HISTORICAL COMMENT

Revisiting a Key Innovation in Evolutionary Biology: Felsenstein's "Phylogenies and the Comparative Method"

Raymond B. Huey,[1,*] Theodore Garland Jr.,[2] and Michael Turelli[3]

1. Department of Biology, University of Washington, Seattle, Washington 98195;   2. Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, California 92521;   3. Department of Evolution and Ecology, University of California, Davis, California 95616

ABSTRACT: The comparative method has long been a fundamental exploratory tool in evolutionary biology, but this venerable approach was revolutionized in 1985, when Felsenstein published "Phylogenies and the Comparative Method" in *The American Naturalist*. This article forced comparative biologists to start thinking phylogenetically when conducting statistical analyses of correlated trait evolution rather than simply applying conventional statistical methods that ignore evolutionary relationships. It did so by introducing a novel analytical method (phylogenetically "independent contrasts") that required a phylogenetic topology with branch lengths and that assumed a Brownian motion model of trait evolution. Independent contrasts enabled comparative biologists to avoid the statistical dilemma of nonindependence of species values, arising from shared ancestry, but came at the cost of needing a detailed phylogeny and of accepting a specific model of character change. Nevertheless, this article not only revitalized comparative biology but even encouraged studies aimed at estimating phylogenies. Felsenstein's characteristically lucid and concise statement of the problem (illustrated with powerful graphics), coupled with an oncoming flood of new molecular data and techniques for estimating phylogenies, led Felsenstein's 1985 article to become the second most cited article in the history of this journal. Here we present a personal review of comparative biology before, during, and after Joe's article. For historical context, we append a perspective written by Joe himself that describes how his article evolved, unedited transcripts of reviews of his submitted manuscript, and a guide to some nontrivial calculations. These additional materials help emphasize that the process of science does not always occur gradually or predictably.

Keywords: adaptation, allometry, comparative method, comparative physiology, phylogeny, statistics.

* Corresponding author; email: hueyrb@uw.edu.
   ORCIDs: Huey, https://orcid.org/0000-0002-4962-8670; Garland, https://orcid.org/0000-0002-7916-3552; Turelli, https://orcid.org/0000-0003-1188-9856.

This paper addresses a complex and important issue, and provides a solution to part of the problem—a very unsatisfactory solution, as the author is well aware, given the degree to which our data will usually fall short of the quality required by the method he proposes. . . . Nevertheless, as far as I can tell the method does what is claimed, and it is probably worth publishing. (From an anonymous review of the manuscript that became Felsenstein [1985*a*])

Introduction

Very few articles not only change the way that we think about biology but also change the way we gather and analyze data to study our favorite questions. Joe Felsenstein's 1985 article titled "Phylogenies and the Comparative Method" (Felsenstein 1985*a*) did both. His article laid out why business as usual in the comparative method was no longer appropriate, he showed graphically why phylogenies were fundamental to comparative biology, and he provided an analytical method that examines correlated evolution of traits on a tree while accounting for the covariances induced by shared phylogenetic history. In so doing, Felsenstein (1985*a*) rapidly and radically changed both evolutionary and organismal biology. Of course, the ideas and methods of his article have been debated, refined, extended, and supplemented over the decades, but its basic philosophy and fundamental message remain solid. No one would consider ignoring phylogeny when analyzing data involving multiple species (actually, we wish that were true, but it still happens, especially in some biomedical fields). Only a third of a century after being published, Joe's article has become the second most cited article in the history of *The American Naturalist* (fig. 1; see also Carvalho et al. 2005).

We had the good fortune to be friends and colleagues of Joe during the period that led to his 1985 article (and thereafter!). To honor Joe (fig. 2) and his 1985 article, we describe our personal recollections of comparative biology in the late 1970s
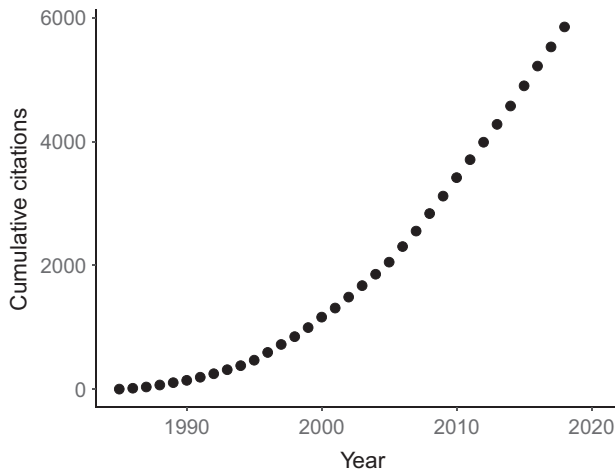
Figure 1: Felsenstein (1985*a*) is the second most cited article in this journal ($N = 5,825$). Only Nei (1972) currently has more citations (6,793). Paine (1966) is third (3,192). Data accessed from Web of Science (December 23, 2018).

and early 1980s, summarize the issues that inspired his article, recount the initial responses (not always positive!) of comparative biologists, and review some extensions of his approach. In the spirit of his article (and the original draft, which Joe kindly shared), we also discuss the assumptions that underlie his model-based solution to the problem of "correcting" for the phenotypic correlations induced by shared evolutionary history. For students, we provide a simple guide (app. C) to the two nontrivial calculations that Joe left as an "exercise" for the reader.

We are delighted that Joe contributed a perspective (app. A) describing how his article evolved. Also, he graciously shared copies of the initial manuscript reviews, which are fascinating reading (app. B). Both Joe's perspective and the appended reviews provide important historical insights on how science evolves.

Comparative Biology Shortly before Felsenstein (1985*a*)

Comparative methods are widely used in many fields, from anthropology to linguistics to zoology. In all of these fields, investigators gather data for some "traits" of multiple "groups," compute statistical relationships among them, and often attempt to infer the evolutionary processes underlying contemporary patterns. In biological sciences, the comparative method has long been the primary way of documenting the diversity of traits, quantifying their intercorrelations, and speculating on their evolution.

Darwin (1859, 1871) was, of course, a master of using interspecific differences to infer evolutionary patterns and processes. By the early twentieth century, statistical methods were being introduced, enabling quantitative evaluations of the correlated evolution of traits. By the mid-twentieth century, comparative biologists were commonly using comparative approaches to explore two generic themes: adaptation to the environment and scaling (how traits change with body size). However, they rarely used phylogenetic information either when selecting species to study or when analyzing their data (according to Huey 1987; Brooks and McLennan 1991; Harvey and Pagel 1991). Below we present examples of both themes from comparative physiology, but parallel examples are found in many biological disciplines (for a breakdown of the disciplines and organisms represented in articles that cite Joe's article, see Carvalho et al. 2005). These examples help circumscribe the world into which Felsenstein (1985*a*) emerged and also help to elucidate its revolutionary impact.

Adaptation to the Environment

A classical approach to studying adaptation involves comparing animals or plants that live in different environments (e.g., high vs. low altitude, deserts vs. rain forests), that have different capacities (e.g., sprinters vs. marathoners), or that have different behaviors (solitary vs. social; Clutton-Brock and Harvey 1977; Garland et al. 2005; Rezende and Diniz-Fihlo 2012). In comparative physiology, adaptation to extreme environments was a common theme (Bartholomew 1987). These studies often required complex physiological measurements, and so comparative physiologists would often study only two species, one from each environment (or with each capacity). In the Panglossian tradition of this era, physiologists would typically interpret any observed difference (or sometime even its absence; see Feder 1987; Huey and Berrigan 1996, p. 210) as evidence of adaptation.

Species selected for comparative studies were chosen for convenience (i.e., were locally available or readily purchased), because data for them were already available in the literature, because they were assumed to be exemplars of special adaptation (e.g., to extreme environments), or because they had some tractable physiological attribute (the August Krogh principle; Krebs 1975). Phylogenetic relationships were not an issue, and so species being compared were often distantly related (e.g., from different families or different orders), producing comparisons between apples and oranges (Huey 1987).

Many real examples exist for that period, but here we describe a semihypothetical example, thus protecting the guilty (see also Garland and Adolph 1994). A physiologist interested in elucidating cardiovascular adaptation to diving in mammals might choose to compare a muskrat (aquatic) and a guinea pig (terrestrial), which are similar in body size, and then be tempted to attribute any observed cardiovascular differences to diving adaptations, assuming implicitly that terrestrial life was ancestral. Other interpreta-

Figure 2: Photograph of Joe Felsenstein ("looking over a back fence, overly serious") in 1985. Photo credit: Eleanor Price (Joe's mother).

tions are, however, reasonable. For example, muskrats live at low elevation, but guinea pigs live at high altitude: thus, cardiovascular differences might reflect adaptation to altitude. In addition, muskrats are wild, but guinea pigs are domesticated: thus, observed differences might reflect adaptation to domestication. Furthermore, muskrats and guinea pigs belong to very different rodent clades, separated by tens of millions of years of independent evolution: thus, observed differences might reflect basal synapomorphies of either clade, evolving long before muskrats evolved an aquatic lifestyle and associated aquatic specializations. Last, differences might merely reflect accumulated drift (Turelli et al. 1988; Garland and Adolph 1994). And, of course, these alternatives are not mutually exclusive!

The muskrat example is partially fabricated, but many (perhaps most) comparative physiological studies prior to Felsenstein (1985*a*) selected species with little or no regard to phylogenetic or even taxonomic relationships. Exceptions exist, of course (e.g., Lasiewski and Dawson 1967; MacMillen and Nelson 1969; McNab 1969; Dawson and Hulbert 1970). Even so, some comparative and ecological physiologists were skeptical of phylogenetic approaches. For example, McNab (1969, p. 664)

concluded that "it is increasingly clear that the efficacy of homeostatic functions is mainly determined by the delicate interactions existing between an organism and its environment, only a small influence at best showing the impress of historical events." Waterman (1975, p. 313) asserted that "phylogenetic explanations in comparative physiology must be rather nonrigorous and speculative," simply because historical events are not "strictly subject to experimental verification or direct observation."

During the 1970s, comparative biologists were beginning to appreciate the strengths of comparing closely related species (Clutton-Brock 1974). (Note: Felsenstein [see app. A] traces the origin of this insight to an overlooked book by Salisbury [1942].) For example, Huey and Pianka (1981) argued that comparisons among confamilial or congeneric species would be much stronger—and less confounded—than those comparing species from different families or orders. But ever since Felsenstein (1985*a*), comparative biologists now often—although not always—pay careful attention to phylogenetic issues when selecting species for comparison. Unfortunately, the taxon sampling that would be ideal for a comparative study is often not possible, given existing patterns

of organismal diversity (or data available; Miles and Dunham 1993; Losos 2011; Huey and Pianka 2017, p. 337).

One might think that at least two species were needed for a study to be considered comparative, although Garland and Adolph (1994) offered multiple reasons why just two species are insufficient. Nevertheless, one-species comparisons do exist. A case in point involves the Galápagos marine iguana, *Amblyrhynchus cristatus*. Early studies considered several physiological capacities of these lizards as unique adaptations to diving, but Dawson et al. (1977) later noted that those early studies of marine iguanas had actually been done before any baseline studies of capacities of nondiving iguanid lizards. But once those baseline studies had been done, Dawson et al. (1977, p. 896) realized that many of the marine iguana's presumed "diving adaptations" were in fact generic among iguanid lizards, such that the marine iguana is best thought of as "a representative of a terrestrial line that was preadapted for exploiting a unique combination of circumstances." Evolutionary baselines do need to be established, not merely conjectured.

Other one-species comparisons exist, especially in the minds of biologists studying model animals. In his delightful presidential address to the American Psychological Association, Frank Beach (1950, p. 119) criticized the field of comparative psychology for being almost exclusively focused on learning behavior in the Norway rat. Beach even proposed that the name of the association's journal, the *Journal of Comparative Psychology*, be changed to the *Journal of Rat Learning*, although he conceded that many fellow comparative psychologists would object because "they appear to believe that in studying the rat they are studying all or nearly all that is important in behavior. . . . And how else are we to interpret Professor Skinner's 457-page opus which is based exclusively upon the performance of rats in bar-pressing situations but is entitled simply 'The Behavior of Organisms'?"

Scaling

A favorite empirical theme in comparative biology is scaling, which quantifies how traits (e.g., metabolic rate, brain size, and locomotor capacity) change with body size. Typically, an investigator would search the literature, compile published data (and sometimes supplement those data), and run a regression analysis (e.g., log metabolic rate on log body mass). The resulting allometric equations provided useful insights into how diverse ecological, physiological, morphological, or biomedically relevant traits change with body size. Similar approaches might apply to any analysis of the correlated evolution or coadaptation of diverse traits, even those independent of body mass.

Sometimes investigators would analyze residuals (vertical deviations from regression equations), attempting to find whether certain taxa (or ecological types) had relatively positive versus negative residuals. For example, McNab (1963) and Schoener (1968) found that predators (mammals, birds) had larger home ranges or territories than did omnivores or herbivores of the same mass. Other studies compared metabolic rates of various related taxa (e.g., passerine vs. nonpasserine birds [Lasiewski and Dawson 1967], marsupials vs. placentals [MacMillen and Nelson 1969], and marsupials vs. eutherian mammals [Dawson and Hulbert 1970]).

Such studies were insightful but suffered from two problems. First, species were not selected in an attempt to balance phylogenetic sampling but rather were compiled from available published data. Inevitably, some taxa (e.g., rodents) were overrepresented, while many other lineages were absent. Unfortunately, few studies—then or now—compare their sampling with the actual diversity of extant species (exceptions include Gorman 1973; Albuquerque et al. 2015). Ideally, comparative articles should report the fraction of species (in a clade) that are sampled and the evenness of sampling of subclades.

Second, standard statistical procedures, including those used routinely for allometry and scaling (e.g., linear regression, analysis of covariance), assume that residuals from the statistical model are independent and identically distributed (see also Grafen 1989). In this context, independence means that related species (e.g., within a genus) are unlikely to be bunched together as high or low residuals. However, shared ancestry of hierarchically related species will often produce phenotypic similarities, clearly violating nonindependence (Felsenstein 1985a). Early attempts to correct for this involved using generic or family averages, not species values (e.g., Harvey et al. 1980).

A few comparative articles were phylogenetically framed. In an underappreciated article (appearing in *Science* but cited only 89 times in Web of Science through December 23, 2018), Greene and Burghardt (1978) examined constricting behavior of snakes on a phylogeny and noted that one mode of constriction was shared by many lineages, implying a common ancestry dating to the early Paleocene. Gittleman (1981) used a parsimony approach to count the number of times parental care had evolved in fishes but had to rely on the Linnaean taxonomy rather than a phylogeny. Similarly, Shine (1985) used a taxonomy to estimate the number of times viviparity had evolved in reptiles. Lauder (1981) used a phylogeny to elucidate evolutionary correlations between morphology and the environment. Ridley (1983) proposed using parsimony to determine whether changes in two qualitative characters were independent.

Independent of Felsenstein (1985a), Cheverud et al. (1985) invoked a relevant quote from Darwin (1859, p. 185) and sounded the alarm about the need to involve phylogenies in comparative analyses. Importantly, they proposed a "phylogenetic autocorrelation model" that quantifies both phylogenetic and "specific" (adaptive) effects on the distribution of trait values of species. However, shortcomings of their method

(see Lynch 1991; Rohlf 2001) limited the impact of this pioneering effort.

Birth of a Notion: Felsenstein (1968) and Felsenstein (1985*a*)

In his perspective (app. A), Felsenstein traces the history of his 1985 article, starting with his dissertation work. As he states, he had realized while writing his dissertation (Felsenstein 1968) how independent contrasts might be constructed if characters evolved in a manner consistent with Brownian motion. However, his goal was to use continuous characters to estimate phylogenies, a topic he developed in Felsenstein (1973*b*). Hence, the task was to compute the likelihood of character data at the tips of a tree, assuming a known tree and postulated model for character evolution. The analogous calculations for discrete characters, including his famous "pruning algorithm" for dealing with internal nodes, appeared in Felsenstein (1973*a*). These fundamental ideas underlie all modern likelihood-based phylogeny estimates from genetic data. As Joe indicates (app. A), he had a beautiful machine waiting for an application. The motivation was provided later by his interactions with Paul Harvey, who introduced him to the need to develop methods that correctly analyze phenotypic data from multiple species.

Joe's dissertation (1968) contained the key, but ironically his initial doctoral project failed. He hoped to develop a tractable model relating entropy and information theory to evolution (see Felsenstein 1978). This would have reflected his extraordinary and precocious talent for population genetics theory (Felsenstein 1965; Bodmer and Felsenstein 1967), initially fostered by his undergraduate mentor, James F. Crow. His PhD mentor, Richard C. Lewontin, suggested, "Why don't you write up that tree stuff?" (J. Felsenstein, personal communication). Joe's consolation-prize dissertation reads more like an extraordinarily ambitious grant proposal than the "three-papers-stapled-together composite" that is our modern doctoral standard. Joe ruminated on how to simultaneously reconstruct both phylogenies and histories of character evolution. His most substantial result was the pruning algorithm as applied to the Brownian motion model—but this appears as a preliminary result, analogous to a lemma proven on the way to establishing an important theorem.

The tone of Joe's dissertation is encapsulated in beautiful illustrations produced by his father (reproduced here as fig. 3). Joe suggested that likelihood could unravel both phylogenies and character evolution. However, as Paul Harvey states in a recent email to Joe (app. A), even in the late 1970s it seemed inconceivable that "accurate phylogenies [would be available] in the foreseeable future." Similarly, as Joe notes, the reviewers of his 1983 manuscript that became Felsenstein (1985*a*) were doubtful that phylogenies would be available
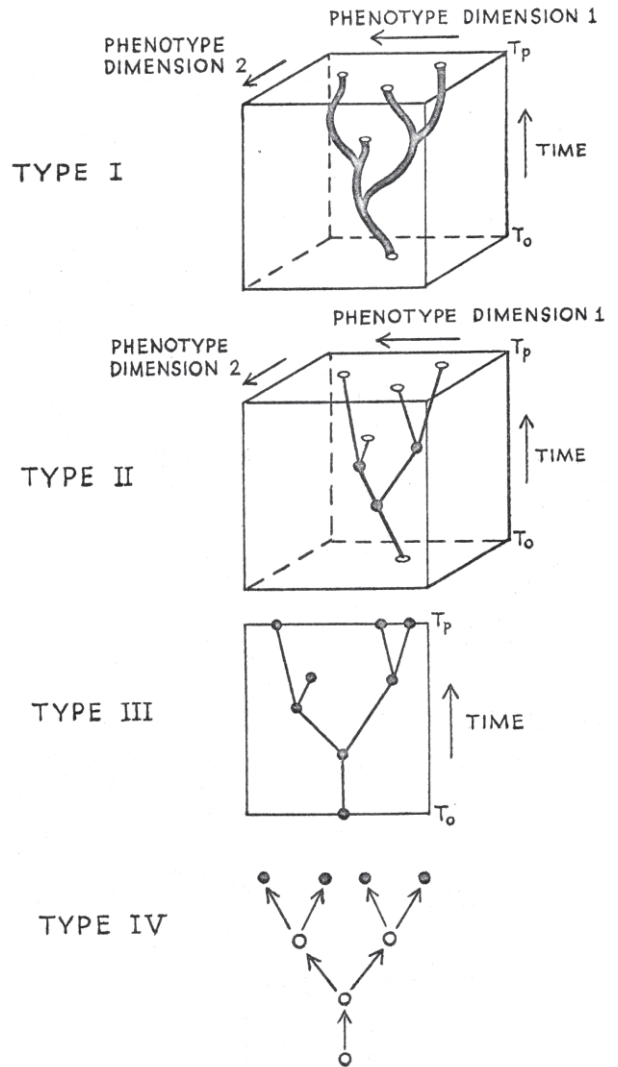
Figure 3: This figure is reproduced from Felsenstein's (1968) dissertation and was labeled "Four types of evolutionary trees." It illustrates Joe's early interest in understanding phylogenies and how he viewed traits ("phenotype dimension") coevolving over evolutionary time.

to calculate the phylogenetic corrections he proposed (see app. B). Moreover, when one of us (M.T.) was Joe's PhD student in the mid-1970s, he felt that Joe's fascination with estimating phylogenies seemed as irrelevant to understanding biology as his erudite views on alternative interpretations of Marxist ideology (for an implicit apology, see Turelli et al. 2018).

Enter Felsenstein (1985*a*)

Felsenstein (1985*a*) consists of several sections woven together to illustrate several key points. Each part is critical

to the whole, but the whole has proven greater than the sum of its parts. We summarize these parts here but strongly encourage people to read the original.

The first section was titled "The Problem" and specifically addressed nonindependence. Joe illustrated the issue via simple but powerful graphics. His figures 1–3 show how a comparative data set can have many or few independent samples, depending on the phylogeny. His figures 5–7 showed an apparent (if weak) relationship between two traits when 40 species were assumed independent (fig. 6) but a complete absence of a relationship when he revealed that the species actually belong to two long-diverged sister clades, each with recent 20-species radiations (fig. 7).

Joe could have used equations to make his points, but his simple graphics made his points elegantly accessible even to people with limited phylogenetic background or intuition. Sometimes, a picture really is worth a thousand words.

One of us (R.B.H.) vividly remembers when Joe showed him the graphs (hot off Joe's dot matrix printer) over lunch. He was aware of the nonindependence issue and of the importance of thinking phylogenetically (Huey and Pianka 1981), but he was stunned by Joe's graphs. He remembers thinking, "Now I get it." Obviously, many other biologists soon had the same reaction, as indicated by the sudden, pervasive appearance of "phylogenetic corrections" completely divorced from any model of character change (e.g., Coyne and Orr 1989).

In a rarely discussed paragraph at the end of this section, Joe notes with characteristic insight that one needs to understand the process(es) underlying character change to determine whether any phylogenetic correction is needed. If character values respond directly to changing phenotypic optima in each lineage, the resulting phenotypic values would indeed be independently distributed if the optima could be viewed as independent draws from a distribution of optima. As Pagel and Harvey (1992, p. 425) noted, "The procedures force users to make decisions about the way they think evolution has proceeded" (e.g., to be more explicit about the presumed model of character change).

The second section was titled "Previous Approaches." Here Joe outlined approaches that had been developed mainly by British comparative biologists, principally Tim Clutton-Brock and Paul Harvey (1977), who noted that nonindependence is a concern "if phylogenetic inertia is strong." Furthermore, as discussed above, they proposed using a nested ANOVA to select the appropriate taxonomic level for independent comparative analyses.

The third section was modestly called "A Possible Solution." Here Joe noted that if one had a true phylogeny (i.e., known branching order), a model of trait evolution (e.g., independent Brownian motion on each branch), and branch lengths corresponding to predicted variances in character change, then one could use routine statistics to analyze the (scaled) differences between adjacent pairs on the tree, because the "contrast" between the trait values for species A and B would be independent of the contrast between trait values for species Y and Z. Working down the nodes of the tree, whose ancestral trait values could be estimated by likelihood under the Brownian model, this method could generate $n - 1$ independent contrasts. The article then develops the mathematics of the method, showing explicitly how identically distributed independent contrasts could be constructed. Warning—here be dragons! Appendix C provides a simple guide to the details.

The fourth section was titled "Difficulties." Here Joe conceded the fact that rarely is a phylogeny well known, and so he reviewed techniques (gene frequencies, molecular sequences, and quantitative characters) that might be used to estimate phylogenies. Then he discussed how to add confidence intervals on inferences as well as the issue of the model of evolution.

The last subsection before the summary was titled "What If We Do Not Take the Phylogeny into Consideration?" Here Joe has a blunt conversation with the reader. It is clear, forceful, and unambiguous (p. 14):

> Some reviewers of this paper felt that the message was "rather nihilistic," and suggested that it would be much improved if I could present a simple and robust method that obviated the need to have an accurate knowledge of the phylogeny. . . . The best we can do is perhaps to use pairs of close relatives as suggested above, although this discards at least half of the data. Comparative biologists may understandably feel frustrated upon being told that they need to know the phylogenies of their groups in great detail, when this is not something they had much interest in knowing. Nevertheless, efforts to cope with the effects of the phylogeny will have to be made. Phylogenies are fundamental to comparative biology; there is no doing it without taking them into account.

The Initial Response

Four reviewers submitted comments on Joe's manuscript (app. B), and their comments are fascinating. The reviewers were clearly respectful of the issues and of Joe's presentation, but they were also clearly dubious that the method would ever prove useful, primarily because robust phylogenies (with branch lengths) were few and far between at that time (see the anonymous quote at the beginning of the introduction).

Initial public response to the published article was also mixed. Huey and Garland know from personal experience that many fellow comparative physiologists were openly hostile to the push to incorporate phylogenetic thinking

into the selection of species for study and into the analysis of comparative data. One told Huey that he and his ilk were "killing comparative physiology." Certainly, the rarity of phylogenies also dampened enthusiasm, even for those willing to accept the new paradigm.

Nevertheless, some influential evolutionists quickly appreciated that phylogenetic corrections must be incorporated into multispecies analyses even when no clear model for character evolution was available to guide such corrections. Sessions and Larson (1987) and Losos (1990) were apparently the first to use Joe's method. Huey and Bennett (1987) did not have a tree with branch lengths and so used squared-change parsimony (a technique suggested by Joe) to compute ancestral values of physiological traits, thereby highlighting ancestral-descendent changes rather than across-species differences. In their pioneering meta-analysis of *Drosophila* speciation, Coyne and Orr (1989) invoked Felsenstein (1985a) to motivate their procedure of averaging data on premating and postzygotic isolation across nodes of crudely estimated allozyme-based phylogenies. Specifically, if A and B are sister clades with n_A and n_B species, respectively, measurements of the $2n_A \times n_B$ possible reciprocal hybrid crosses between each pair of A versus B species cannot be viewed a priori as independent observations. Coyne and Orr (1989) adopted a conservative approach of collapsing all available comparisons between these clades into a single average observation to determine the robustness of their conclusions about the accumulation of reproductive isolation as a function of divergence time and geographic isolation.

These early adopters were soon followed by many others. We suspect that several factors promoted a groundswell of support for phylogenetic thinking. Grafen (1989) provided extensions that dealt with unknown (or unrecognized) phylogeny and a lack of information on branch lengths. Harvey and Pagel's (1991) book was, of course, the major catalyst: this book included diverse explorations and extensions of independent contrasts and related approaches. Importantly, Harvey and Pagel were bona fide comparative biologists, giving acceptance and publicity to phylogenetic thinking and analyses. Simultaneously, computer simulations by Martins and Garland (1991) demonstrated that independent contrasts worked as advertised, and they also provided the first stand-alone computer programs (PDAP) to implement independent contrasts. Grafen (1989) also conducted simulations and provided code for use with the GLIM package. The availability of these and subsequent programs (e.g., Purvis and Rambaut 1995), including implementation in PHYLIP, greatly facilitated applications of independent contrasts.[1]

Numerous articles, chapters, and symposia (e.g., Systematics Agenda 2000 [1994] and the American Physiological Society symposium in 1994) helped publicize the utility and necessity of phylogenetic approaches. But another huge stimulus was the synchronous explosion of research in phylogeny estimation. In part, this was made feasible by technical advances in sequencing: in a remarkable coincidence, polymerase chain reaction was invented in 1985. In addition, computer processing speeds were accelerating rapidly, and accessible programs enabled phylogenetic estimation (PHYLIP, MacClade, PAUP, Mesquite, etc.).

We suspect that Felsenstein (1985a) also was a contributor here, as this article made phylogenies relevant (indeed, essential!) to diverse researchers who had formerly had little interest in phylogenies. That changed. Comparative biologists not only began talking with systematists (!) but some—even with no phylogenetic background—began estimating phylogenies to enable comparative studies of interest (e.g., Block et al. 1993).

The end result was that within a decade after 1985 independent contrasts (or complementary phylogenetically informed approaches) had become the modus operandi in much of comparative biology. To be sure, phylogenetic comparative analyses are sometimes inappropriate (Losos 2011; Huey and Pianka 2017), and debates (see below) continue, especially as to the validity of a Brownian motion of evolutionary change. But phylogenies are now an integral part of comparative biology.

Advances since 1985

Many advances have occurred in accounting for phylogenetic effects and for modeling trait evolution. We cannot thoroughly review the many articles implementing and extending the ideas and methods of Felsenstein (1985a). However, given the complexity and breadth of this literature, we attempt an overview of the questions addressed and alternative approaches taken. We consider three nonexclusive categories of developments: (1) alternative implementations of phylogenetic corrections based on Brownian motion; (2) methods of testing for the "phylogenetic signal" expected under Brownian motion, estimating the relative magnitude of this signal, and using transformations to better fit data to models of trait evolution; and (3) alternative models of trait evolution. We make no pretense of being comprehensive,

---

1. As an aside, we note that some of the early computer programs did not actually perform calculations as specified by Felsenstein (1985a). For example, some did not correctly lengthen branches leading to internal nodes (as described here in app. C). Moreover, some commercial statistical packages, when instructed to cal-

culate a regression through the origin (as is required with independent contrasts; Garland 1992; Rohlf 2001), did so, but they did not calculate the corresponding correlation coefficient through the origin. One of us (T.G.) spent countless hours and exchanged innumerable emails with Joe trying to reconcile discrepant numbers from independent contrast analyses. It made one wish that Joe's original article had included a worked example with numbers! And typos in his table 1, indicated in app. C, also led to some confusion. Worked examples can now be found in Garland and Adolph (1994) and Rezende and Diniz-Filho (2012).

and we apologize for not citing all of the key publications in these rapidly expanding areas. For relatively recent reviews, see Rezende and Diniz-Filho (2012) and Ives (2018).

Alternative Implementations of the Brownian Motion Model

As noted above, implementing the ideas of Felsenstein (1985a) involved developing molecular and statistical tools for phylogenetic inference and producing software to calculate the independent and identically distributed contrasts derived from the Brownian motion model (e.g., Martins and Garland 1991). Felsenstein's (1985a) goal in constructing contrasts was to account for correlations induced by phylogenetic history. An alternative approach is to use an estimated phylogeny to infer the covariance structure expected for character data (e.g., Grafen 1989; Lynch 1991). Once the correlation structure is known, the data can be analyzed properly without explicitly constructing contrasts. Standard likelihood methods can be used to estimate parameters, including ancestral character states, correlations between the changes in different characters, and (relative) rates of character evolution. This "model the dependence" approach leads naturally to likelihood-based tests of alternative hypotheses, including independence of and constant relative rates for multiple traits. Moreover, it links inferences for tree-based data to general likelihood-based statistical frameworks, such as the general linear model (Grafen 1989) and the phylogenetic "mixed model" (Lynch 1991; Housworth et al. 2004), in which phylogenetic effects can be considered along with other covariates (e.g., see Hadfield and Nakagawa 2010). These approaches also generalize to Bayesian analyses that account for uncertainty in the phylogeny (e.g., Huelsenbeck and Rannala 2003; Lartillot and Poujol 2011). Grafen (1989) introduced his "phylogenetic regression" as a way to account for phylogenetic effects when the phylogeny and branch lengths are incompletely known. He attempted to compensate for uncertainty in both the tree, involving soft polytomies, and the model of character evolution by introducing a parameter ρ that scales the relative lengths of internal branches. Such transformations are discussed below.

Transformations, Testing for Phylogenetic Signal, and Quantifying Phylogenetic Effects

The development of computer programs to calculate contrasts or otherwise deal with phylogenetically induced correlations was accompanied by statistical tests for the fit of data to the Brownian motion model (Grafen 1989; Garland et al. 1992). Lack of fit was addressed in various ways, including alternative procedures for modifying sets of starter branch lengths. Alternative metrics were introduced to quantify phylogenetic signal and the fraction of trait variance across taxa that could be explained by Brownian motion.

Branch-length transformations were first suggested by Grafen (1989), who raised the relative lengths of internal branches to a power ρ after scaling the tree height to 1. In this framework, $\rho = 0$ corresponds to no phylogeny-induced correlation between the characters in separate taxa. Pagel (1994) and Blomberg et al. (2003) introduced transformations that have been more widely applied. Freckleton et al. (2002) review and apply Pagel's approach. Pagel (1999) introduced a linear scaling parameter ρ into the covariance between trait values, denoted X_i and X_j, in taxa i and j. Assuming that their most recent common ancestor occurred t_{ij} time units in the past, $\mathrm{Cov}(X_i, X_j) = \sigma^2 t_{ij}$ under a constant-rate Brownian motion model, with variance accumulating at rate σ^2 per unit time. Pagel suggested multiplying all covariances by λ, so that $\lambda = 0$ corresponds to no effect of shared history (i.e., no phylogenetic signal), whereas $\lambda = 1$ produces the covariances expected under Brownian motion. As with Grafen's ρ, λ can be estimated by likelihood, assuming Brownian motion. These scalings are analogous to data transformations made to better approximate the assumptions of standard statistical tests. Blomberg and Garland (2002) suggested that the goal of such analyses is to explain patterns associated with data rather than the processes that produced them, but Blomberg et al. (2003) introduced two alternative transformations that assumed non–Brownian motion trait evolution with an apparent goal of making inferences about both pattern and process. Blomberg et al. (2003) introduced parameters into the expressions for $\mathrm{Cov}(X_i, X_j)$ and $\mathrm{Var}(X_i)$ that correspond to either (1) an Ornstein-Uhlenbeck (OU) model (Hansen and Martins 1996), in which the Brownian motion random walk is replaced by a tendency to return to a specific trait value (usually interpreted as stabilizing selection), or (2) a modified Brownian motion model (accelerating/decelerating evolution [ACDC]), in which variance accumulation either accelerates or decelerates through time. In both cases, an extreme parameter value ($d = 0$ for OU and $g \ll 1$ for ACDC) corresponds to (essentially) no effect of phylogeny on trait covariances.

These alternative scaling parameters (ρ of Grafen [1989], λ of Pagel [1999], d and g of Blomberg et al. [2003]) also provide measures of the amount of phylogenetic signal present in the data. The H^2 statistic of Lynch (1991; Housworth et al. 2004) provides explicit estimates of the amount of phylogenetic signal in a continuous-valued trait. As shown by Revell et al. (2008), these estimates of "signal" provide little if any information about underlying evolutionary processes. Blomberg et al. (2003) also provide a randomization-based nonparametric test for whether phylogenetic relatedness implies greater phenotypic similarity than expected by chance. Felsenstein (1985a, p. 6) recognized that tip data should be independent and identically distributed when microevolutionary processes

within species lead to trait values determined solely by lineage-specific selection that is independent of phylogenetic history (including branch lengths). Price (1997) found—and Freckleton and Harvey (2006) confirmed—that some characteristics of this "independent tips" model can be approximated by ecologically motivated models of "niche filling." Nevertheless, the vast majority of data analyses suggest that phylogenetic signal is pervasive (Freckleton et al. 2002; Blomberg et al. 2003), demanding some correction to deal with phylogeny-induced correlations.

As Felsenstein (1985*a*) observed, accurate correction requires knowing both the phylogeny and a relevant model of character change. The problem of estimating phylogenies, which seemed virtually insurmountable in 1985, has become increasingly tractable, with ever-cheaper genomic data (and faster computers) providing previously unimaginable resolving power. Unfortunately, there is no analogous magic bullet for understanding character evolution. Multiple characters do not provide replication of the sort provided by individual nucleotides, which can be thought of as interchangeable (the essential assumption underlying Felsenstein's [1985*b*] bootstrap-based confidence intervals for phylogenies). Hence, determining the best model for character evolution is fundamentally more difficult than estimating the species phylogeny, and most researchers happily follow the default convention of assuming Brownian motion.

The subtlety of model choice is illustrated by Felsenstein's (1985*a*) treatment of branch lengths. In Felsenstein (1985*a*), "branch length" quantifies the expected variance of character change along the branch. Sister branches are allowed to differ in length (see Felsenstein's fig. 9), as routinely observed in phylograms based on DNA data. However, Felsenstein (1985*a*) was silent on how trait-relevant lengths might be estimated, leading many researchers to assume (or explicitly propose; e.g., Grafen 1989; Garland et al. 1992) that essentially any set of branch lengths (e.g., those obtained from a phylogram based on mitochondrial DNA) could provide informative estimates of the rate parameters that enter Felsenstein's formulas. We can accurately estimate relative rates of molecular evolution along sister branches because we have data from many nucleotides. Nothing comparable exists for individual phenotypic traits, such as body size. The simplest Brownian motion model for character change assumes constant rates, in which case branch lengths should be obtained from chronograms. But chronograms are fundamentally more difficult to estimate than phylogenies, because of difficulties associated with calibration (Heath and Moore 2014) and variable absolute and relative rates of molecular evolution (Langley and Fitch 1974).

As Joe has repeatedly told one of his former students (M.T.), "If you don't like my model, come up with a better one." Yet the mosaic nature of trait evolution (Simpson 1944), with dif-ferent traits evolving at different rates across different clades (and sister lineages), makes modeling the evolution of continuous traits much more difficult than modeling molecular evolution. Our typical Markov models of sequence evolution were motivated by the empirical observation of approximate constancy of divergence (Zuckerkandl and Pauling 1965; Yang 2014). Despite uncertainty about how best to model character evolution, the utility of Brownian motion as a null model for character state change is unquestionable (Martins and Garland 1991). Nevertheless, consumers should at least know what is being assumed, should routinely test for the adequacy of the model (Bollback 2002; Boettiger et al. 2012), and should explore the robustness of their conclusions to alternative models. Boettiger et al. (2012) describe relatively straightforward simulation procedures for determining the power of specific data sets to estimate parameters of interest, and they also show that some widely used information-based criteria for model choice (e.g., Akaike information criterion, Bayesian information criterion) produce artifactual inferences for data sets of typical sizes (cf. Cressler et al. 2015). The problem of model choice is associated with more general questions about the connection between hypotheses being tested and the models applied in comparative analyses (Uyeda et al. 2018).

Alternative Models for Character Change

Felsenstein (1985*a*) also described a radical alternative to Brownian motion for character changes in which each species evolves to a value independent of phylogenetic history and drawn from a Gaussian distribution. As he noted, under this model no phylogenetic correction is needed. Probably the most influential and widely applied alternative to Brownian motion is the OU model, in which the trait has a preferred value, θ, that is assumed to apply across the tree (Garland et al. 1993; Hansen and Martins 1996). This model has also been generalized to allow for multiple values of θ, corresponding to Simpson's (1944) vision of alternative peaks in the fitness landscape for phenotypes (Lande 1976; Butler and King 2004), and it continues to be developed in various ways (e.g., Adams and Collyer 2018; Bastide et al. 2018). But this more complex model demands more data and more caution concerning parameter estimation and inference (Boettiger et al. 2012; Cressler et al. 2015). The Brownian motion model has also been generalized to ask whether character evolution is better approximated by uniform rates across the tree or by rates that accelerate or decelerate through time (Blomberg et al. 2003) or vary across groups (Garland et al. 1992; O'Meara et al. 2006).

In addition to understanding how traits change across phylogenies, character evolution models are central to studies of anagenesis (e.g., Hunt 2007; Hunt and Rabosky 2014; Landis and Schraiber 2017). New models are being proposed

to better fit micro- and macroevolutionary data, with a recent emphasis on "jump" processes that combine continuous and discontinuous change, corresponding to punctuated evolution on a macroevolutionary timescale (Landis and Schraiber 2017). These Levy-process models are likely to make their way into future comparative methods.

Independent contrasts can be used to identify "unreplicated" evolutionary events that have occurred only once or a few times within a given clade. To illustrate how, imagine a log-log plot of some trait versus body size. The corresponding plot for independent contrasts, with a regression line forced through the origin, may reveal one or more statistical outliers that were not apparent in the conventional plot (the converse is also possible). Outliers can occur when the rate of evolution has been unusually high along a single branch (McPeek 1995) or during the divergence of lineages (Garland 1992). For example, during an independent contrast investigation of many Carnivora, Garland and Janis (1993) noted that the contrast between polar and grizzly bears in relative hind-limb length was too large to be an artifact of measurement error (in morphology or divergence time): thus, rapid divergence of these two bears might reflect their differential adaptation to swimming or long-distance travel (polar bears) versus digging (grizzly bears). The same approaches can be used to test a priori hypotheses about rates of evolution along branches leading to particular tip species or as clades diverged. Further discussions, examples, and methodological details can be found elsewhere (Ackerly 1999; Garland and Ives 2000; Revell 2008; Uyeda et al. 2018).

Beyond Correlated Evolution and Adaptation

The perspective and scope of phylogenetic comparative methods has broadened from improving our inferences about scaling, correlated evolution, and adaptation to testing a wide range of hypotheses about macroevolutionary patterns and processes. For example, phylogenetically informed statistical methods are now routinely incorporated into community ecology (Kembel and Hubbell 2006; Rafferty and Ives 2013; Ordonez 2014), global change and conservation biology (Faith 1992; Davis et al. 2010; Buckley and Kingsolver 2012), neontological and paleontological analyses of variation in speciation and extinction rates (e.g., Harrington et al. 2018; Martins et al. 2018), and studies of species interactions that may lead to character displacement or convergent evolution (Drury et al. 2018).

Concluding Remarks

As noted above, our article is a personal perspective on a key innovation in evolutionary biology, namely, Felsenstein's (1985a) method of phylogenetically independent contrasts. Its major contributions include convincing comparative biol-

ogists to appreciate the importance of thinking phylogenetically at all stages of a study, making phylogenies relevant to diverse groups of biologists, motivating new phylogenetic derivations, drawing attention to the importance of specific models of evolution, and providing a quantitative—if challenging—method for analyzing comparative data. Today it continues to not only inspire the derivation of many new phylogenies and new comparative explorations but also inspire many new methods of analysis and even debates over phylogenies and the comparative method. Few if any evolutionary articles in our lifetimes can approach Felsenstein (1985a) for impact. Nothing more need be said.

APPENDIX A

Felsenstein's Perspective, by Joseph Felsenstein

In 1967 I was working on my PhD thesis at the University of Chicago in the laboratory of Richard Lewontin. As a result of contact with Anthony Edwards, I had become aware of the Brownian motion model for gene frequency evolution and of his efforts to use it to infer phylogenies by maximum likelihood. He was having difficulties owing to attempting to infer the states of the interior nodes of the tree at the same time. I took a different approach, summing over the states of the interior nodes. This worked better, but I still needed a way to calculate the likelihood of a given tree, preferably without doing matrix inversion.

I had seen in texts of multivariate statistics that it was sometimes possible to transform variables two at a time to come up with a set of uncorrelated variables, so I attempted to do that. Sister species seemed most promising. I took the difference between their coordinates and tried to find a weighted average of the two that would be uncorrelated with that. This worked, but there was still the issue of how those correlated with the values in the other species. To my delight, the difference proved to be uncorrelated with the coordinates of all the others. As for the weighted average, it turned out to have covariances with the other values that were those we would

expect from a species that was located where the latest common ancestor of the two had been but with a modest amount of extra branch length added there.

Using this, it proved possible to decompose the tree into independent variables by working down the tree, taking differences between sister species, replacing their common ancestor by an appropriately weighted average of the two, and lengthening the branch leading to the common ancestor. These contrasts, done in each locus, enabled rapid computation of the likelihood. This was my first use of the "pruning" algorithm for likelihoods. The algorithm was included in my thesis (1968) and published in my article (Felsenstein 1973*b*) on inferring phylogenies from gene frequency data, which was published in the *American Journal of Human Genetics* in 1973. In the meantime, I had encountered the use of "peeling" algorithms for likelihoods in pedigrees in statistical genetics, and I realized that this was analogous to my contrasts for the Brownian motion case. It was immediately clear how to construct a pruning algorithm for discrete characters, including DNA sequences. I published the discrete-states pruning algorithm in an article in *Systematic Zoology* (Felsenstein 1973*a*). All of these attracted little attention.

In my 1973 article on gene frequency phylogenies (Felsenstein 1973*b*), I also included some discussion of whether this would also work for the values of continuous characters. The worrisome problems were character correlation and unequal rates of change in different characters. One could sensibly assume that gene frequencies underwent genetic drift independently at different loci, but quantitative characters were obviously liable to be correlated. For the case in which all of the gene frequencies changed only by genetic drift, there would be a proportionality between the within-population additive genetic covariances of characters and the covariances of the changes through time in a lineage. This might enable us to scale the latter by the former. But when the changes through time were caused by selection, there would be no guarantee of such proportionality. I discussed this in the 1973 article, ending without a general method for using likelihoods to infer phylogenies from values of quantitative characters. Although I published another article in *Evolution* (Felsenstein 1981) that repeated much of the material from the 1973 article, there I did not discuss quantitative characters. In a discussion article in the *Journal of the Royal Statistical Society* (Felsenstein 1983), I noted the following as one of the "unsolved problems" (p. 260):

> Correcting for correlation of characters in the Brownian motion model. The model of Brownian motion used for gene frequencies could also be applied to quantitative characters, a case of great practical importance, but only if some way could be found to correct for the correlations in the evolutionary changes of different characters. The within-species covariances will not nec-

essarily be helpful, as there is no guarantee that covariances of evolutionary changes are proportional to within-species covariances.

In effect, I was all dressed up with nowhere to go. I had the machinery for coping with character correlations in computing the likelihoods of trees but no practical way of inferring those correlations.

In 1982 a related problem was brought to my attention. I had met Paul Harvey when he visited Seattle in the late 1970s and had helped him solve a multivariate statistics problem. In 1982 he returned again for a six-month stay in our Department of Zoology. During it, Paul gave a seminar on the puzzle of what taxonomic level to use when studying the correlations between characters in evolution. In his work with Tim Clutton-Brock and Georgina Mace (e.g., Clutton-Brock and Harvey 1980; Harvey et al. 1980), they had been worried by the close similarities of members of the same genus. They tried to correct for this by using averages for genera rather than using values from individual species (using nested ANOVA to identify the appropriate taxonomic level for analysis). But this, of course, raised the issue of whether genera in different orders were comparable taxonomic levels.

Paul recently wrote me the following:

> In 1977 and the few years afterward I had been responsible for the statistical analysis on the Clutton-Brock and Harvey papers. I was aware that species were not independent data for analysis and I tried to identify where the variation lay. To start with, we used genera. By the time of my 1982 talk, I was arguing that we should use nested analysis of variance and covariance based on taxonomy to identify independence. By that time I was publishing on the issue in methodological papers with Georgina Mace [Harvey et al. 1980] and Bob Martin [Martin and Harvey 1985]. . . . The day or so after I gave my talk, you pointed out in a follow up lunch time seminar that independent contrasts based on phylogeny rather than taxonomy was the way to go. You were correct. (It had never occurred to me that we might get accurate phylogenies in the foreseeable future so that we could escape from taxonomies.) (Email from Paul Harvey to Joe Felsenstein, January 24, 2018, used by permission; citations inserted by us)

In 1982 I went on a year's sabbatical leave to the Institute of Animal Genetics at the University of Edinburgh. During that time I gave a seminar on the use of phylogenies in comparative methods at University College London and in the PopGroup (Population Genetics Group) meeting at the University of Birmingham. At the former seminar a woman who was a postdoc or student made the excellent point that another way of coping with phylogenies

was to find pairs of species that were close in the phylogeny, for which the paths between them did not intersect, and basing an analysis on their differences, which would then be independent. I did not catch her name and later asked others if they knew who she was—none did. When I submitted the article, I cited her as anonymous, name to be supplied later. Some of the referees wondered what obscure conspiratorial motive led me to deliberately conceal her name and rob her of the credit. A few years later I heard that Salisbury (1942) had earlier made the same suggestion and could cite that instead.

The article was submitted to *The American Naturalist* on the 23rd of November in 1983. In those years I had to carefully avoid most systematics journals, as they had editors and reviewers who were hard-core cladists who were strongly prejudiced against statistical inference and against me in particular. *The American Naturalist* did not have this problem, and over the years it had published many important articles on related topics like allometry. Rereading the reviews, I see that the reviewers were not raising issues from the systematics wars of the era. Instead, their doubts came from a skepticism that phylogenies would be available at all. If I had submitted the article a year earlier, I suspect that it might not have been accepted. But 1985 turned out also to be the year that polymerase chain reaction was discovered; the tide of molecular data was becoming a flood.

The editors did insist that one excessively arrogant attempt at humor be removed: I had said at the very end of the article that if someone insisted on doing comparative methods without using phylogenies, it might be more useful if they took up selling real estate. The editors were wise not to allow this. The article appeared as the first one in the January 1985 issue. I do not know whether this placement was accidental or whether the editors deliberately placed it there. The method got a major boost when Paul Harvey and Mark Pagel featured it in their 1991 book, *The Comparative Method in Evolutionary Biology*. Over the years this article has been seen as central to phylogenetic approaches to analyzing comparative data. Many comparative biologists assume that it must be my most highly cited article (it is my fourth most highly cited article). Having the year in which it got the most citations be 2016 is gratifying.

APPENDIX B

The Manuscript Reviews

Joe Felsenstein kept the four anonymous reviews and kindly has given us permission to transcribe them here. We have resisted the temptation to highlight several sentences. Note that reasons behind certain comments are addressed in Felsenstein's commentary (app. A).

Reviewer 1

This is an interesting problem. Its importance is described very well by the author.

Page 7, last two lines: There is a problem with subscripts due to the fact that the printer was typing too low on the page.

Page 13, paragraph 2, line 2: I can't wait to find out who this mysterious person is!

Page 15, last line: I agree, but you may not really want to say that. They might double their fee the next time you try to sell your house!

In general, this is a very interesting paper. The main problem is simply the fact that it is a very difficult problem with no real solution unless one is willing to believe that one has a good independent estimate of the phylogeny of the organisms one is dealing with. About all one can do in a practical study is simply to add some footnotes to one's paper indicating that there are problems with such uses of the comparative method.

Reviewer 2

This paper addresses a complex and important issue, and provides a solution to part of the problem—a very unsatisfactory solution, as the author is well aware, given the degree to which our data will usually fall short of the quality required by the method he proposes. In fact, Felsenstein does such a good job of pointing out the difficulties of applying his method at present that the paper comes across as rather nihilistic. Can he offer us no more practical advice than to close down comparative biology until the phylogeny of everything known becomes known in substantial detail? Nevertheless, as far as I can tell the method does what is claimed, and it is probably worth publishing, both as a stimulus to further work and because some few people may actually have data sets sufficiently detailed that it will be of immediate use.

I have three major problems with this paper. First, the issue of phylogenetic effects is only a subset of the more general problem of assuring statistical independence of one's data in comparative studies. The strength of Felsenstein's remarks, and the success of his proposed solution, depend entirely upon the degree to which "phylogenetic inertia" may have affected the relationship between the variables being investigated. This may be expected to differ in different types of studies. For example, if the Y variable depends more or less instantaneously upon the X variable (as might expected in regressions of ecological variables on morphology or on body mass), the problem of phylogenetic inertia largely disappears. There may still be problems with independence of the data (see below), but phylogeny per se is irrelevant. In other types of studies, such as those involving body pro-

portions or brain/body-mass studies, "phylogenetic inertia" is obviously going to be more of a problem. I wish that the paper had acknowledged this, rather than lumping all "comparative" studies together and assuming that the problem appears with equal force in all.

Even in the case of the ecological variables the relationship between X and Y may depend on other factors than the degree of relatedness, such as the habitat occupied. If degree of relatedness covaries with preferred habitat, there may be covariation between the phylogeny and the relationship between the variables under study. This would produce an apparent phylogenetic effect, but clearly its removal (though desirable) would only partially correct for dependence in the data. Given the number of such factors that could affect the independence of data in these kinds of comparative studies, and the difficulty of predicting their effects a priori, it seems almost quixotic to expect to achieve absolute statistical independence in any real case. Correcting for real or apparent phylogenetic inertia does not guarantee it.

Second, I found the exposition of the method (pp. 6–10) a little difficult to follow. S_x and S_y are not clearly defined, once it is mentioned that they are not sample standard deviations on page 8. Also, I am not sure throughout the discussion when and whether the subscripts X and Y are referring to the original variables or to the contrasts. I think I can figure it out, but more explicit notation or, better yet, a simple numerical example of the regression analysis described in equations (2) would greatly benefit most readers.

Finally, the last section of the paper (p. 15) is arrogant and snide. While this may fairly represent the views of the author, it seems out of place in a scientific work. At least, he could tell us exactly who the naughty, naughty, BAD comparative biologists are who have ADAMANTLY REFUSED to refer to phylogenies—and on what occasions they have done so (in response to reviewer's comments? After a few beers at the local bar?). In any case, the American Naturalist is devoted to the unification of the biological sciences, and should not allow its pages to be used for employment counseling.

Reviewer 3

This is an excellent manuscript that I recommend for publication in the American Naturalist. The author describes a statistical problem associated with numerical studies using the comparative method. For a certain class of models, it is shown that a standard method can lead to erroneous conclusions. A method of correctly analysing data is presented which requires information concerning the phylogeny of the taxa involved. The problem discussed is important. The means of resolving it are clearly presented, at least for one class of models.

The author does not indicate that there are any alternative plausible models for which the standard methods are appropriate. I believe that such alternatives exist. This paper makes it clear, however, that any alternative would have to be explicitly justified before it could be used as a basis for analysis. One would have to demonstrate that the problems described by the author do not apply. I hope that this paper would stimulate a discussion of such alternatives and the means to test their appropriateness.

Reviewer 4

The subject matter of this paper is certainly timely, and the matter needs airing. The paper doesn't quite 'get there', but perhaps that can be dealt with. I have a number of specific suggestions and questions below, but before I get to that, I'd like to deal with a pair of (seemingly) political issues.

First, the withholding of the name of someone who shared (personal communication) with the authors suggests a concern that (a) some poor soul is concerned for his(her) reputation or (b) that the releasing of the name would endanger the author's chances of having the manuscript accepted. Ridiculous on both counts! Second, the blast at the end of the paper is an uncalled for 'cheap shot.' Frankly, it's surprising from someone of the author's professional stature. Now, on to some constructive commentary.

The usual statistical concern over 'independence' is a consideration about the sampling frame, not about deeper connections among the extant sample units. If one could sample OUT's from, but without reference to, the phylogeny, one could be said to have met the requirements of statistical independence. Of course, one of the problems is that OUT's are not sampled at random and without regard for the phylogeny; they are sampled because of it. Somehow, that distinction needs to be made clear in the manuscript.

The question of normality is a separate issue from independence and requires another solution. One could easily come up with non-normal techniques if one could be assured of independence, as pointed out in the paper.

The interpretation (phyletic) that one puts on any resultant 'correlation' certainly does depend on phylogenetic considerations, but that fact (in and of itself) is anything but regrettable. I'm getting a scrambled version of these various messages from the paper.

On page 9, below Eq (3), I believe the statement should read ". . . weights being proportional to the inverses . . .". It isn't intuitively clear why one is lengthening the branch below the i-th note. A few words of explanation would probably clear it up.

On predetermining the phylogeny, I have several comments. The gene frequency strategy is alright as long as one does not want to correlate gene frequencies, and numerous

people are eager to do so. The quantitative character suggestion isn't very helpful. Why not settle for a pooled 'within-taxon' covariance matrix as 'the best we can do'? Given proper scaling, it probably would give us a reasonable reference frame.

The likelihood ratio test on page 13 is so sensitive to violated assumptions that I wonder at the wisdom of using it in this situation of almost hopeless ignorance. I would think that, leaving the taxonomic interpretation aside for the moment, the best way to obtain some sort of significance statement from the data is to use a permutation test (on population ID's for one or the other of the variables).

In summary, I think the paper has a lot of potential, but more attention needs to be paid to how one is to function in a world of imperfect taxonomic (and phyletic) knowledge. The problems could be stated more clearly, but are nevertheless real enough to be compelling. Even the clever solutions presented here, however, fall regrettably short of dealing with the real ignorance we practitioners face. What we really need is a statistical treatment that allows for undetermined sampling frame and phyletic pattern.

APPENDIX C

A Child's Guide to Two Calculations Underlying Table 1 of Felsenstein (1985a), by Michael Turelli

Figures 1–7 of Felsenstein (1985a) lucidly illustrate the origin and possible consequences of phylogeny-induced correlations between trait values across species. In contrast, Felsenstein's proposed solution, independent contrasts derived under a Brownian motion model, is summarized in his more challenging table 1. Here we suggest a modification of its notation, point out two typos, and present derivations of both types of formulas that appear at the bottom of table 1, namely, estimates of ancestral trait values, corresponding to the internal nodes of the tree presented in figure 9, and the variances associated with estimated character changes along internal branches (the branch-lengthening formulas).

We start with some notation. In table 1, the trait values at the tips of the tree and those at internal nodes are all denoted X_i; that is, no distinction is made between tip values (associated with extant species), which can be directly observed (or at least estimated from samples), and ancestral values, which must be estimated from tip data and the Brownian motion model. Following standard statistical convention, we use capital letters, X_i, to denote the trait value of taxon i when considered as a random variable. If a specific value for X_i is observed, then we denote the observed value by lowercase x_i. In contrast, if we estimate X_i from observed values, we denote the estimate $\hat{X}_i$. For tip values we have observations, x_i, whereas for internal nodes we have only estimates, $\hat{X}_i$. (Yes, we are ignoring the fact that tip values are also estimates, but our convention is consistent with the variance formulas that Felsenstein [1985a] provides— and that we derive below. For methods that include information on within-species variation, see Ives et al. [2007] and Felsenstein [2008].) One final piece of notation is useful. Felsenstein (1985a) uses v_i to denote the variance associated with Brownian motion character change along branch i (which, by definition, leads to taxon i, whether it is tip or an internal node). We use Δ_i to denote the random variable corresponding to this character change. Felsenstein's Brownian motion model assumes that the Δ_i are all independent (i.e., that changes along distinct branches are independent) and normally distributed, with $E(\Delta_i) = 0$ and $Var(\Delta_i) = v_i$.

As noted by Ted Garland, two typos occur in the last two equations below table 1. First, in the expression for X_8 (which we will denote $\hat{X}_8$), the denominator should be $v_3 + v_6'$ (not $v_3 + v_6$). Second, the expression for v_8' should start with $v_8' = v_8 + \cdots$ (not $v_8' = v_7' + \cdots$). Both expressions are explained below.

Estimates of ancestral character states. Referring to Felsenstein's figure 9, X_6 denotes the trait value for the most recent common ancestor (MRCA) of extant sister taxa 4 and 5. To calculate the maximum likelihood estimate for X_6, note that under Felsenstein's Brownian model, the distribution of trait values in descendant taxa 4 and 5, conditioned on trait value $X_6 = x_6$ in taxon 6, their MRCA (i.e., $(X_4, X_5 | X_6 = x_6)$) is bivariate normal with mean (x_6, x_6), variances (v_4, v_5), and covariance 0. This follows from the fact that conditional on the trait value in their MRCA, the random variables corresponding to the character values in the descendants can be described by $(X_4 | X_6 = x_6) = x_6 + \Delta_4$ and $(X_5 | X_6 = x_6) = x_6 + \Delta_5$. The expected character change along each branch is 0 (i.e., $E(\Delta_i) = 0$), the change along branch i has variance v_i (i.e., $Var(\Delta_i) = v_i$), and the changes along each branch are independent and normally distributed. Thus, the joint probability density function for $(X_4, X_5 | X_6 = x_6)$ is

$$f_{X_4,X_5}(x_4, x_5) = \frac{1}{\sqrt{2\pi v_4}} \text{Exp}\left(\frac{(x_4 - x_6)^2}{2v_4}\right)$$
$$\times \frac{1}{\sqrt{2\pi v_5}} \text{Exp}\left(\frac{(x_5 - x_6)^2}{2v_5}\right),$$

and the log likelihood, denoted $\ln(L(x_6))$, associated with $X_6 = x_6$, given $X_4 = x_4$ and $X_5 = x_5$, is just the natural log of this probability density function, namely,

$$\ln L(x_6) = \text{constant} - \frac{(x_4 - x_6)^2}{2v_4} - \frac{(x_5 - x_6)^2}{2v_5}. \quad \text{(C1)}$$

Finding the maximum of equation (C1) with respect to x_6, we obtain

$$\hat{X}_6 = \frac{x_4/v_4 + x_5/v_5}{1/v_4 + 1/v_5} = \frac{x_4 v_5 + x_5 v_4}{v_4 + v_5}. \quad \text{(C2)}$$

Note that this is an estimate of X_6 (hence the hat). Felsenstein also presents formulas for $\hat{X}_7$ and $\hat{X}_8$. The expression for $\hat{X}_7$ is precisely analogous to equation (C2). However, the expression for $\hat{X}_8$ involves a new variance parameter, denoted v_6', which is associated with the fact that $\hat{X}_8$ depends on the trait value for taxon 6, which is not observed but only estimated. The new variance parameter is explained below.

Why "lengthen the branches" associated with internal nodes? Associated with estimates of character states at internal nodes, Felsenstein (1985a) presents variance values, v_i', that increase the variances v_i associated with changes Δ_i along each branch. Why? The parameter v_6 is the variance accumulated along the branch leading to taxon 6; that is, $v_6 = \text{Var}(\Delta_6) = \text{Var}(X_6|X_8)$. In contrast, v_6' is the variance of the estimate of X_6 given x_8 (i.e., given the trait value in the MRCA of the two taxa in the contrast $X_3 - \hat{X}_6$); that is, Felsenstein's v_6' is just $\text{Var}(\hat{X}_6|X_8 = x_8)$. When expressed in terms of the random variables X_4 and X_5, whose observed values enter $\hat{X}_6$, we have

$$\hat{X}_6 = \frac{X_4 v_5 + X_5 v_4}{v_4 + v_5}. \quad \text{(C3)}$$

To calculate the variance of this random variable conditional on the ancestral value X_8, just note that for $i = 4$ or 5, $(X_i|X_8 = x_8) = x_8 + \Delta_6 + \Delta_i$. Substituting these expressions into equation (C3) and simplifying, we obtain

$$
\begin{aligned}
v_6' &= \text{Var}(\hat{X}_6|X_8 = x_8) \\
&= \text{Var}\left(x_8 + \Delta_6 + \frac{\Delta_4 v_5 + \Delta_5 v_4}{v_4 + v_5}\right) \quad \text{(C4)} \\
&= v_6 + \frac{v_4 v_5}{v_4 + v_5}.
\end{aligned}
$$

This is the expression for v_6' given below table 1 of Felsenstein (1985a).

In general, contrasts are differences between trait values at internal nodes or tips. When (estimated) values associated with internal nodes are used, the variances associated with the branches leading to those nodes must be increased as described by equation (C4) to account for the additional variance associated with estimating the ancestral trait values.

Literature Cited

Ackerly, D. D. 1999. Comparative plant ecology and the role of phylogenetic information. Pages 391–414 *in* M. C. Press, J. D. Scholes, and M. G. Braker, eds. Physiological plant ecology. Blackwell Science, Oxford.

Adams, D. C., and M. L. Collyer. 2018. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Systematic Biology 67:14–31.

Albuquerque, R. L., K. E. Bonine, and T. Garland Jr. 2015. Speed and endurance do not trade off in phrynosomatid lizards. Physiological and Biochemical Zoology 88:634–647.

Bartholomew, G. A. 1987. Interspecific comparison as a tool for ecological physiologists. Pages 11–37 *in* M. E. Feder, A. F. Bennett, W. W. Burggren, and R. B. Huey, eds. New directions in ecological physiology. Cambridge University Press, Cambridge.

Bastide, P., C. Ané, S. Robin, and M. Mariadassou. 2018. Inference of adaptive shifts for multivariate correlated traits. Systematic Biology 662–680.

Beach, F. A. 1950. The Snark was a Boojum. American Psychologist 5:115–124.

Block, B. A., J. R. Finnerty, A. F. Stewart, and J. Kidd. 1993. Evolution of endothermy in fish: mapping physiological traits on a molecular phylogeny. Science 260:210–214.

Blomberg, S. P., and T. Garland Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. Journal of Evolutionary Biology 15:899–910.

Blomberg, S. P., T. Garland Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:714–745.

Bodmer, W. F., and J. Felsenstein. 1967. Linkage and selection: theoretical analysis of the deterministic two locus random mating model. Genetics 57:237–265.

Boettiger, C., G. Coop, and P. Ralph. 2012. Is your phylogeny informative? measuring the power of comparative methods. Evolution 66:2240–2251.

Bollback, J. P. 2002. Bayesian model adequacy and choice in phylogenetics. Molecular Biology and Evolution 19:1171–1180.

Brooks, D. R., and D. A. McLennan. 1991. Phylogeny, ecology, and behavior: a research program in comparative biology. University of Chicago Press, Chicago.

Buckley, L. B., and J. G. Kingsolver. 2012. Functional and phylogenetic approaches to forecasting species' responses to climate change. Annual Review of Ecology and Systematics 43:205–226.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. American Naturalist 164:683–695.

Carvalho, P., J. A. F. Diniz-Filho, and L. M. Bini. 2005. The impact of Felsenstein's "Phylogenies and the Comparative Method" on evolutionary biology. Scientometrics 62:53–66.

Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. Evolution 39:1335–1351.

Clutton-Brock, T. H. 1974. Primate social organization and ecology. Nature 250:539–542.

Clutton-Brock, T. H., and P. H. Harvey. 1977. Primate ecology and social organization. Journal of Zoology 183:1–39.

———. 1980. Primates, brains, and ecology. Journal of Zoology 190:309–323.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

Cressler, C. E., M. A. Butler, and A. A. King. 2015. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. Systematic Biology 64:953–968.

Darwin, C. 1859. The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. J. Murray, London.

———. 1871. The descent of man, and selection in relation to sex. J. Murray, London.

Davis, C. C., C. G. Willis, R. B. Primack, and A. J. Miller-Rushing. 2010. The importance of phylogeny to the study of phenological response to global climate change. Philosophical Transactions of the Royal Society B 365:3201–3213.

Dawson, T. J., and A. J. Hulbert. 1970. Standard metabolism, body temperature and surface areas of Australian marsupials. American Journal of Physiology 218:1233–1238.

Dawson, W. R., G. A. Bartholomew, and A. F. Bennett. 1977. A reappraisal of the aquatic specializations of the Galapagos marine iguana (*Amblyrhynchus cristatus*). Evolution 31:891–897.

Drury, J., G. F. Grether, T. J. Garland Jr., and J. Morlon. 2018. An assessment of phylogenetic tools for analyzing the interplay between interspecific interactions and phenotypic evolution. Systematic Biology 67:413–427.

Faith, D. P. 1992. Conservation evaluation and phylogenetic diversity. Biological Conservation 61:1–10.

Feder, M. E. 1987. The analysis of physiological diversity: the prospects for pattern documentation and general questions in ecological physiology. Pages 38–70 *in* M. E. Feder, A. F. Bennett, W. W. Burggren, and R. B. Huey, eds. New directions in ecological physiology. Cambridge University Press, Cambridge.

Felsenstein, J. 1965. The effect of linkage on directional selection. Genetics 52:349–363.

———. 1968. Statistical inference and the estimation of phylogenies. PhD diss. University of Chicago, Chicago.

———. 1973a. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. Systematic Zoology 22:240–249.

———. 1973b. Maximum-likelihood estimation of evolutionary trees from continuous characters. American Journal of Human Genetics 25:471–492.

———. 1978. Macroevolution in a model ecosystem. American Naturalist 112:177–195.

———. 1981. Evolutionary trees from gene-frequencies and quantitative characters—finding maximum-likelihood estimates. Evolution 35:1229–1242.

———. 1983. Statistical-inference of phylogenies. Journal of the Royal Statistical Society A 146:246–272.

———. 1985a. Phylogenies and the comparative method. American Naturalist 125:1–15.

———. 1985b. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791.

———. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. American Naturalist 171:713–725.

Freckleton, R. P., and P. H. Harvey. 2006. Detecting non-Brownian trait evolution in adaptive radiations. PLoS Biology 4:e373.

Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. American Naturalist 160:712–726.

Garland, T., Jr. 1992. Rate tests for phenotypic evolution using phylogenetically independent contrasts. American Naturalist 140:509–519.

Garland, T., Jr., and S. C. Adolph. 1994. Why not to do two-species comparative studies: limitations on inferring adaptation. Physiological Zoology 67:797–828.

Garland, T., Jr., A. F. Bennett, and E. L. Rezende. 2005. Phylogenetic approaches in comparative physiology. Journal of Experimental Biology 208:3015–3035.

Garland, T., Jr., A. W. Dickerman, C. M. Janis, and J. A. Jones. 1993. Phylogenetic analysis of covariance by computer simulation. Systematic Biology 42:265–292.

Garland, T., Jr., P. H. Harvey, and A. R. Ives. 1992. Procedures for the analysis of comparative data using phylogenetically independent contrasts. Systematic Biology 41:18–32.

Garland, T., Jr., and A. R. Ives. 2000. Using the past to predict the present: confidence intervals for regression equations in phylogenetic comparative methods. American Naturalist 155:346–364.

Garland, T., Jr., and C. M. Janis. 1993. Does metatarsal/femur ratio predict maximal running speed in cursorial mammals? Journal of Zoology 229:133–151.

Gittleman, J. 1981. The phylogeny of parental care in fishes. Animal Behaviour 29:936–941.

Gorman, G. C. 1973. The chromosomes of the Reptilia, a cytotaxonomic interpretation. Pages 349–424 *in* A. B. Chiarelli and E. Capanna, eds. Cytotaxonomy and vertebrate evolution. Academic Press, New York.

Grafen, A. 1989. The phylogenetic regression. Philosophical Transactions of the Royal Society B 326:119–157.

Greene, H. W., and G. M. Burghardt. 1978. Behavior and phylogeny: constriction in ancient and modern snakes. Science 200:74–77.

Hadfield, J. D., and S. Nakagawa. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. Journal of Evolutionary Biology 23:494–508.

Hansen, T. F., and E. P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404–1417.

Harrington, S. M., J. M. de Haan, L. Shapiro, and S. Ruane. 2018. Habits and characteristics of arboreal snakes worldwide: arboreality constrains body size but does not affect lineage diversification. Biological Journal of the Linnean Society 125:61–71.

Harvey, P. H., T. H. Clutton-Brock, and G. M. Mace. 1980. Brain size and ecology in small mammals and primates. Proceedings of the National Academy of Sciences of the USA 77:4387–4389.

Harvey, P. H., and M. D. Pagel. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

Heath, T. A., and B. R. Moore. 2014. Bayesian inference of species divergence times. Pages 487–533 *in* L. K. Ming-Hui Chen and P. Lewis, eds. Bayesian phylogenetics: methods, algorithms, and applications. Sinauer, Sunderland, MA.

Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. American Naturalist 163:84–96.

Huelsenbeck, J. P., and B. Rannala. 2003. Detecting correlation between characters in a comparative analysis with uncertain phylogeny. Evolution 57:1237–1247.

Huey, R. B. 1987. Phylogeny, history, and the comparative method. Pages 76–98 *in* M. E. Feder, A. F. Bennett, W. W. Burggren, and R. B. Huey, eds. New directions in ecological physiology. Cambridge University Press, Cambridge.

Huey, R. B., and A. F. Bennett. 1987. Phylogenetic studies of coadaptation: preferred temperatures versus optimal performance temperatures of lizards. Evolution 41:1098–1115.

Huey, R. B., and D. Berrigan. 1996. Testing evolutionary hypotheses of acclimation. Pages 205–237 *in* I. A. Johnston and A. F. Bennett, eds. Phenotypic and evolutionary adaptation to temperature. University of Cambridge Press, Cambridge.

Huey, R. B., and E. R. Pianka. 1981. Ecological consequences of foraging mode. Ecology 62:991–999.

———. 2017. Body temperature distributions of active diurnal lizards in three deserts: skewed up or skewed down? Functional Ecology 32:334–344.

Hunt, G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. Proceedings of the National Academy of Sciences of the USA 104:18404–18408.

Hunt, G., and D. L. Rabosky. 2014. Phenotypic evolution in fossil species: pattern and process. Annual Review of Earth and Planetary Science 42:421–441.

Ives, A. R. 2018. Mixed and phylogenetic models: a conceptual introduction to correlated data. Leanpub.

Ives, A. R., P. E. Midford, and T. Garland Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. Systematic Biology 56:252–270.

Kembel, S. W., and S. P. Hubbell. 2006. The phylogenetic structure of a Neotropical forest tree community. Ecology 87:S86–S99.

Krebs, H. A. 1975. The August Krogh principle: "for many problems there is an animal on which it can be most conveniently studied." Journal of Experimental Zoology 194:221–226.

Lande, R. 1976. Natural selection and genetic drift in phenotypic evolution. Evolution 30:314–334.

Landis, M. J., and J. G. Schraiber. 2017. Pulsed evolution shaped modern vertebrate body sizes. Proceedings of the National Academy of Sciences of the USA 114:13224–13229.

Langley, C. H., and W. M. Fitch. 1974. An examination of the constancy of the rate of molecular evolution. Journal of Molecular Evolution 3:161–177.

Lartillot, N., and R. Poujol. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. Molecular Biology and Evolution 28:729–744.

Lasiewski, R. C., and W. R. Dawson. 1967. A re-examination of the relation between standard metabolic rate and body weight in birds. Condor 69:13–23.

Lauder, G. V. 1981. Form and function: structural analysis in evolutionary morphology. Paleobiology 7:430–442.

Losos, J. B. 1990. Ecomorphology, performance capability, and scaling of West Indian *Anolis* lizards: an evolutionary analysis. Ecological Monographs 60:369–388.

———. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology. American Naturalist 177:709–727.

Lynch, M. 1991. Methods for the analysis of comparative data in evolution biology. Evolution 45:1065–1080.

MacMillen, R. E., and J. E. Nelson. 1969. Bioenergetics and body size in dasyurid marsupials. American Journal of Physiology 217:1246–1251.

Martin, R., and P. H. Harvey. 1985. Brain size allometry ontogeny and phylogeny. Pages 147–174 *in* W. L. Jungers, ed. Size and scaling in primate biology. Springer, Boston.

Martins, E. P., and T. Garland Jr. 1991. Phylogenetic analysis of the correlated evolution of continuous characters: a simulation study. Evolution 45:534–557.

Martins, M. J. F., T. M. Puckett, R. Lockwood, J. P. Swaddle, and G. G. Hunt. 2018. High male sexual investment as a driver of extinction in fossil ostracods. Nature 556:366–369.

McNab, B. K. 1963. Bioenergetics and the determination of home range size. American Naturalist 97:133–140.

———. 1969. The economics of temperature regulation in Neotropical bats. Comparative and Biochemical Physiology 31:227–268.

McPeek, M. A. 1995. Testing hypotheses about evolutionary change on single branches of a phylogeny using evolutionary contrasts. American Naturalist 145:686–703.

Miles, D. B., and A. E. Dunham. 1993. Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. Annual Review of Ecology and Systematics 24:587–619.

Nei, M. 1972. Genetic distance between populations. American Naturalist 106:283–292.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:992–933.

Ordonez, A. 2014. Functional and phylogenetic similarity of alien plants to co-occurring natives. Ecology 95:1191–1202.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society B 255:37–45.

———. 1999. Inferring the historical patterns of biological evolution. Nature 401:877–884.

Pagel, M., and P. H. Harvey. 1992. On solving the correct problem: wishing does not make it so. Journal of Theoretical Biology 156:425–430.

Paine, R. T. 1966. Food web complexity and species diversity. American Naturalist 110:65–75.

Price, T. 1997. Correlated evolution and independent contrasts. Philosophical Transactions of the Royal Society B 352:519–529.

Purvis, A., and A. Rambaut. 1995. Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analysing comparative data. Computer Applications in the Biosciences 11:247–251.

Rafferty, N. E., and A. R. Ives. 2013. Phylogenetic trait-based analyses of ecological networks. Ecology 94:2321–2333.

Revell, L. J. 2008. On the analysis of evolutionary change along single branches in a phylogeny. American Naturalist 172:140–147.

Rezende, E. L., and J. A. F. Diniz-Fihlo. 2012. Phylogenetic analyses: comparing species to infer adaptations and physiological mechanisms. Comprehensive Physiology 2:639–674.

Ridley, M. 1983. The explanation of organic diversity: the comparative method and adaptations for mating. Clarendon, Oxford.

Rohlf, F. J. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. Evolution 55:2143–2160.

Salisbury, E. J. 1942. The reproductive capacity of plants. G. Bell, London.

Schoener, T. W. 1968. Sizes of feeding territories among birds. Ecology 49:123–141.

Sessions, S. K., and A. Larson. 1987. Developmental correlates of genome size in plethodontid salamanders and their implications for genome evolution. Evolution 41:1239–1251.

Shine, R. 1985. The evolution of viviparity in reptiles: an ecological analysis. Pages 605–694 *in* C. Gans and F. Billet, eds. Biology of the Reptilia. Wiley, New York.

Simpson, G. G. 1944. Tempo and mode in evolution. Columbia University Press, New York.

Systematics Agenda 2000. 1994. Systematics Agenda 2000: charting the biosphere. Technical report. New York: Systematics Agenda.

Turelli, M., B. S. Cooper, K. M. Richardson, P. S. Ginsberg, B. Peckenpaugh, C. X. Antelope, K. J. Kim, et al. 2018. Rapid global spread of *w*Ri-like *Wolbachia* across multiple *Drosophila*. Current Biology 28:R269–R271.

Turelli, M., J. H. Gillespie, and R. Lande. 1988. Rate tests for selection on quantitative characters during macroevolution and microevolution. Evolution 42:1085–1089.

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

Waterman, T. H. 1975. Expectation and achievement in comparative physiology. Journal of Experimental Zoology 194:309–344.

Yang, Z. 2014. Molecular evolution: a statistical approach. Oxford University Press, Oxford.

Zuckerkandl, E., and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. Pages 97–166 *in* M. Kasha and B. Pullman, eds. Evolving genes and proteins. Academic Press, New York.

Editor: Daniel I. Bolnick

H.H.NICHOLS, SC.
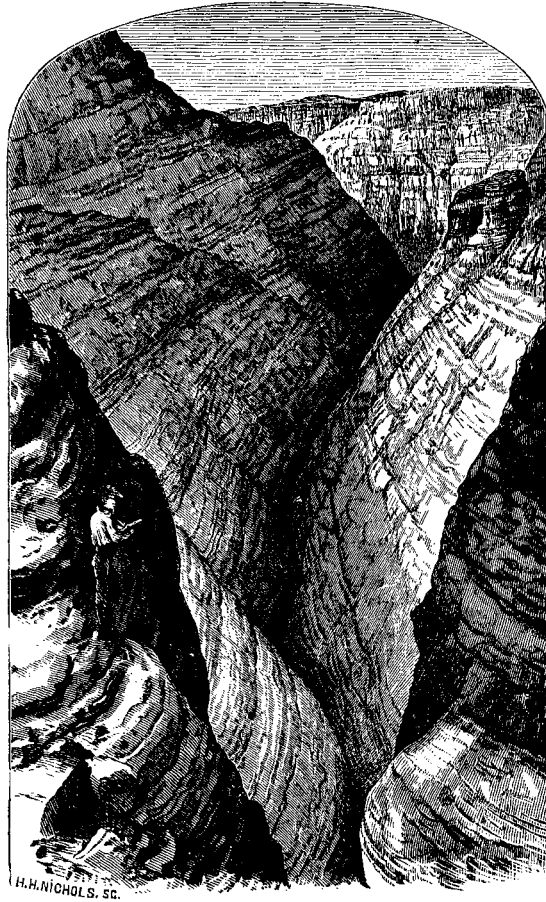
"The narrative is one of the most thrilling records of personal adventure we have ever read; the interest of the reader is intense from the first to the last page, the story being told in a modest, unpretending way, so that the dangers do not seem exaggerated, and the impression produced by the rare exhibition of courage and endurance is not lessened by any straining for effect in the words of the narrator." From the review of Powell's *Exploration of the Colorado* (*The American Naturalist*, 1876, 10:102–105).