

Using Integrative Biology to Infer Adaptation from Comparisons of Two (or a Few) Species

Christian L. Cox^{1,*}

Michael L. Logan²

¹Institute for the Environment and Department of Biological Sciences, Florida International University, Miami, Florida;

²Department of Biology, University of Nevada Reno, Reno, Nevada

Accepted 12/26/2020; Electronically Published 4/6/2021

ABSTRACT

Phylogenetic comparative methods represent a major advance in integrative and comparative biology and have allowed researchers to rigorously test for adaptation in a macroevolutionary framework. However, phylogenetic comparative methods require trait data for many species, which is impractical for certain taxonomic groups and trait types. We propose that the philosophical principle of severity can be implemented in an integrative framework to generate strong inference of adaptation in studies that compare only a few populations or species. This approach requires (1) ensuring that the study system contains species that are relatively closely related; (2) formulating a specific, clear, overarching hypothesis that can be subjected to integrative testing across levels of biological organization (e.g., ecology, behavior, morphology, physiology, and genetics); (3) collecting data that avoid statistical underdetermination and thus allow severe tests of hypotheses; and (4) systematically refining and refuting alternative hypotheses. Although difficult to collect for more than a few species, detailed, integrative data can be used to differentiate among several potential agents of selection. In this way, integrative studies of small numbers of closely related species can complement and even improve on broadscale phylogenetic comparative studies by revealing the specific drivers of adaptation.

Keywords: adaptation, integrative biology, severe testing, strong inference, phylogenetic comparative methods.

The Problem with Comparisons of Two (or a Few) Species

Over the past 40 years, there has been an increasing focus on statistical rigor in evolutionary biology (Clutton-Brock and Harvey 1979; Arnold 1983; Mayr 1983; Garland et al. 1992; Garland and Adolph 1994). New analytical tools and techniques have substantial power to leverage biological diversity to gain insight into the evolution of organisms and functional traits. Research in comparative evolutionary biology relies on the collection of data from different species, but species are not statistically independent entities. Rather, all species on Earth are connected by an evolutionary tree of life, and phylogenetic relationships can structure trait differences among closely related taxa.

Beginning with Felsenstein (1985) and others (Grafen 1989; Maddison 1990; Harvey and Pagel 1991; Garland et al. 1992), a variety of statistical tools have been developed that use the statistical nonindependence of species on a phylogeny to account for evolutionary relationships during comparative analyses. As an early part of the movement to drive adoption of statistical phylogenetic comparative methods in comparative physiology, Garland and Adolph (1994) published a seminal paper titled “Why Not to Do Two-Species Comparative Studies: Limitations for Inferring Adaptation.” The central argument of this paper was that comparisons of two species (or two populations) have limited inference because any difference between a pair of species could exist for many reasons and is not limited to the hypothesized driver of the difference. Importantly, this same criticism can be applied to any study that has few enough species that phylogenetic comparative analyses are inappropriate. However, this critique does not apply to studies that are not attempting to identify the specific cause of differences among taxa (Sanford et al. 2002), such as descriptive studies that demonstrate variation or similarity in either traits or functional mechanisms across taxonomic groups and do not invoke adaptation.

Regardless, Garland and Adolph’s (1994) critique was an important corrective to a glut of physiological research that contained two-species comparisons and improperly attempted to infer adaptation from differences between the species. However, we suggest that there are contexts where comparisons of two (or a few) species can be used to gain meaningful insight into adaptation. Indeed, sometimes studies of fewer species can result in more powerful inference into the role of adaptation than can phylogenetic comparative methods, especially in fields like comparative physiology, where it is often difficult to collect data on many species.

*Corresponding author; email: ccox@fiu.edu.

Common Methods for Inferring Adaptation

The definition of adaptation and methods of inferring adaptation have been a subject of great interest and controversy in organismal and evolutionary biology for over half a century (Ghiselin 1966; Williams 1966; Clutton-Brock and Harvey 1979; Gould and Lewontin 1979; Mayr 1983; Harvey and Pagel 1991; Rezende and Diniz-Filho 2011; Olson and Arroyo-Santos 2015). Although our goal is not to go into detail regarding all methods of inferring adaptation—and not all methods of inference are relevant to the fields (e.g., comparative physiology) that have traditionally suffered from the two-species comparison problem—we will briefly highlight some of the major methods of inference.

First, adaptation can be measured directly as shifts in the frequencies of alleles that are linked to fitness or a fitness proxy (Geffeney et al. 2002; Nachman et al. 2003; Balanyá et al. 2006; Hoekstra et al. 2006) by detecting the signature of selection in DNA (e.g., ratios of synonymous to nonsynonymous mutations; Tajima 1989; McDonald and Kreitman 1991) or through multivariate regression techniques that correlate some measure of fitness with phenotypic trait values (Lande and Arnold 1983; Arnold and Wade 1984; Bronikowski 2000). Second, adaptation can be inferred indirectly using phylogenetic comparative methods that identify correlations between traits and other properties of organisms or their environments while accounting for phylogenetic nonindependence (Felsenstein 1985; Harvey and Pagel 1991; Garland et al. 1992; Garland and Adolph 1994; Rezende and Diniz-Filho 2011). Put another way, phylogenetic comparative methods identify co-occurring evolutionary transitions in traits to infer adaptation. While each of these approaches can be powerful, they may be difficult to implement or irrelevant for certain study systems. For example, genetic markers that are linked to traits of interest are available for only a few organisms and traits, making genetic approaches powerful but less applicable across the tree of life. While genomes of nonmodel organisms are increasingly available, linking genotypes to phenotypes requires large sample sizes and substantial effort (Korte and Farlow 2013; Narum et al. 2013). Similarly, mark-recapture methods are often required to estimate fitness for multivariate regression approaches to quantifying selection in nature. But marking and recapturing may be impractical for small-bodied, secretive, highly mobile, or long-lived organisms. Finally, for both multivariate regression and phylogenetic comparative methods, there may be lurking variables that drive fitness-trait correlations (for multivariate regression) or evolutionary correlations (for phylogenetic comparative methods). This combination of challenges to the implementation of each of these methods highlights that there is no single method available that can perfectly isolate the exact evolutionary cause of trait evolution in all taxa (Olson and Arroyo-Santos 2015).

Nevertheless, there has been a recent focus on the application of phylogenetic comparative methods in integrative and comparative biology to infer adaptation in systems where direct measurements of adaptation are impractical (Losos 2011a; Ord

and Summers 2015). These methods have been a great boon to comparative physiologists and have led to substantial insights into the evolution and adaptation of functional traits (Cooper and Purvis 2010; Rabosky et al. 2013; Schenk 2013; Uyeda et al. 2017). Importantly, the advent of phylogenetic comparative analyses supplied a new tool that could allow inference of adaptation without field studies of survival and reproduction or detailed knowledge regarding the genetics of focal traits. Before the development of phylogenetic methods, comparative studies of adaptation could be statistically underdetermined (data could not distinguish between alternative hypotheses) and subject to tautological reasoning (famously critiqued by Gould and Lewontin 1979; also see Olson and Arroyo-Santos 2015).

Despite the many advantages of phylogenetic approaches for revealing important agents of selection, these analyses employ broad patterns of biological diversity to gain insight into adaptation and thus require both trait data and a phylogeny composed of many species. Although the sample size required for phylogenetic comparative analyses will vary according to tree characteristics, trait variation, and type of analysis, it is safe to say that sample sizes in the dozens to hundreds are generally needed (Ackerly 2000; Maddison et al. 2007; Heath et al. 2008a, 2008b; Boettiger et al. 2012). Yet there is a fundamental tension between the number of species on which one can collect data and the depth and richness of data collection that is possible for each species. Indeed, many phylogenetic comparative analyses focus on easily studied or commonly measured traits, such as body size (Cooper and Purvis 2010; Kahrl et al. 2016), physiological traits that are commonly measured (Cox and Cox 2015; Uyeda et al. 2017), color pattern (Penney et al. 2012; Davis Rabosky et al. 2016), and antipredator defenses (Arbuckle and Speed 2015; Stankowich and Campbell 2016). However, there are many physiological and functional traits that have not been the frequent focus of study for decades (necessary to build a comparative sample of data in the literature) and require substantial amounts of time, effort, and money to measure across a broad range of species. For example, microbiome and RNA sequencing data are increasingly important for understanding adaptation (Wood and Stinchcombe 2017; Rudman et al. 2019), yet both types of data require high-throughput sequencing at significant cost and effort. These costs make generating comparative data for dozens of species in the typical time frame for scientific studies nearly impossible in most laboratories. Similarly, some lineages have low extant diversity (e.g., ratite birds and coelacanths), such that sampling many species would be impossible.

Given the intrinsic drawbacks of phylogenetic comparative methods, should comparative biologists limit themselves to species-rich clades and easily measured traits as the only options to infer adaptation? We argue that there is scope for inferring adaptation from studies of two or a few species if researchers use an integrative approach that leverages multiple independent lines of evidence to make strong inferences about the potential role of selection in driving trait differences among taxa. Thus, comparative studies of small numbers of species should not be dismissed out of hand.

The Severity Principle and Strong Inference in Integrative Biology

We suggest that the philosophy of scientific inference can be deployed to infer adaptation from species-depauperate but data-rich studies. There are several principles that can frame an integrative approach to inferring adaptation. First, the concept of strong inference (Platt 1964) suggests that significant advances in science are made in research programs that sequentially and systematically test all feasible alternative hypotheses (box 1). Platt (1964) emphasizes the important role of experiments, especially those that have clear or even binary outcomes regarding the likelihood that a hypothesis is false. A research program that is guided by strong inference thus follows a process of systematic testing of hypotheses with further refinement of those hypotheses in light of new knowledge. The principle of severity (Mayo 1997; Mayo and Spanos 2010, 2011) similarly focuses on making “severe” tests of hypotheses. A hypothesis is considered to have passed a severe test when the test would have produced a result substantially less consistent with the hypothesis if the hypothesis was false or incorrect (box 1). In this context, a severe test does not refer to a specific statistical test but rather emphasizes an approach to data collection that permits hypothesis tests that are not statistically underdetermined (fig. 1). Indeed, the principles of strong inference and severe testing are consistent with any method of evaluating evidence against hypotheses, including Bayesian statistics, frequentist approaches, and model-based inference using information theoretics (e.g., Akaike information criteria).

Research programs in integrative biology, which focus on assimilating techniques and data from diverse fields of study, have long engaged in severe tests of hypotheses (Brodie and Brodie 1990; Karasov and Levey 1990; Hutchinson et al. 2007; Ott and Secor 2007). However, the value of this approach is sometimes appreciated only for single-species studies, while studies comparing two or a few species can be considered problematic. We contend that there is no nadir of inference power centered on two-species comparisons and that an integrative

approach employing severe tests to make strong inferences can be deployed to study adaptation when comparing only two or a few species.

Inferring Adaptation through Integration

We assert that the principles of strong inference and severe testing can be deployed within an integrative framework to infer adaptation. This approach to study design can be used to infer adaptation on multispecies data sets that have too few species for phylogenetic comparative methods and requires (1) ensuring that the study system contains species that are relatively closely related; (2) formulating a specific, clear, overarching hypothesis that can be subjected to testing across the biological hierarchy (e.g., ecology, behavior, morphology, physiology, and genetics); (3) collecting data that avoid statistical underdetermination and thus allow severe tests of hypotheses; and (4) systematically refining and refuting alternative hypotheses. We elaborate on each of these points below.

First, the study system should contain species that are relatively closely related, as comparisons that span huge gaps in evolutionary time will limit inference. This is because observed similarities or differences cannot be attributed to any single factor in two species that have been separated by millions of years of evolution (Rezende and Diniz-Filho 2011). For example, early comparative biology is peppered with comparisons of species that have been diverged for many millions of years, such as mammalian ungulates and carnivores or birds and mammals (Bartholomew et al. 1968; Nielsen et al. 1970; Schmidt-Nielsen 1972; Schmidt-Taylor and Weibel 1981; Roberts et al. 1996). Because the evolutionary distances between these taxa pairs are so vast, nonadaptive processes like phylogenetic and developmental constraints might be more important than adaptation in shaping functional relationships among traits (Rezende and Diniz-Filho 2011). Using closely related species minimizes differences and facilitates tests of adaptation on those traits that do differ among species.

Box 1

Statistical philosophies that can guide comparative studies of adaptation

Strong inference. Strong inference was first described by Platt (1964) as a strict application of inductive reasoning to the scientific method. Strong inference relies on a sequential procedure of (1) devising alternative hypotheses, (2) testing these hypotheses using studies that exclude one or more hypotheses, and (3) iteratively refining and testing additional alternative hypotheses that follow from steps 1 and 2.

Severity principle. The severity principle was developed by Deborah Mayo and colleagues (Mayo 1996, 1997; Mayo and Spanos 2010, 2011) as part of the broader philosophy of error statistics. According to Mayo and Spanos (2011, p. 164), “A hypothesis H passes a severe test T with data x_0 if, (S-1) x_0 accords with H , (with a suitable notion of accordance) and (S-2) with very high probability, test T would have produced a result that accords less well with H than does, if H were false or incorrect.” The terms (S-1) and (S-2) are conditions of the severe test. In other words, a hypothesis is considered to have passed a severe test when the test would have produced a result that would be substantially less consistent with the hypothesis if the hypothesis was false or incorrect. The key component of the severity principle is careful design of studies to avoid collecting data that lead to statistically underdetermined tests of hypotheses.

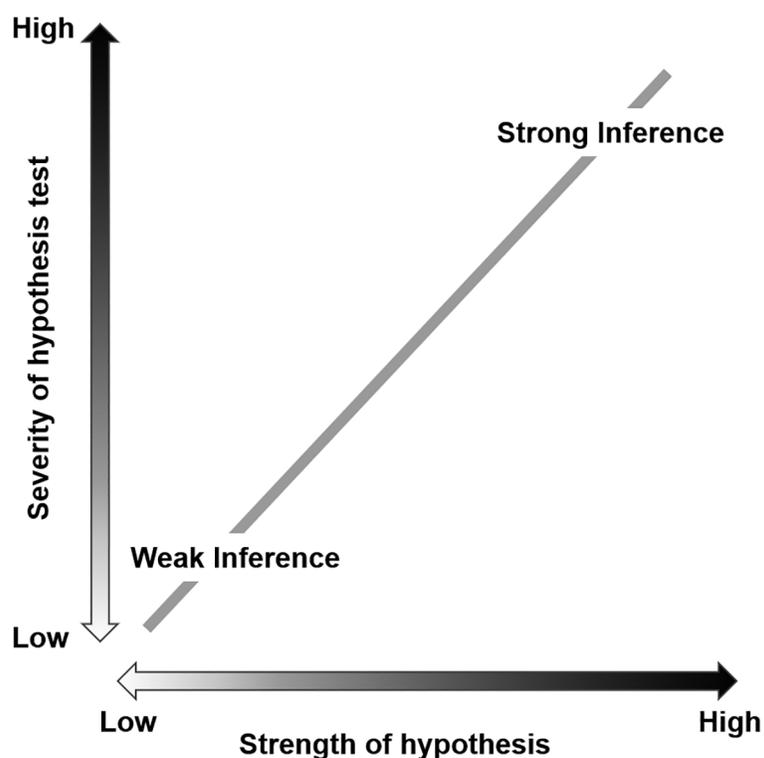
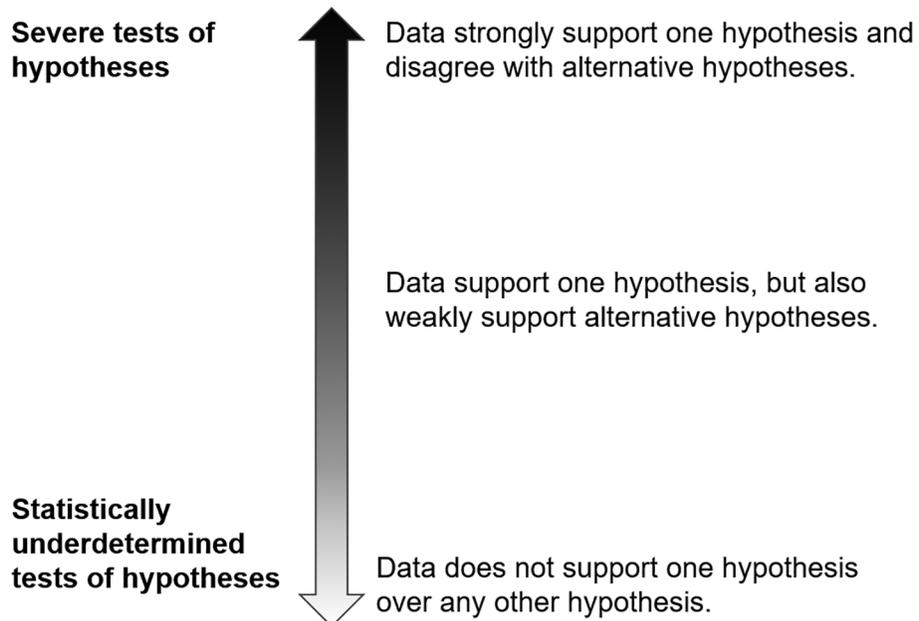
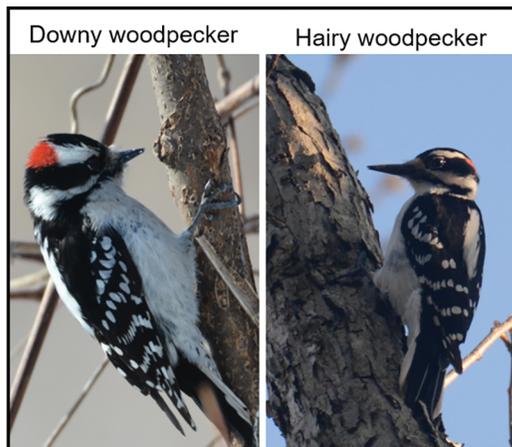


Figure 1. *Top*, severity of hypothesis testing falls along a continuum from statistically underdetermined to extremely severe. *Bottom*, strong inference is generated by combining strong hypotheses with severe testing of those hypotheses.

Second, the hypothesis for the agent of selection should be specific and, if possible, make testable predictions across different levels of the biological hierarchy or trait modules (fig. 2). For example, any hypothesis stating that a putative agent of selection will cause variation among species in a single trait would be stronger if the hypothesis could also predict variation

among species in multiple trait types (e.g., morphology, behavior, physiology, and genetics).

Third, studies should attempt to make severe tests of hypotheses by collecting data that have the potential to be more consistent with one hypothesis while at least partially refuting another. For example, testing whether a trait differs between



System and Research Question: Downy and hairy woodpeckers are closely-related species of insectivorous birds that co-occur throughout North American woodlands. Why do these species have different beak sizes?

Weak Hypothesis: Downy and hairy woodpeckers differ in beak size, which must be adaptive because of ecological difference between species.

Strong Hypothesis: Differences in beak size arose from adaptation to the biomechanics of feeding in different microhabitats.

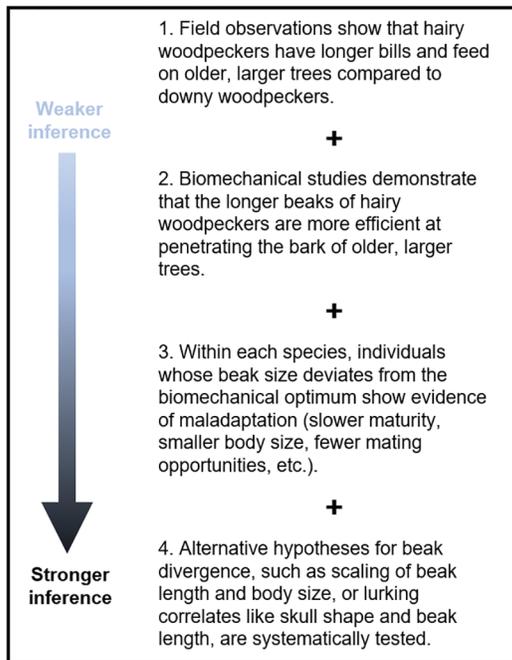


Figure 2. Example of an integrative approach to studying adaptation in comparisons of only two species. Photos by John David Curlis.

species is not a severe test, as there are multiple generative mechanisms that could cause species to differ in trait values. Rather, information about the direction of the trait difference, how that difference is linked to specific environmental factors, and whether the trait difference is linked to physiological performance within each species would strengthen the analysis

into a severe test. The key component of a severe test is that the data gathered will be more consistent with one hypothesis and less consistent with an alternative hypothesis. However, there is no a priori threshold whereby a hypothesis test suddenly becomes severe (fig. 1). Rather, scientists should strive to design studies that conduct hypothesis tests that are as severe as possible given logistical constraints.

Fourth, this integrative approach relies on generating multiple lines of evidence to evaluate alternative hypotheses on the role of adaptation in the divergence of a given trait. In this way, integrative studies can carefully test and discard alternative hypotheses in an iterative fashion to build evidence for adaptation. Yet even when researchers leverage severe testing in an integrative framework to answer a scientific question, they may not produce a complete answer to the question in a single study. Instead, effective research programs can increase the severity and specificity of their hypothesis tests as results accumulate over multiple studies. It is worth noting that the kind of synthetic data that is necessary for an integrative approach would be difficult to collect for more than a few species in a reasonable time frame. Thus, this approach is best suited to in-depth studies of two (or a few) species where detailed field and laboratory data collection are possible.

Consider a hypothetical example of the integrative approach to inferring morphological adaptation—the evolution of beak size in woodpeckers. We use this as an illustrative example that is only loosely rooted in real observations about these species, and the following discussion should not be interpreted as an actual analysis of the evolution of beak size in this system. Rather, we have chosen some familiar species and ideas to demonstrate the application of an integrative approach. Hairy woodpeckers (*Dryobates villosus*) and downy woodpeckers (*Dryobates pubescens*) are closely related species of insectivorous birds with sympatric distributions in similar habitats, and they are strikingly similar in morphology (fig. 2). However, these species differ in several morphological traits, including beak length, which is longer in hairy woodpeckers. How should a biologist study beak evolution in this situation? First, the species must be reasonably closely related to apply the integrative approach. Downy and hairy woodpeckers are congeners or are at least in the same clade (Shakya et al. 2017; Miles et al. 2018) and are thus sufficiently closely related so that a comparison is reasonable.

Second, the hypotheses about adaptation of beak size should be clear and specific. As a somewhat obvious example, simply inferring adaptation by testing whether beak length differs between the species would represent a test of a weak hypothesis because beak length might differ between these species for any number of reasons. Refining the hypothesis by including ecological drivers of beak adaptation (e.g., hairy woodpeckers feed on older trees with thicker bark compared with downy woodpeckers) and the direction of the adaptive difference (e.g., hairy woodpeckers have longer beaks that are more effective for feeding on thicker bark than downy woodpeckers) would strengthen the hypothesis. This specificity can be expanded to include predictions about other trait modules occurring at different levels of biological

organization. For example, the hypothesis that beak size is adapted to bark thickness also makes predictions about the energetic costs and biomechanics of the relationship between beak size and bark thickness that can be tested. Third, experimental tests should be constructed to be severe by collecting data that can favor one hypothesis and disfavor another. In the case of biomechanical tests of feeding in the laboratory, one could test whether long beaks are more energetically efficient than short beaks at penetrating the thicker bark of older, larger trees. If longer beaks do not have an energetic advantage at penetrating thicker bark or are even less efficient than short beaks at penetrating thicker bark, then the hypothesis that longer beaks are an adaptation to thicker bark is at least partially refuted.

Finally, while strong hypotheses and severe tests are important for strong inference, it is also crucial to systematically test (and discard) alternative hypotheses using the same principle of severity. Perhaps beak size is longer in hairy woodpeckers, and they indeed feed on older trees with thicker bark. However, hairy woodpeckers are also slightly larger than downy woodpeckers, and beak length might simply scale with body size and have nothing to do with the feeding environment. It would thus be important to include body size as a covariate in models that test for a relationship between beak length and microhabitat type. In a less simplistic example, perhaps beak length is strongly associated with feeding microhabitats. However, beak length may simply be correlated with tongue length or be an inevitable result of the scaling relationship between beak length and skull shape. In this case, it would be worth testing the alternative hypothesis that it is skull shape or tongue length that is adapted to older, thicker trees rather than beak length per se. Ultimately, iterative rounds of severe testing followed by refinement of hypotheses would lead to strong inferences about the agents of selection underlying beak morphology in woodpeckers.

The utility of this integrative approach for inferring adaptation is not limited to comparative physiology and is applicable to other fields, such as behavioral ecology, that are interested in understanding the causes of adaptation. Furthermore, this approach is not even limited to evolutionary biology more generally and in fact could be used to test any hypothesis about the causes of differences between taxa, including causes (e.g., ecological) that have nothing to do with adaptation per se. Regardless of the specific questions that are being investigated, carefully designed studies can still produce results to test multiple, perhaps even mutually exclusive, hypotheses. In these cases, strong inference can be achieved through the refinement of hypotheses in the light of new data. This stepwise process of refining, severely testing, and discarding alternative hypotheses should inexorably lead to strong inference.

By their nature, phylogenetic comparative studies are correlative because only a few traits or environmental variables can reasonably be measured across dozens or hundreds of species. The kind of detailed, integrative data that is necessary to rule out lurking variables is difficult to collect for more than a few species. Even in the rare cases where deep data sets are available for a broad range of taxa, these data are usually collected by different research groups using a variety of methods, reducing the reproducibility and reliability of results (McKechnie and Wolf 2004;

Genoud et al. 2018). Thus, integrative studies of small numbers of closely related species can even have advantages over phylogenetic comparative studies in that they may be more effective at inferring the specific agents of selection that drive adaptation and trait divergence.

Phylogenetic Comparative Methods, Integrative Biology, and Adaptation

This perspective should not be misconstrued as a broad critique of phylogenetic comparative methods or a call for relaxing standards in comparative biology. The central message of Felsenstein (1985) and Garland and Adolph (1994) is that branching relationships among evolutionary trees can cause errors that can mislead comparative analyses, and this message is still important. We agree that whenever there are a sufficient number of taxa, phylogenetic comparative analyses should generally be employed to infer adaptation. A return to the “just so” adaptationist storytelling of yesteryear, driven by superficially “integrative” two-species comparisons would be a counterproductive outcome of the shift in perspective we are advocating here. However, we think that a truly integrative approach that leverages multiple data types to conduct severe tests of hypotheses and generate strong inference represents another avenue by which adaptation can be inferred when phylogenetic comparative methods are not possible and has its own set of advantages. Moreover, integrative and phylogenetic studies are not necessarily mutually exclusive and, in some contexts, may complement each other. For example, phylogenetic studies can be used to generate hypotheses that can then be investigated via deep, integrative comparisons of fewer species. The combination of phylogenetic perspectives with detailed studies on a few species or populations has led to substantial insights into some of the most familiar and important examples of evolutionary adaptation in nature, such as adaptation of snakes to toxic prey (Brodie and Brodie 1990; Geffeney et al. 2002; Feldman et al. 2009), adaptation of mouse pelage to habitat variation (Nachman et al. 2003; Hoekstra et al. 2005, 2006), and adaptation of *Anolis* lizards to extreme thermal environments (Campbell-Staton et al. 2012, 2016, 2017, 2020).

Despite their advantages, comparative studies containing only two or a few species are often criticized without consideration of the fact that they often leverage rich, integrative data sets for inferring adaptation. We are certainly not the first to consider the costs and benefits of different methods of studying adaptation (Olson and Arroyo-Santos 2015) or to consider the limitations of phylogenetic comparative methods for studying adaptation (Losos 1999, 2011b; Uyeda et al. 2018). Indeed, elements of the basic approach we are highlighting were mentioned in Garland and Adolph (1994; see the section “Enhancing Two-Species Comparisons through a Multivariate Approach” and the last two paragraphs of the paper) when these authors originally noted the potential limitations of two-species studies. However, we suggest that the principles of strong inference and severity illuminate the advantages of comparisons of small numbers of species in some contexts. Given the trade-off between data depth and breadth for any given taxonomic group,

an integrative approach will continue to be useful in the future as a legitimate avenue for inferring adaptation.

Acknowledgments

The development of our thoughts around the application of the severity principle and strong inference for inferring adaptation in integrative biology was influenced by two posts on the *Dynamic Ecology* blog (Fox 2012, 2016). We thank John Schenk for thoughtful discussions about adaptation and phylogenetic comparative methods and critical comments on an earlier version of the manuscript.

Literature Cited

- Ackerly D.D. 2000. Taxon sampling, correlated evolution, and independent contrasts. *Evolution* 54:1480–1492.
- Arbuckle K. and M.P. Speed. 2015. Antipredator defenses predict diversification rates. *Proc Natl Acad Sci USA* 112:13597–13602.
- Arnold S.J. 1983. Morphology, performance and fitness. *Am Zool* 23:347–361.
- Arnold S.J. and M.J. Wade. 1984. On the measurement of natural and sexual selection: theory. *Evolution* 38:709–719.
- Balanyá J., J.M. Oller, R.B. Huey, G.W. Gilchrist, and L. Serra. 2006. Global genetic change tracks global climate warming in *Drosophila subobscura*. *Science* 313:1773–1775.
- Bartholomew G.A., R.C. Lasiewski, and E.C. Crawford. 1968. Patterns of panting and gular flutter in cormorants, pelicans, owls, and doves. *Condor* 70:31–34.
- Boettiger C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? measuring the power of comparative methods. *Evolution* 66:2240–2251.
- Brodie E.D., III, and E.D. Brodie Jr. 1990. Tetrodotoxin resistance in garter snakes: an evolutionary response of predators to dangerous prey. *Evolution* 44:651–659.
- Bronikowski A.M. 2000. Experimental evidence for the adaptive evolution of growth rate in the garter snake *Thamnophis elegans*. *Evolution* 54:1760–1767.
- Campbell-Staton S.C., Z.A. Cheviron, N. Rochette, J. Catchen, J.B. Losos, and S.V. Edwards. 2017. Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. *Science* 357:495–498.
- Campbell-Staton S.C., S. Edwards, and J. Losos. 2016. Climate-mediated adaptation after mainland colonization of an ancestrally subtropical island lizard, *Anolis carolinensis*. *J Evol Biol* 29:2168–2180.
- Campbell-Staton S.C., R.M. Goodman, N. Backström, S.V. Edwards, J.B. Losos, and J.J. Kolbe. 2012. Out of Florida: mtDNA reveals patterns of migration and Pleistocene range expansion of the green anole lizard (*Anolis carolinensis*). *Ecol Evol* 2:2274–2284.
- Campbell-Staton S.C., K.M. Winchell, N.C. Rochette, J. Fredette, I. Maayan, R.M. Schweizer, and J. Catchen. 2020. Parallel selection on thermal physiology facilitates repeated adaptation of city lizards to urban heat islands. *Nat Ecol Evol* 4:652–658.
- Clutton-Brock T.H. and P.H. Harvey. 1979. Comparison and adaptation. *Proc R Soc B* 205:547–565.
- Cooper N. and A. Purvis. 2010. Body size evolution in mammals: complexity in tempo and mode. *Am Nat* 175:727–738.
- Cox C.L. and R.M. Cox. 2015. Evolutionary shifts in habitat aridity predict evaporative water loss across squamate reptiles. *Evolution* 69:2507–2516.
- Davis Rabosky A.R., C.L. Cox, D.L. Rabosky, P.O. Title, I.A. Holmes, A. Feldman, and J.A. McGuire. 2016. Coral snakes predict the evolution of mimicry across New World snakes. *Nat Commun* 7:11484.
- Feldman C.R., E.D. Brodie, and M.E. Pfrender. 2009. The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. *Proc Natl Acad Sci USA* 106:13415–13420.
- Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat* 125:1–15.
- Fox J. 2012. Why, and how, to do statistics (it's probably not why and how you think). *Dynamic ecology* (blog). November 28, 2012. <https://dynamicecology.wordpress.com/2012/11/28/why-and-how-to-do-statistics-its-probably-not-why-and-how-you-think/>.
- . 2016. Why don't more ecologists use strong inference? *Dynamic ecology* (blog). June 1, 2016. <https://dynamicecology.wordpress.com/2016/06/01/obstacles-to-strong-inference-in-ecology/>.
- Garland T., Jr., and S.C. Adolph. 1994. Why not to do two-species comparative studies: limitations on inferring adaptation. *Physiol Zool* 67:797–828.
- Garland T., Jr., P.H. Harvey, and A.R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. *Syst Biol* 41:18–32.
- Geffeney S., E.D. Brodie, and P.C. Ruben. 2002. Mechanisms of adaptation in a predator-prey arms race: TTX-resistant sodium channels. *Science* 297:1336–1339.
- Genoud M., K. Isler, and R.D. Martin. 2018. Comparative analyses of basal rate of metabolism in mammals: data selection does matter. *Biol Rev* 93:404–438.
- Ghiselin M.T. 1966. On semantic pitfalls of biological adaptation. *Philos Sci* 33:147–153.
- Gould S.J. and R.C. Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc B* 205:581–598.
- Grafen A. 1989. The phylogenetic regression. *Philos Trans R Soc B* 326:119–157.
- Harvey P.H. and M.D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford University Press, Oxford.
- Heath T.A., S.M. Hedtke, and D.M. Hillis. 2008a. Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 46:239–257.
- Heath T.A., D.J. Zwickl, J. Kim, and D.M. Hillis. 2008b. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst Biol* 57:160–166.
- Hoekstra H.E., R.J. Hirschmann, R.A. Bunday, P.A. Insel, and J.P. Crossland. 2006. A single amino acid mutation

- contributes to adaptive beach mouse color pattern. *Science* 313:101–104.
- Hoekstra H.E., J. Krenz, and M.W. Nachman. 2005. Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity* 94:217–228.
- Hutchinson D.A., A. Mori, A.H. Savitzky, G.M. Burghardt, X. Wu, J. Meinwald, and F.C. Schroeder. 2007. Dietary sequestration of defensive steroids in nuchal glands of the Asian snake *Rhabdophis tigrinus*. *Proc Natl Acad Sci USA* 104:2265–2270.
- Kahrl A.F., C.L. Cox, and R.M. Cox. 2016. Correlated evolution between targets of pre- and post-copulatory sexual selection across squamate reptiles. *Ecol Evol* 6:6452–6459.
- Karasov W.H. and D.J. Levey. 1990. Digestive system trade-offs and adaptations of frugivorous passerine birds. *Physiol Zool* 63:1248–1270.
- Korte A. and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods* 9:1–9.
- Lande R. and S.J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Losos J.B. 1999. Uncertainty in the reconstruction of ancestral character states and limitations on the use of phylogenetic comparative methods. *Anim Behav* 58:1319–1324.
- . 2011a. Convergence, adaptation, and constraint. *Evolution* 65:1827–1840.
- . 2011b. Seeing the forest for the trees: the limitations of phylogenies in comparative biology: (American Society of Naturalists address). *Am Nat* 177:709–727.
- Maddison W.P. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44:539–557.
- Maddison W.P., P.E. Midford, and S.P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. *Syst Biol* 56:701–710.
- Mayo, D.G. 1996. Error and the growth of experimental knowledge. University of Chicago Press, Chicago.
- . 1997. Severe tests, arguing from error, and methodological underdetermination. *Philos Stud* 86:243–266.
- Mayo D.G. and A. Spanos. 2010. Error and inference: recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science. Cambridge University Press, Cambridge.
- . 2011. Error statistics. Pp. 153–198 in P.S. Bandyopadhyay, ed. *Philosophy of statistics*. Elsevier, Amsterdam.
- Mayr E. 1983. How to carry out the adaptationist program? *Am Nat* 121:324–334.
- McDonald J.H. and M. Kreitman. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.
- McKechnie A.E. and B.O. Wolf. 2004. The allometry of avian basal metabolic rate: good predictions need good data. *Physiol Biochem Zool* 77:502–521.
- Miles M.C., E.R. Schuppe, R.M. Ligon IV, and M.J. Fuxjager. 2018. Macroevolutionary patterning of woodpecker drums reveals how sexual selection elaborates signals under constraint. *Proc R Soc B* 285:20172628.
- Nachman M.W., H.E. Hoekstra, and S.L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci USA* 100:5268–5273.
- Narum S.R., C.A. Buerkle, J.W. Davey, M.R. Miller, and P.A. Hohenlohe. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Mol Ecol* 22:2841.
- Olson M.E. and A. Arroyo-Santos. 2015. How to study adaptation (and why to do it that way). *Q Rev Biol* 90:167–191.
- Ord T.J. and T.C. Summers. 2015. Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evol Biol* 15:137.
- Ott B.D. and S.M. Secor. 2007. Adaptive regulation of digestive performance in the genus *Python*. *J Exp Biol* 210:340–356.
- Penney H.D., C. Hassall, J.H. Skevington, K.R. Abbott, and T.N. Sherratt. 2012. A comparative analysis of the evolution of imperfect mimicry. *Nature* 483:461.
- Platt J.R. 1964. Strong inference. *Science* 146:347–353.
- Rabosky D.L., F. Santini, J. Eastman, S.A. Smith, B. Sidlauskas, J. Chang, and M.E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat Commun* 4:1–8.
- Rezende E.L. and J.A.F. Diniz-Filho. 2011. Phylogenetic analyses: comparing species to infer adaptations and physiological mechanisms. *Comp Physiol* 2:639–674.
- Roberts T.J., J.-M. Weber, H. Hoppeler, E.R. Weibel, and C.R. Taylor. 1996. Design of the oxygen and substrate pathways. II. Defining the upper limits of carbohydrate and fat oxidation. *J Exp Biol* 199:1651–1658.
- Rudman S.M., S. Greenblum, R.C. Hughes, S. Rajpurohit, O. Kiratli, D.B. Lowder, S.G. Lemmon, D.A. Petrov, J.M. Chaston, and P. Schmidt. 2019. Microbiome composition shapes rapid genomic adaptation of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 116:20025–20032.
- Sanford G.M., W.I. Lutterschmidt, and V.H. Hutchison. 2002. The comparative method revisited. *Bioscience* 52:830–836.
- Schenk J.J. 2013. Evolution of limited seed dispersal ability on gypsum islands. *Am J Bot* 100:1811–1822.
- Schmidt-Nielsen K. 1972. Locomotion: energy cost of swimming, flying, and running. *Science* 177:222–228.
- Schmidt-Nielsen K., F.R. Hainsworth, and D.E. Murrish. 1970. Counter-current heat exchange in the respiratory passages: effect on water and heat balance. *Respir Physiol* 9:263–276.
- Shakya S.B., J. Fuchs, J.-M. Pons, and F.H. Sheldon. 2017. Tapping the woodpecker tree for evolutionary insight. *Mol Phylogenet Evol* 116:182–191.
- Stankowich T. and L.A. Campbell. 2016. Living in the danger zone: exposure to predators and the evolution of spines and body armor in mammals. *Evolution* 70:1501–1511.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Taylor C.R. and E.R. Weibel. 1981. Design of the mammalian respiratory system. I. Problem and strategy. *Respir Physiol* 44:1–10.

Uyeda J.C., M.W. Pennell, E.T. Miller, R. Maia, and C.R. McClain. 2017. The evolution of energetic scaling across the vertebrate tree of life. *Am Nat* 190:185–199.

Uyeda J.C., R. Zenil-Ferguson, and M.W. Pennell. 2018. Rethinking phylogenetic comparative methods. *Syst Biol* 67: 1091–1109.

Williams G. 1966. *Adaptation and natural selection*. Princeton University Press, Princeton, NJ.

Wood C.W. and J.R. Stinchcombe. 2017. A window into the transcriptomic basis of genotype-by-genotype interactions in the legume-rhizobia mutualism. *Mol Ecol* 26:5869–5871.