment site was occupied by species $j$ at the previous time step:

$$C_{ik} = (1 - |H_k - R_i|)S_{ij}.$$  

Thus, assuming that $S_{ij} = 1$ (i.e., no effect of species $j$ and species $i$ through plant-soil feedback, as explained below), a species would have a high value of $C_{ik}$ if it had a value of $R_i$ that is close to that of $H_k$ as in Mouquet et al.’s (2002) model. In our simulation, a close match between $H_k$ and $R_i$ is most likely when a species’ $R_i$ value is 0.5, since $H_k$ also has a peak at 0.5 (fig. 1B).

Plant-soil feedback moderates competitive ability, with the value of $S_j$ defining the direction and strength of the feedback. Specifically, $S_j$ defines the effect of species $j$ occupying a given recruitment site during a given time step on the competitive ability of species $i$ in that microsite during the following time step. Thus, $S_j$ affects the competitive ability of species $i$ independently of both recruitment site condition (as defined by $H_i$) and species phenotype (as defined by $R_i$) and represents the amount by which the difference between the abiotic environment and the species’ match to the environment is improved ($S_j > 1$), worsened ($S_j < 1$), or not affected ($S_j = 1$) via plant-soil feedback (Fukami and Nakajima 2013). That is, if $S_j = 1$, there is no net effect of plant-soil feedback, whereas $S_j > 1$ and $S_j < 1$ represent positive and negative plant-soil feedback, respectively.

We set the values of $S_j$ to attempt to represent the interactions within and between AM and EM plants that are observed in nature (fig. 1A). There are many ways in which plant-soil feedback could be structured, but here we assume that plants facilitated the growth of conspecific individuals via local accumulation of specific mycorrhizal fungi that are beneficial to them, via modification of soil nutrient composition to the species’ own benefit via their specific mycorrhizal association (Corrales et al. 2016) or via better protection from pathogens through added physical or chemical defenses in mycorrhizal roots (Duchesne et al. 1989; Newsham et al. 1995; Bennett et al. 2017). Plants can also negatively affect conspecifics by, for example, accumulating soil pathogens (Kulmatiski et al. 2008). However, studies that examine net effects of pathogens plus mycorrhizae often find overall positive effects from soil biota (e.g., fig. 1B in Mangan et al. 2010; see also Cortois et al. 2016; Bennett et al. 2017; Teste et al. 2017). Similarly, in a large analysis of forest plots in North America and New Zealand, Dickie et al. (2014) found that most tree species exhibited positive density dependence. We therefore assumed in our model that positive effects outweigh negative ones to result in net positive plant-soil feedback for conspecifics. Accordingly, we set $S_{ij} = 1.1$ when $i = j$.

We also assumed that EM tree species shared similar nitrogen economy (Corrales et al. 2016) as well as similar associations with EM fungi. As a result, EM plant species can benefit each other to some extent, but the strength of interspecific plant-soil feedback is variable among EM plant species pairs and is never greater than that of intraspecific feedback. We believe that these assumptions are plausible because evidence suggests that EM fungi tend to be generalists with high overlap between co-occurring species (Kennedy et al. 2003; Peay et al. 2015; but see Bennett et al. 2017) and that shared EM fungi can facilitate interspecific recruitment (Horton et al. 1999; Nara 2006b). In addition, there is evidence that different combinations of EM fungi and plants can have different effects on plant growth, even among closely related species (Nara 2006a; Fransson et al. 2015). Consequently, when $i \neq j$, we assigned $S_{ij}$ values by taking a uniform randomly drawn value between 1 and 1.1 for each $S_{ij}$. For simplicity, we assume that this same pattern of intra- and interspecific feedback strengths also holds for AM plant species. In other words, EM and AM host species have similar capacities for positive plant-soil feedback in our model. Finally, we also assumed that EM and AM plant species did not facilitate each other through plant-soil feedback. Thus, $S_{ij} = 1$ between all pairs of an EM species and an AM species.

The direction and strength of plant-soil feedback in tropical forests have not been thoroughly characterized, and it is uncertain whether the assumptions specified above are always realistic. To examine whether model results are sensitive to these assumptions, we ran two sets of additional simulations. In the first, we assumed that among EM host plants, $S_{ij}$ took a uniform randomly drawn value between 1 and 1.1, even when $i = j$, in order to relax the assumption that all intraspecific feedbacks were more or equally positive compared with any interspecific feedbacks (fig. S1B; figs. S1–S5 are available online). In a second set, we used a combination of alternative assumptions. Specifically, we assumed (1) variable positive intra- and interspecific feedbacks for EM host plants, (2) variable intra- and interspecific feedbacks (both positive and negative) for AM host plants, and (3) negative effects of EM host plants on AM host plants. For assumption 1, we tried two variants. In one (variant 1), we used the original assumption as above, that is, that the strength of interspecific plant-soil feedback is variable among EM plant species pairs and is never greater than that of intraspecific feedback (fig. S1C). The other (variant 2) is that interspecific and intraspecific feedbacks are all positive but...
equally variable, such that some intraspecific feedbacks are less positive than some interspecific feedbacks (fig. S1D).

Community Assembly
At each time step, each species in the species pool immigrated to the local patch with probability \( I \). For each of the 100 AM species in the pool, we used \( I = 0.05 \). For each EM species, we used \( I = 1/(\text{number of EM species in the pool}) \). For example, when we had two or 20 EM species in the pool, \( I \) was 0.5 or 0.05 for each EM species, respectively. This way, we standardized for the total immigration frequency for all EM plants across the gradient of the size of the EM species pool. We also did additional simulation in which \( I = 0.05 \) for each species, regardless of the size of the EM species pool used, in order to decouple the effects of regional species richness per se from those of regional relative abundance.

At each recruitment site in the local patch, species \( i \) arrived with probability \( 1 - \exp[-(P_i + FN)/1000] \) (total number of recruitment sites, i.e., 1,000) at each time step. Here, \( P_i \) is the number of propagules of species \( i \) that immigrate from the species pool (20 propagules for species chosen for that time step for immigration from the species pool, and 0 seed for all other species), \( F \) is fecundity (50 for all species), and \( N_i \) is the number of plants belonging to species \( i \) in the local patch (0 for all species in the first time step, i.e., at \( t = 1 \), which is analogous to modeling recruitment after a stand-replacing disturbance). The specifications were chosen so as to have values that likely fall within a plausible range. When the number of recruitment sites that were assigned to receive a propagule of species \( i \) exceeded \( P_i + FN \), (which rarely happens), \( P_i + FN \), recruitment sites were randomly selected from these recruitment sites, and a propagule of the species were assigned only to the selected recruitment sites.

Given this probability, there were three possibilities regarding plant establishment and seed production in each recruitment site. First, if the recruitment site were already occupied by a plant, that plant remained there. In other words, we assumed that seeds could not displace established adults. Second, if the recruitment site were empty and no species arrived at that recruitment site, it remained empty. Third, if the recruitment site were empty and one or more species arrived at that recruitment site, of those species that arrived, the one with the greatest value of \( C_{\text{irr}} \) (independent of the number of propagules of each species) was assumed to occupy the recruitment site and produce propagules starting the following time step. The seed-to-adult recruitment processes are important in many of the dynamics modeled here, but processes operating during those stages per se were not the focus of our model, which was intended to identify the minimal processes required to produce monodominance.

After plant establishment was completed for all recruitment sites, plants occupying a recruitment site died with probability \( m \), which was 0.4 for all species. Thus, competitive ability in our model is not related to dispersal, fecundity, or baseline survival; it dictates only a species’ ability to win a recruitment site. We assembled communities for 400 time steps. From visual inspection of results, 400 time steps seemed long enough for most communities to reach an equilibrium state (fig. 2).

Manipulating Species Pool Size and Species Variability
We manipulated two factors—the number of EM species in the species pool and the amount of variation in \( R_i \) among

![Figure 2: Examples of simulated population dynamics for two forest patches, showing ectomycorrhizal (EM) monodominance (A) and arbuscular mycorrhizal (AM) dominance (B) as alternative stable states. Both A and B show results for one instance of local community assembly under the same species pool, and each line indicates a species, with warm colors denoting EM tree species and cold colors denoting AM tree species. In these examples, there were 100 AM tree species and 10 EM tree species in the species pool, and species variability among EM species (\( \nu \)) was set to be small (\( \nu = 0.0001 \), with all EM species having an \( R_i \) value very close to the optimal, 0.5 (see fig. 1B). Depending on assembly history, local communities develop either as EM monodominance (A), in which a single EM species becomes dominant (orange line) and all others are AM species, or as an AM-only community (B), in which a few EM species were able to colonize the forest patch initially but became locally extinct because they were not common initially and were excluded competitively by AM species through positive plant-soil feedback among AM species.](image-url)
EM host species (i.e., species variability $v$, as defined in “Species Pools”)—to examine their effects on EM abundance in assembled local communities. To this end, under each of the species pools we used (i.e., 16 pools in fig. 3), we ran the simulation 100 times, with each run considered one instance of community assembly in a local forest patch. Under each species pool, the replicated 100 patches had the same set of recruitment sites, which allowed us to evaluate whether plant communities that develop in different patches that share the same set of $H_i$ values in them diverged in species composition as a result of random variation in the history of species arrival from the species pool. If they did diverge, that would be evidence for alternative stable states. We were particularly interested in assessing the conditions under which EM monodominance arose. We defined EM monodominance as the case in which only one dominant EM species occupied one-fourth (25%) or more of the available recruitment sites in a patch, with the other sites being occupied by AM species.

**Simulation Results and Discussion**

Simulation results verified that communities with (fig. 2A) and without (fig. 2B) monodominant EM species could develop as alternative stable states in our model. Under the same species pool, communities were either dominated by AM species without any EM species present or had at least 25% EM individuals, representing two alternative stable states, similar to those seen in some Neotropical rainforests. By varying species pool size and phenotypic variability (fig. 3), we found that EM monodominance arose as an alternative state whenever trait variability among EM species was low. In these cases, all EM species in the species pool had an $R_i$ value that was nearly optimal for the abiotic conditions available in the forest patch. Under this condition, a single dominant EM species was competitive enough to form monodominant stands when it arrived early during local community assembly but not when it arrived late, as indicated by the strongly bimodal pattern in the top row in figure 3. One of the two modes had no EM species, and the other mode had only one EM species, even when the species pool contained multiple EM species.

Regardless of species pool size, as $R_i$ values became more variable among EM species in the local species pool (i.e., moving from top to bottom rows in fig. 3), the EM total abundance became more variable among forest patches, and the bimodal pattern became increasingly obscure. The frequency of monodominance declined with greater phenotypic variability when there was more than one EM species in the pool because, although the number of EM trees in a patch might be high, these patches consisted of multiple EM species.

The additional simulation indicated that these results generally held under the alternative assumptions examined, but only when all intraspecific feedbacks were more or equally positive compared with interspecific feedbacks (fig. S1C). Otherwise, multiple EM plant species often coexisted as a result of mutual facilitation among them, with only a small number of monodominance patches developing (fig. S1B, S1D). Therefore, one condition for monodominance formation in our model is stronger positive intraspecific relative to interspecific feedback among EM trees, which is consistent with some recent empirical data (Bennett et al. 2017). Strong intraspecific feedback is possible if, for example, a tree species cultivates an assemblage of highly beneficial EM fungi. The additional simulation in which immigration rate $I$ was 0.05, regardless of the size of the EM species pool, indicated that the number of monodominant patches is positively correlated with the immigration rate per EM species (fig. S2A, S2C).

Together, these results suggest that there is no effect of species pool size per se on whether EM monodominance will emerge or on how strong the bimodal pattern reflecting monodominance formation will be. Rather, it is the amount of phenotypic variability among EM species that shapes the pattern of EM abundance in local patches. Specifically, reduced variability tends to cause greater frequency of monodominance (fig. 3). In order to directly test for the relationship between species pool size and phenotypic variability of EM species, we ran additional simulations to examine how species pool size influences the relationship between species phenotypic value ($R_i$) and the abundance of EM species when EM species variability in the species pool is large ($v = 1$). We found that when species pool size was small, there was a relatively good correspondence between $R_i$ value and EM species abundance (fig. 4A). We call this correspondence 'trait-environment matching', meaning that the closer species are in their phenotype $R_i$ to the optimal (i.e., 0.5 in our simulation), the more abundant they tend to become. In this case, the abundances of species follow what is expected on the basis of their match to the environment (i.e., $|R_i - H_i|$).

As species pool size was increased, this correspondence became weaker (fig. 4), such that species with an optimal $R_i$ value (closer to 0.5) were not necessarily more abundant than those with a suboptimal value (away from 0.5). When the species pool was diverse, two additional abundance peaks appeared (at $R_i$ of around 0.2 and 0.8), so that species that did not have an $R_i$ close to 0.5 could nevertheless be as abundant as those with an optimal $R_i$, as most clearly seen in figure 4D. In short, for EM species in our model, it is adaptive to have traits that best match the underlying abiotic environmental conditions if there are not many EM species in the species pool. However, if the species pool has many EM species, our results suggest that there can be disruptive natural selection that keeps some species away from having the abiotically optimal $R_i$ value if the species pool has many species. As a
Figure 3: Histograms showing ectomycorrhizal tree abundance in a local forest patch across a gradient of the number of ectomycorrhizal tree species in the species pool and the species phenotypic variability ($v$) among ectomycorrhizal tree species in the species pool (see fig. 1B). The $S_i$ values used are as specified in figure 1A. The number above each bar indicates the mean number of ectomycorrhizal tree species that were present in the patch, averaged across all patches belonging to the bar. The number in the top right corner of each histogram indicates the percentage of the patches that showed monodominance by our criteria (only one ectomycorrhizal tree species in the patch and more than 25% of the stems belong to that species). Each histogram shows results for one species pool, but we replicated species pools for each histogram and found that results were consistent across replicated pools.
result, convergent trait evolution that results in low interspecific trait variability and tight trait-environment matching in the species pool may be more likely when the EM species pool is small, whereas divergent trait evolution may arise when the EM species pool is large.

Why do EM species with suboptimal phenotypes obtain comparable densities to those with optimal phenotypes when the species pool is large (fig. 4D)? We suspected that it was because the abundance of each EM species is affected so strongly by the plant-soil feedback—which is made more complex by the larger number of EM species that each have different strengths of plant-soil feedback—that the underlying influence of abiotic conditions is overwhelmed and obscured. To test this idea, we reran the simulation, but this time with the strength of plant-soil feedback among EM and/or AM species (S_e) all set to 1.1 instead of being drawn uniform-randomly from [1, 1.1]. Setting all S_e values to 1.1 makes interspecific feedback strengths all the same and equal to intraspecific feedback strength, eliminating the complexity of feedback strength. In this additional simulation, we found good trait-environment matching—such as the one seen under small species pools (fig. 4A)—if all S_e values among EM species were set to 1.1 (fig. 5B). Poor trait-environment matching was not, however, eliminated when S_e values among AM species were set to 1.1 (fig. 5C, 5D). These results support our hypothesis that it is the variation in interspecific feedback strengths among EM species that makes trait-environment matching poor and monodominance unlikely when the species pool is large.

In summary, our results suggest that two interdependent processes influence the likelihood of trait and community convergence. First, trait-environment matching that leads to trait convergence among EM species may be less likely under a larger EM species pool because of local plant-soil feedbacks that vary in strength. Second, weak trait convergence may in turn impede the development of EM monodominance, which represents an extreme case of community divergence.

**Empirical Trait Data Analysis**

**Study Systems and Methods**

To begin to assess whether empirical data are consistent with the hypothesis developed with the model above, we analyzed leaf trait data from trees in two geographic regions: Southeast Asia (Borneo), where EM trees are common in most lowland forests, and South America (Peruvian Amazon), where EM trees are generally rare but can become monodominant. Both regions contain large areas of lowland tropical rainforests with extraordinary levels of tree di-
versity. The Peruvian Amazon has greater regional tree diversity (approximately 5,000 species; Pennington et al. 2004) than Borneo (approximately 3,000 species) probably because it is connected to the rest of the Amazon basin, the world’s largest contiguous area covered by lowland rainforest. Thus, at the scale of a local patch of forest (e.g., Lambir Hills in Borneo or Allpahuayo-Mishana in Peru), we expect the regional tree species pool to be reasonably similar, with Peru being slightly larger. However, the pool of EM trees is much larger in Borneo than in Peru.

For our analysis, trait data were taken from two separate studies of plant functional ecology in Peruvian Amazon (Fortunel et al. 2012, 2014) and Borneo (Russo et al. 2005, 2013). Detailed methods on data collection can be found from Fortunel et al. (2012, 2014) and in the appendix, respectively. Data in Peru were collected from multiple 0.5-ha Gentry plots, whereas those in Borneo were collected from within a single 52-ha forest dynamics plot in Lambir Hills National Park. We used data on four leaf traits that are involved in the fast-slow plant economics spectrum (Reich 2014) and that are thought to be strongly filtered by environmental conditions (Fortunel et al. 2014): specific leaf area (SLA), leaf carbon and nitrogen concentrations by mass, and leaf C:N ratio. Leaf C:N ratio is considered a particularly important trait in determining the strength of plant-soil feedbacks (De Deyn et al. 2008; Ke et al. 2015). Data from individual plants were averaged within species. Species’ mycorrhizal status was classified at the genus or family level on the basis of Brundrett (2009) and Peay et al. (2013). Data are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.c0kr7 (Fukami et al. 2017).

For each region, we analyzed trait data from lower fertility (sand) and higher fertility (clay) soil habitats separately because these habitats differ in species pool composition within each region. The trait data set from Peru included 365 AM and 13 EM species from clay soils and 140 AM and 3 EM species from sandy soils. The trait data set from Borneo included 98 AM and 17 EM species from clay soils and 111 AM and 35 EM species from sandy soils. We took three approaches to look for differences in EM traits between the two species pools (Peru and Borneo) within soil types. First, we compared mean trait values for EM versus AM host species within each species pool, using a two-sample t-test, to evaluate our prediction that EM species are functionally distinct from AM species. Second, we compared the variance of trait values for EM versus AM species within each species pool, using a one-sided variance ratio F-test, to evaluate our prediction that EM species have larger trait variability than AM species. Third, we compared the variance in trait values between EM species pools in Borneo and Peru, using a one-sided variance ratio F-test, to evaluate our prediction that there should be greater variation in trait values among EM host species in Borneo than in Peru, where EM monodominance is found. We do not approach

---

**Figure 5:** Effect of changing $S_j$ values on the relationship between species value ($R_i$) and species abundance. Symbols are as in figure 4, except that data for each panel are based on 60 (not 100) replicate species pools. EM, ectomycorrhizal; AM, arbuscular mycorrhizal.
this question in an explicit phylogenetic framework, but our test is conservative because we predict larger trait variation in the narrower phylogenetic pool, that is, in Borneo, which had only one EM family (Dipterocarpaceae).

**Trait Analysis Results and Discussion**

We found significant differences in some mean foliar trait values between AM and EM species in both Peru and Borneo that were consistent with our model prediction (table 1; fig. 6). In Peru, foliar C:N ratio was significantly different between AM and EM species on both soil habitats, and foliar nitrogen concentration was different only on clay soil. In Borneo, foliar carbon concentration differed significantly between AM and EM species on both habitats, but no other significant differences were observed. The difference found in Peru between AM and EM host species in foliar C:N was particularly large, in contrast to virtually identical means for this trait in Borneo (fig. 6).

One possible explanation for these inconsistencies lies in the relative differences in the sizes of EM and AM host species pools in each forest. To the extent that greater richness of the species pool causes more complex plant-soil feedbacks within each mycorrhizal group, in Peru (with fewer EM host species), their traits may track environmental conditions closely, causing a large divergence from those of AM host species. Since the pool of AM species is larger, they may be influenced strongly by the complex plant-soil feedbacks, reducing their match to underlying abiotic conditions. In contrast, in Borneo, where both mycorrhizal groups are species rich, complex plant-soil interactions might have prevented species traits from tracking the environment in both EM and AM host species.

Another explanation lies in the difference in the relative soil fertility of these regions. Averaging over soil types and mycorrhizal status, foliar N was significantly lower ($F_{1,71} = 133.7, P < .001$) and C:N significantly higher ($F_{1,71} = 109.8, P < .001$) in Borneo than in Peru, and both soils at the Bornean site are particularly low in soil N, P, K, and other nutrients (Baillie et al. 2006). Low soil fertility is strongly associated with overall low foliar nutrient concentrations and higher C:N ratios (Chapin et al. 1993), which could give less latitude for differences between AM and EM host species in Borneo compared with more fertile sites. Trait differences between AM and EM host species in Borneo may be more tightly linked to leaf structure and physical defense, as suggested by the significant differences between mycorrhizal groups in foliar C, which is likely to affect leaf litter decomposition and, in turn, the strength of plant-soil feedbacks and the extent of trait-environment matching.

There were significant differences in the variance of leaf traits between AM and EM host species in Peru and Borneo. However, in contrast to our expectations, variance in traits was often larger for AM than EM host species in Borneo (three of four traits), whereas no significant differences were observed in trait variances between AM and EM host species in Peru (table 2). Furthermore, results for trait variance of EM species pools did not support our model prediction either. The variance in species traits was not larger among EM host species in Borneo than in Peru for any of the traits measured for clay ($F_{16,12} = 0.56, P = .86$; leaf carbon concentration [LCC] $F_{16,12} = 0.52, P = .89$; leaf nitrogen concentration [LNC] $F_{16,12} = 0.05, P = 1$; C:N $F_{16,12} = 0.87, P = .61$) or sandy (SLA $F_{34,2} = 2.32, P = .35$; LCC $F_{34,2} = 5.13, P = .18$; LNC $F_{34,2} = 0.03, P = 1$; C:N $F_{34,2} = 0.49, P = .85$) habitats.

How could these apparent discrepancies between model predictions and empirical patterns be reconciled? Our focus in this article has been on contrasting alternative stable states between diverse AM and monodominant EM patches, but as a family, the Dipterocarpaceae dominate this and many other forests in Southeast Asia in terms of basal area. One potential explanation for their trait convergence may then be that phylogenetic relatedness of dipterocarp species have caused them to act as if they belonged to one species, thereby allowing for trait-environment matching in the absence of complex interspecific plant-soil feedback. As for the apparent lack of trait convergence among EM relative

<table>
<thead>
<tr>
<th>Region and habitat</th>
<th>No. species</th>
<th>SLA</th>
<th>LCC</th>
<th>LNC</th>
<th>C:N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Peru:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clay</td>
<td>362</td>
<td>0.06</td>
<td>1.59</td>
<td>-4.18</td>
<td>6.15</td>
</tr>
<tr>
<td>Sand</td>
<td>140</td>
<td>0.07</td>
<td>-0.15</td>
<td>-2.87</td>
<td>4.68</td>
</tr>
<tr>
<td><strong>Borneo:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clay</td>
<td>98</td>
<td>0.08</td>
<td>-4.43</td>
<td>1.06</td>
<td>1.9</td>
</tr>
<tr>
<td>Sand</td>
<td>111</td>
<td>1.29</td>
<td>-4.55</td>
<td>0.58</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Note: SLA, specific leaf area; LCC, leaf carbon concentration; LNC, leaf nitrogen concentration; C:N, leaf C:N ratio.

* $P < .05$. 

This content downloaded from 132.236.027.111 on July 22, 2017 05:14:57 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).
to AM host species in Peru, this result may reflect low statistical power due to small sample size. The number of EM species was much smaller in Peru (13 and three species) than in Borneo (17 and 35 species).

Discussion

Taken together, our simulation results suggest one potential reason why EM monodominance is favored when the species pool has a small number of EM species. Specifically, results summarized in figure 3 indicate that monodominance can emerge only when EM species phenotypes \((R_i)\) closely match the environmental conditions available in local forest patches \((H_i)\). Results presented in figure 4 indicate in turn that such optimal trait-environment matching is more likely under small species pools because EM species can then more easily track the environment. As a result, trait evolution of EM species converges toward abiotically

Figure 6: Probability density plots showing the frequency of tree species having particular foliar C:N ratios under two soil conditions (less fertile sand or more fertile clay) in Borneo and Peru. Dark gray indicates ectomycorrhizal (EM) host species. Light gray indicates all other species, most of which are arbuscular mycorrhizal (AM) hosts. Dashed lines indicate means.
optimal conditions. Finally, results in figure 5 show that the reason why species do not track the environment under diverse species pools is because the complex plant-soil feedback among the many EM species obscures the effect of environmental heterogeneity on local species abundances. In other words, environmental filtering (Kraft et al. 2015) is weaker than priority effects.

Although we have focused on mycorrhizal associations, monodominance may arise through other pathways (Peh et al. 2011). For example, a defensive ant-plant mutualism generates monodominant stands of *Duroia hirstuta* known as devil’s gardens in the Amazon (Frederickson et al. 2005). Still, many tree species that form monodominant stands in the tropics are EM. Studies have uncovered evidence supporting a range of ways in which EM associations may generate positive plant-soil feedbacks that lead to monodominance, including common mycorrhizal networks (McGuire 2007), alteration of local nutrient economies (Corrales et al. 2016), variation in leaf chemistry (Torti et al. 2001), and EM mining of organic nitrogen (Orwin et al. 2011). Our model does not distinguish between these different mechanisms. Rather, our work is complementary to these studies in that we ask why geographic differences in species pools may lead to some tropical regions containing EM monodominant stands and other regions showing local codominance of multiple tree species of both mycorrhizal types.

We chose one simple way to represent plant-soil feedbacks. Namely, we assumed that plant-soil feedbacks were net positive and most beneficial to conspecifics. Choosing appropriate values for plant-soil feedbacks is not trivial, because in nature these feedbacks are hard to estimate. Empirical measurement of their strength should take into account multiple contributing factors, such as abiotic effects (Waring et al. 2015) and the combined effects of pathogens and mutualists (Klironomos 2002) and natural enemies (Bagchi et al. 2014). In addition, whether net feedbacks are positive or negative also depends on the choice of other heterospecific comparisons and sterilization methods used to generate the reference point. A number of other factors may influence feedback strength, such as phylogenetic relatedness (Gilbert and Webb 2007; Liu et al. 2012), local density (Comita et al. 2010; Mangan et al. 2010; Liu et al. 2012), and mycorrhizal type (Johnson et al. 2012; Dickie et al. 2014; Bennett et al. 2017).

In this study, we attempted to keep the model focused on differences in species pool size by keeping the nature of plant-soil feedback identical between AM and EM species. Through additional simulation (figs. S1, S2), we have made an initial effort to examine the robustness of our model predictions to assumptions regarding plant-soil feedback. Future iterations of this model could explore more complex plant-soil feedback structure, such as more negative (or less positive) intraspecific feedbacks and stronger positive feedbacks in EM compared with AM plants. It is also possible that EM plants are better able to make the environment less suitable for AM plants, owing to the abilities of EM fungi to access organic forms of nutrients and preempt nutrient uptake by AM fungi. This scenario can be evaluated further with our model. Finally, we assigned *R* and *S* values independently, but in nature, species with similar traits—as represented by more similar values of *R*—might have the tendency to share fungal symbionts and therefore have more positive *S*. Phylogenetic relationships of plant species, which we did not consider in this article, may prove useful in some cases as a proxy for estimating *R* and *S* values.

It is well established in plant-soil feedback theory that positive feedback causes alternative stable states, whereas negative feedback facilitates species coexistence (Bever et al. 2012). Our simulation results are consistent with this prior theory. However, most previous plant-soil feedback theories have considered interactions between only two plant species. Here, we have studied interactions among many plant species (see also Fukami and Nakajima 2011, 2013), which is the only way to directly address the effect of the number of species in the regional species pool on local community assembly driven by plant-soil feedback.

One benefit of our model is that it makes specific predictions about trait variation of plant species and about

---

### Table 2: Difference in leaf trait variance between arbuscular mycorrhizal (AM) and ectomycorrhizal (EM) tree species across two edaphic habitats in Peru and Borneo

<table>
<thead>
<tr>
<th>Region and habitat</th>
<th>No. species</th>
<th>SLA</th>
<th>LCC</th>
<th>LNC</th>
<th>C:N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AM</td>
<td>EM</td>
<td>F ratio</td>
<td>P</td>
<td>F ratio</td>
</tr>
<tr>
<td>Peru:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clay</td>
<td>362</td>
<td>13</td>
<td>1.11</td>
<td>.45</td>
<td>1.03</td>
</tr>
<tr>
<td>Sand</td>
<td>140</td>
<td>3</td>
<td>5.71</td>
<td>.16</td>
<td>4.24</td>
</tr>
<tr>
<td>Borneo:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clay</td>
<td>98</td>
<td>17</td>
<td>1.15</td>
<td>.40</td>
<td>3.17</td>
</tr>
<tr>
<td>Sand</td>
<td>111</td>
<td>35</td>
<td>1.42</td>
<td>.12</td>
<td>4.07</td>
</tr>
</tbody>
</table>

Note: SLA, specific leaf area; LCC, leaf carbon concentration; LNC, leaf nitrogen concentration; C:N, leaf C:N ratio.  
* P < .05.
mycorrhiza-driven plant-soil feedbacks under different species pool sizes, specifically that divergence in the values of traits related to local environmental fit should be greater in more diverse species pools. Our trait analysis did not yield strong support for this prediction, perhaps because these are difficult effects to measure empirically, since many factors not included in our model influence trait variation in natural settings. Moreover, uncertainties regarding the mycorrhizal status (e.g., EM or AM) of plant species in different geographical locations may also have caused the poor correspondence of model predictions and empirical patterns (Forrestel et al. 2017). In addition, detecting patterns in traits across a gradient of species pool size is made difficult by the inherent correlation between species pool size and statistical power for any analysis that uses species as a unit of replication.

Our trait analysis is only a first step to begin to evaluate the empirical relevance of the hypothesis we have developed here through the simulation modeling. Research on the links between plant traits, mycorrhizal associations, and the strength of plant-soil feedback is still at an early stage of development (Laughlin et al. 2015; Cortois et al. 2016; Bennett et al. 2017), and the dearth of relevant data may in part explain the poor correspondence of our empirical results with the model predictions. We list three future directions of empirical research. First, with respect to the determinants of monodominance, we need to better understand which traits are important for tree species’ fit to the local environment (Kraft et al. 2008; Fortunel et al. 2014; Laughlin et al. 2015; Forrestel et al. 2017) and how correlated different traits are with competitive abilities and the strength of plant-soil feedbacks (Uriarte et al. 2010; Fortunel et al. 2016). For example, some traits may be more closely related to competitive ability and others to niche differences (Mayfield and Levine 2010; Kraft et al. 2015b; Kunstler et al. 2016). Likewise, the patterns observed in the relationship between trait variability and monodominance may depend on which traits are involved in which kinds of interactions in the system. This possibility is supported by the fact that in our analyses of trait variation, some traits showed patterns consistent with our predictions, whereas others did not. To wit, plant-soil feedbacks involving EM species in Borneo may be mediated more strongly by the effects of trait variation in leaf structure and physical defense on litter decomposability than by the effects of foliar N because of the generally low fertility of soils in this Bornean forest. Thus, whether patterns of variation consistent with trait-environment matching are observed may not only depend on how correlated the traits examined are with plant-soil feedbacks but also be constrained by the local environment. Second, to complement the indirect inference of species performance and interactions through trait analysis, more direct measurements of the $C_{gk}$ values could be helpful. One approach may be leveraging trait values in neighborhood models (Uriarte et al. 2010; Fortunel et al. 2016) as estimates of the $C_{gk}$ in our model to test the effects of AM and EM trees on each other’s growth and survival. Third, we compared just two regions here, but it will be more informative to analyze data from more regions—including Africa, where monodominance also arises—to assess the potential applicability of our hypothesis across geographical regions with different evolutionary histories.

If species pool properties are key determinants of community assembly (Ricklefs and Schluter 1993; Zobel 2016), as we have considered in this article, a fundamental question is what causes differences in species pools among tropical regions in the first place. For example, what might explain the unusually high prevalence of EM lineages in the Asian tropics? Any understanding of why the EM pool is so large in Asia must reckon with the radiation of the Dipterocarpoideae (Dipterocarpaceae), which numbers more than 470 species (Ashton 2002). More generally, many hypotheses have been put forward to explain the disparities in tree species diversity among different tropical regions. However, there is no consensus explanation (Couvreur 2014). Some authors emphasize the relative importance of extrinsic factors, such as climate or climatic stability (Couvreur 2014), whereas others highlight intrinsic factors, such as history of diversification of different lineages in different regions, which have resulted in, for example, a diverse understory tree flora only in South America (Gentry 1993; Terborgh et al. 2016). Another potentially important intrinsic factor is EM association and its influence on the origin and maintenance of tropical tree diversity, as outlined here. Clearly, more research is needed to understand EM influences on patterns of tree monodominance and how this relates to the community assembly and diversity of regional tree floras.

Conclusions

To our knowledge, this is the first study to develop a hypothesis to explain why alternative stable states of either local EM monodominance or relative rarity arise in some tropical regions and not others. Specifically, our hypothesis is that an increase in the size of the EM tree species pool results in an increase in the complexity of biotic interactions, which in turn prevents monodominance by a single EM species. We have proposed that under a diverse EM host species pool, tree species maintain divergence in traits because of the complex plant-soil feedbacks among the many different tree species, with no species achieving a monodominant status. In contrast, under a depauperate EM host species pool, tree species experience strong selection for evolution toward the optimal abiotic conditions, allowing them to become competitive enough to form monodominant stands via positive plant-soil feedback that causes priority effects. Overall, this study highlights the importance of geographical variation in species pools in understanding the conditions under which
community divergence results in vastly different alternative stable states.

Returning to the general topic of convergence, our goal in this article has been to point out the utility of studying the link between trait convergence and community convergence. To this end, we have explored how trait convergence may affect community convergence, using the case of tropical monodominance as an illustrative example. Although we have developed only one hypothesis, one thing that is clear from our work is that the natural historical knowledge of how convergent traits influence species interactions is essential to understanding community convergence.

Acknowledgments

We thank A. A. Agrawal for the invitation to contribute this article to the Vice Presidential Symposium of the American Society of Naturalists held in Austin, Texas, in June 2016. Comments from A. A. Agrawal, P.-J. Ke, and three anonymous reviewers improved the article. T.F. thanks the National Science Foundation (NSF; DEB 1555786) and the Terman Fellowship at Stanford University for financial support as well as the Center for Macroecology, Evolution, and Climate and the Section of Microbiology at the University of Copenhagen for sabbatical support. K.G.P. was supported in part by the NSF (DEB 1249342 and RAPID 1361171) and a Department of Energy Early Career Grant (DE-SC0016097). S.E.R. was supported in part by the NSF (RAPID 1361171). Author participation: T.F. and K.G.P. designed the study; M.N. conducted and analyzed simulation in collaboration with T.F.; C.B., P.V.A.F., and C.F. collected data from Peru; S.E.R. collected data from Peru; K.G.P. and S.E.R. analyzed data from Peru and Borneo; T.F. wrote the first draft of the manuscript; and all authors contributed to writing the manuscript.

APPENDIX

Methods for Quantifying Leaf Functional Traits in Bornean Rain Forest

Study System and Species Selection

Lambir Hills National Park (Lambir) is located in the northwestern part of Borneo in the Malaysian state of Sarawak (4°20’N, 113°50’E). Lambir receives approximately 3,000 mm of rainfall annually, with all months averaging >100 mm (Watson 1985). The region has the highest tree species richness recorded in the Paleotropics (Ashton 2005), with species in the Dipterocarpaceae dominating the forest (Lee et al. 2002). The soils and geomorphology of Lambir have been previously described (Bailie et al. 2006; Tan et al. 2009). The soils range from coarse loams that are sandstone derived, leached, nutrient depleted, and well drained with substantial raw humus, to clays that are shale derived, less nutrient depleted, and less well drained, with little raw humus. Tree community composition and demography vary across soil types, with most species exhibiting soil habitat specialization and tree demography varying across soil types (Davies et al. 2005; Russo et al. 2005). Lambir is the site of a 52-ha research plot that was established in 1991 as part of the Center for Tropical Forest Science Forest Global Earth Observatory plot network (Anderson-Teixeira et al. 2015) to monitor woody plants. All trees ≥ 1 cm in diameter at breast height are tagged, mapped, and dentified, and their diameters are measured to the nearest 1 mm.

Tree species were sampled for the quantification of leaf functional traits from within and near the Lambir plot (Russo et al. 2013). The species identity of individuals sampled outside of the plot was verified using an on-site herbarium and by consultation with local botanists (S. Tan). Species were selected to encompass a wide range of families—targeting taxa contributing substantially to forest basal area in the Lambir plot—to target species-rich genera, such as Shorea (Dipterocarpaceae) and Diospyros (Ebenaceae), and to include species with a range of shade tolerance niches. For each species, juvenile (1–5 cm in diameter) and adult (>10 cm in diameter) trees were sampled as much as possible, with one to 22 individuals sampled per species.

Quantification of Leaf Functional Traits

From each tree, three to five mature, sunlit, minimally damaged leaves were harvested. The petiole was cut from the lamina, which was gently cleaned of debris. Fresh leaf laminas were scanned (Canon LiDE 110), and the images were analyzed with ImageJ software (Schneider et al. 2012) to estimate the area of each. After oven drying at 60°C for 3 days, the dry weight of each lamina was recorded. The SLA (cm²/g) was calculated as fresh area divided by dry weight, and SLA was averaged across leaves of each individual. After drying, the midvein was removed from each lamina. Laminas from each individual were ground together to a fine, uniform powder and analyzed by elemental combustion for carbon and nitrogen content (Costech Analytical Elemental Combustion System 4010). For each individual, percent carbon (C) and nitrogen (N) were calculated as the mass of C (or N) in the sample/dry mass of the sample × 100, and C:N ratio was calculated as percent C/percent N. Trait values were averaged across individuals to obtain mean values for each tree species.

Literature Cited


This content downloaded from 132.236.027.111 on July 22, 2017 05:14:57 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).


References Cited Only in the Online Appendixes


Symposium Editor: Anurag A. Agrawal