Annual Review of Ecology, Evolution, and Systematics

Causes and Consequences of Apparent Timescaling Across All Estimated Evolutionary Rates

Luke J. Harmon,[1] Matthew W. Pennell,[2] L. Francisco Henao-Diaz,[2] Jonathan Rolland,[3] Breanna N. Sipley,[4] and Josef C. Uyeda[5]

[1]Institute for Bioinformatics and Evolutionary Studies (IBEST) and Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844, USA; email: lukeh@uidaho.edu

[2]Department of Zoology and Biodiversity Research Centre, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

[3]Laboratoire Evolution et Diversité Biologique, CNRS, UMR5174, Université Toulouse III–Paul Sabatier, 31062 Toulouse, France

[4]Program for Bioinformatics and Computational Biology, University of Idaho, Moscow, Idaho 83844, USA

[5]Department of Biological Sciences, Virginia Tech University, Blacksburg, Virginia 24061, USA

Annu. Rev. Ecol. Evol. Syst. 2021. 52:587–609

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

https://doi.org/10.1146/annurev-ecolsys-011921-023644

Keywords

speciation, extinction, genetics, phenotypes, microevolution, macroevolution

Abstract

Evolutionary rates play a central role in connecting micro- and macroevolution. All evolutionary rate estimates, including rates of molecular evolution, trait evolution, and lineage diversification, share a similar scaling pattern with time: The highest rates are those measured over the shortest time interval. This creates a disconnect between micro- and macroevolution, although the pattern is the opposite of what some might expect: Patterns of change over short timescales predict that evolution has tremendous potential to create variation and that potential is barely tapped by macroevolution. In this review, we discuss this shared scaling pattern across evolutionary rates. We break down possible explanations for scaling into two categories, estimation error and model misspecification, and discuss how both apply to each type of rate. We also discuss the consequences of this ubiquitous pattern, which can lead to unexpected results when comparing rates

over different timescales. Finally, after addressing purely statistical concerns, we explore a few possibilities for a shared unifying explanation across the three types of rates that results from a failure to fully understand and account for how biological processes scale over time.

1. INTRODUCTION

Evolutionary rates play a central role in the quest to bridge micro- and macroevolution (Gingerich 1983, 2009; Hansen & Martins 1996; Rolland et al. 2018). The central question is whether rates of microevolutionary changes that we measure over a few generations are sufficient to explain the diversity of life that has accumulated on the planet over the past 4 billion years (Kinnison & Hendry 2001). Or, in slightly different terms, can we explain all the diversity of life via models of population and quantitative genetics extrapolated along the branches of the tree of life (Charlesworth et al. 1982, Estes & Arnold 2007)? The opposing point of view is that we need special macroevolutionary phenomena, such as species selection or times of dramatically accelerated diversification, to explain broadscale phenomena (Erwin 2000, Gould & Eldredge 1972, Hendry & Kinnison 2001, Reznick & Ricklefs 2009, Rolland et al. 2018). Measuring and comparing rates of evolution across timescales could provide some resolution for this debate.

Empirical measurements of evolutionary rates have revealed a remarkably consistent pattern of scaling across micro- and macroevolutionary time. Rate estimates depend predictably on the timescale over which they are measured, with the highest rates measured over the shortest time interval (Henao Diaz et al. 2019, Ho et al. 2011, Uyeda et al. 2011). Across rates of molecular, trait, and lineage diversification, the fastest rates tend to be measured over short time intervals, whereas rates measured over longer time intervals tend to be slower (**Figure 1**). As a consequence, patterns of change over short timescales predict that evolution has tremendous potential to create variation, and that potentiality is barely tapped by the present-day diversity of life (Gould 1985). In other words, the short-term potential of evolution, in terms of rates of change of DNA sequences, traits, and the accumulation of species, is much greater than what is manifest over long timescales. The causes of this ubiquitous scaling pattern remain mysterious and pose a challenge to the unification of evolutionary studies.

There is one obvious candidate explanation for this scaling pattern: Perhaps evolution is constantly accelerating. In a world where evolution is speeding up through time, we are making our measurements when all evolutionary rates are at their long-term maximum. There are many unique aspects of the world today, from human impacts to maximal levels of biodiversity, and these could potentially explain long-term patterns of increasing rates. Indeed, some evolutionary dynamics seem to be changing in the Anthropocene (Otto 2018). However, there are a few reasons why one might be skeptical of this as an explanation for rate scaling. One primary counterargument is that the patterns of rate scaling are not limited to extant taxa and also appear in the fossil record (Gingerich 1983, Henao Diaz et al. 2019, Uyeda et al. 2011); in other words, elevated results are common whenever time intervals are short, regardless of whether that short time interval is particularly close to the present day. Instead, we view this pattern as an artifact caused by our failure to accurately measure and model the world. Most of our review, then, is an analysis of this artifact, although we do return to this idea of acceleration in Section 5.1.

In this review, we promote a view of evolutionary rates as model parameters whose interpretation depends on the accuracy and adequacy of the model used to estimate rates (Hunt 2012). To that end, rate scaling can come from two main sources: estimation error, which involves statistical error in the data or in the way the data is fit to a model, and model misspecification, in which the

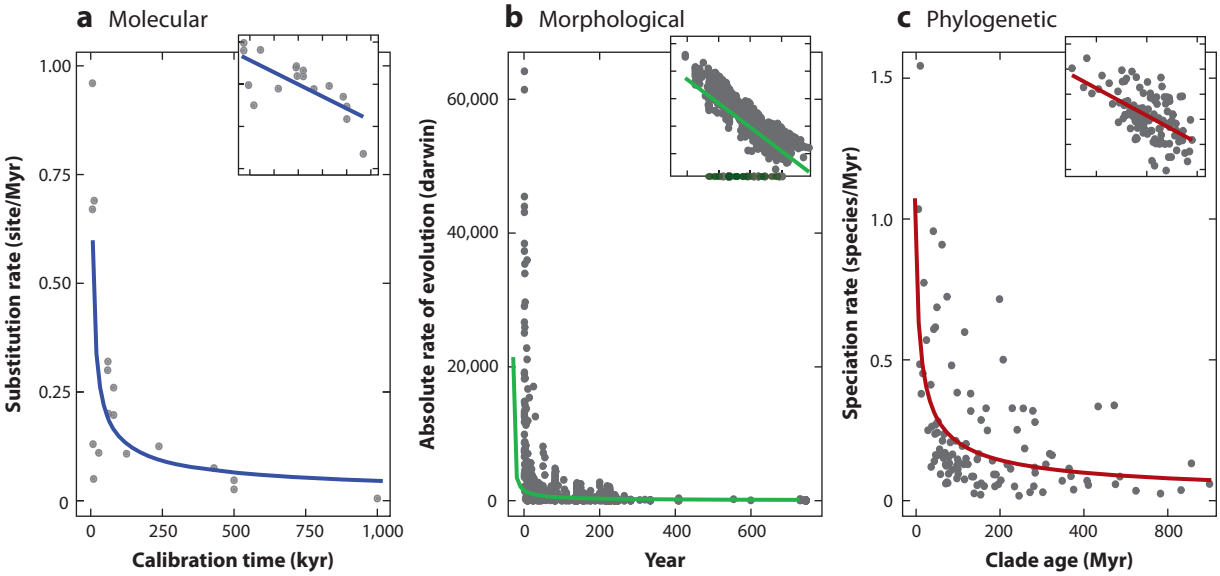

Figure 1

Rate-scaling patterns showing rate (*y*-axis) versus time interval (*x*-axis); insets are on a log-log scale. (*a*) Rates of molecular evolution versus time interval. Estimated mean mitochondrial substitution rates compared with their calibration intervals or oldest sampled tip. Rates were calculated on contemporaneous and ancient sequences using Bayesian or coalescent methods under the best substitution model. Data compiled from Ho et al. (2007). (*b*) Rates of trait evolution versus time interval. Absolute morphometric divergence rate in darwins compared with its measurement time interval estimated for 30 orders among tetrapods, arthropods, mollusks, radiolarians, and foraminifera. Data replotted from Uyeda et al. (2011). (*c*) Rates of speciation versus time interval. Estimated speciation rates from 104 time-calibrated molecular phylogenies compared with a clade's crown age (Henao Diaz et al. 2019); diversification rates were calculated on representatives of seed plants, ferns, vertebrates, and invertebrates using Bayesian analysis of macroevolutionary mixtures (BAMM) (Rabosky 2014) on representatives of seed plants, ferns, vertebrates, and invertebrates.

model does not capture the main behavior of the evolutionary process. Both of these phenomena can lead to rate-scaling patterns, and authors sometimes argue past one another due to a lack of understanding of these potential sources of scaling patterns. We advocate, then, a direct attempt to consider estimation error and model misspecification when considering rates (see the sidebar titled Statistical Issues and Misconceptions Associated with Scaling of Rates).

STATISTICAL ISSUES AND MISCONCEPTIONS ASSOCIATED WITH SCALING OF RATES

Analyses of rate scaling with time intervals are subject to a particular type of estimation error, variously referred to as spurious self-correlation (Jackson & Somers 1991, Kenney 1982, Prairie & Bird 1989) or as a mathematical coupling of data (Moreno et al. 1986). The issue arises when one variable is plotted or regressed against another but both share a common mathematical term (Long 1980). In this example, an evolutionary rate (amount of change/time) is plotted against its own denominator (time) (Sheets & Mitchell 2001). The shared component of these coupled variables can then lead to peculiar scaling relationships (Long 1980).

(*Continued*)

One might be tempted to simply avoid coupled variables. However, in many cases, correlations between ratios and their own denominators are of interest, even in spite of their mathematical coupling (Long 1980). The clearest cases are when the ratio itself has particular meaning, as here when comparing evolutionary rates (Prairie & Bird 1989). However, we cannot ignore the presence of shared terms in our regression and scaling plots (Long 1980).

One can find three important take-home points in the extensive literature on this topic. First, we should expect a relationship between rates (change/time; Y/X) and time (X)—a scaling rule—whenever the original variables (time and amount of change) show no relationship (Kenney 1982). Such a finding redirects the question to, Why is there no correlation between X and Y? In the case of evolutionary rates, we can evaluate this possibility by analyzing the amount of evolutionary change (or species diversity accumulated) versus time. We find that in some cases—most commonly in the case of lineage diversification—there is little or no relationship between the original variables, such that the scaling rule emerges as a consequence of their shared term (Sheets & Mitchell 2001, Wiens 2011). This lack of a relationship between the original variables, though, does not provide a mechanistic answer or even tell us whether the result is due to estimation error or model misspecification. Additionally, sometimes there is a relationship between X and Y, and even in that case, peculiar scaling patterns can still follow due to coupling.

Second, measurement and sampling error, particular types of estimation errors, can have a nonrandom effect on comparisons of rates across time, resulting in spurious relationships under some circumstances (Long 1980). Focusing on the case in which we relate Y/X to X, when measurement and/or sampling error is particularly high in X compared with Y, then negative scaling relationships can emerge due to error alone. Therefore, it is particularly important to explicitly account for the measurement of both variables when considering scaling relationships. In the specific case of rates versus time, we should expect problems to emerge when the error in estimating time is greater than the error in estimating the amount of change that has taken place (Long 1980). We see echoes of this explanation for all three rates in Section 4 and conclude that this factor plays a prominent role in explaining the ubiquity of scaling across rates.

Third, the fundamental question of any particular scaling pattern is, Why does the relationship take the particular form that it does? That is, why does the scaling plot show the particular shape and slope (Estes & Arnold 2007, Gingerich 2019)? We believe that this answer is best addressed via assessment of model adequacy (Brown & Thomson 2018). For evolutionary rates, a wide range of evolutionary models predict change to accumulate with time. The lack of relationship between change and time discussed in the first point deviates dramatically from many simple evolutionary models, whether they be based on constant-rate birth–death (diversification), time-homogeneous stochastic substitution models (genetics), or Brownian motion (traits). We can explore this deviation by considering the rate-scaling properties of macroevolutionary models as a consequence of model misspecification, and proceed either analytically or via simulation (Pennell et al. 2015). For example, due to model misspecification, negative relationships may occur that are stronger than expected given the statistical properties of ratios even in cases where X and Y are uncorrelated. In those cases, one can still detect deviations from a model—model inadequacy—by comparing the observed slope of the scaling with that expected under the null model (Henao Diaz et al. 2019). One must then search for a different model that is adequate to explain the observed pattern. This type of approach also allows one to separate scaling due to estimation error—that is, error due to the act of measuring nature to estimate evolutionary rates—from scaling due to model misspecification. This distinction reflects the extremely common desire in the literature to uncover the role of biological processes (as opposed to measurements and statistical analyses) in driving observed patterns.

Our goal is to provide a unified statistical and biological perspective on the scaling of evolutionary rates of molecular evolution, trait evolution, and diversification. We first review how each type of rate is typically measured and then review the literature on rate scaling for that type of rate. We discuss common explanations for scaling patterns, separating them into explanations based on estimation error and model misspecification. In Section 6, we discuss the practical consequences

of rate-scaling patterns. Scaling patterns have strong impacts on testing hypotheses about evolutionary rates, especially when comparing different time periods. It is common for researchers to compare speciation rates, for example, across lineages of different ages, and all such comparisons are compromised by ubiquitous rate-scaling patterns. In Section 7, we conclude and attempt to characterize and identify the potential causes of rate scaling. Although a definitive mechanism is still lacking, we highlight a range of explanations that are likely candidates and discuss how they can be evaluated. We hope that this review helps clarify work comparing rates across timescales, suggests ways to improve models used to estimate rates, and, eventually, helps researchers identify and describe the biological processes that affect long-term evolutionary patterns.

2. ESTIMATING RATES OF EVOLUTION

Estimating an evolutionary rate seems straightforward at first. After all, isn't a rate simply the amount of change divided by the time over which that change occurred? In this simplistic view, rates are simply summary statistics of the net change divided by time. However, while the measurements themselves are descriptors of net rate over a certain timescale, there is no particular reason to think that rates are directly comparable over different timescales (Gingerich 2019). Indeed, it is important to remember that a particular pattern of rate scaling may be explained by many different models and dynamics of change, both biological and otherwise. Thus, inference of process from scaling patterns of net change over time must be approached with caution, with careful consideration given to alternative sources of rate scaling.

We draw a distinction between these summary statistics of net rate and parameter estimates of rates obtained by fitting process-based models (Hunt 2012). Process-based models generally assume constant (or at least locally constant), instantaneous rates of change that are thus independent of timescale—so long as the model assumed is adequate and uniquely identifiable. We discuss this in more detail in Sections 4 and 5.

2.1. Estimating Rates of Molecular Evolution

Following previous literature (Ho et al. 2005), when we discuss rates of molecular evolution, we are focused on the nucleotide substitution rate, defined as the rate at which new mutations become fixed in populations (Holmes et al. 2016). Rates of molecular evolution could be measured simply by comparing two sequences, counting the number of differences between them, and dividing by time. This approach would underestimate rates, especially between older sequences, by undercounting when multiple substitutions have occurred at a single site (Yang 2006). Even in the early days of modeling sequence evolution, this was recognized as a potential problem (Jukes & Cantor 1969). Since then, rates of molecular evolution have generally been based on fitting continuous-time Markov models (Gillespie 1986).

Continuous-time Markov models work by considering transitions among possible character states at each locus independently. For DNA, these states are represented by A, C, T, and G. Transitions between these states are described by a rate matrix (usually denoted Q) that contains, as off-diagonal elements, the rates of transition from each state to the three others (e.g., q_{AG} denotes the transition rate from state A to state G),

$$Q = \begin{bmatrix} -q_A & q_{AG} & q_{AC} & q_{AT} \\ q_{GA} & -q_G & q_{GC} & q_{GT} \\ q_{CA} & q_{CG} & -q_C & q_{CT} \\ q_{TA} & q_{TG} & q_{TC} & -q_T \end{bmatrix}.$$

By convention, the diagonal elements are set so that the rows sum to zero. That is, for example, $q_A = q_{AG} + q_{AC} + q_{AT}$ (Yang 2006).

In the molecular evolution literature, various restricted forms of this Q-matrix go by different names (Yang 2006). For example, if all the rates are set equal to one another, the model is called the Jukes-Cantor model (Jukes & Cantor 1969); if each pair of states has a unique rate, but the forward and reverse rates are equal (e.g., the rate of $A \rightarrow T$ is equal to the rate of $T \rightarrow A$), then this is called the general-time reversible model (Yang 2006). Modifications also allow rates to vary across sites or across branches in the tree.

To obtain an overall rate, one estimates the best multiplier, r, such that the tree branch length times r, applied to a normalized Q matrix best predicts patterns of sequence divergence. If time is measured in absolute units, say millions of years, then r gives average substitution rates as the expected number of substitutions per site per million years. Using this protocol, one can calculate rates for any specified model of molecular evolution (Yang 2006). Note that the procedure here also works for any set of discrete characters evolving on a phylogeny under a Markov process, including phenotypic traits (Lewis 2001). However, rate estimates for morphology are more often measured from continuously valued data, which we address in Section 2.2.

2.2. Estimating Rates of Trait Evolution

In paleontology and evolutionary biology, it is common to study the rates of evolution of continuously valued traits (Gingerich 2009, Hunt 2007). There are two main ways to estimate evolutionary rates: either directly by using one of two expressions that measure change in trait values over time or indirectly by estimating a model parameter that represents the instantaneous rate of change (Harmon 2014).

There are two classic units for measuring evolutionary rates, darwins and haldanes (Gingerich 1983, Haldane 1949). These can be viewed as either summary statistics of net rate over a time interval or an estimate of instantaneous evolutionary rate under a model of strict linear directional change. Darwins express change in proportional change per million years, and haldanes express change in units of within-population phenotypic standard deviation per generation—the scale that is more relevant when considering genetic constraints on evolutionary change (Harmon 2014). These two metrics have been widely used to quantify rates across a range of scales, from single generations to millions of years over phylogenies (Hendry & Kinnison 1999). Importantly, however, both darwins and haldanes can be interpreted only as an instantaneous rate under the highly restrictive scenario of a linear trend in trait change over time. Under many other models, both of these rate metrics are expected to scale with the time interval over which they are measured.

In most modern studies, rates of evolution are calculated as the parameters of a model of evolution. The most common of these models is Brownian motion, a model in which trait change follows a random walk through time (Felsenstein 1973). Although sometimes equated with genetic drift, Brownian motion can result from a range of evolutionary scenarios, including randomly varying selection, drift, and other factors (Hansen & Martins 1996). For example, if populations are under strong selection following a peak on the adaptive landscape, and that peak is changing via Brownian motion, then we expect traits to also evolve following that same Brownian motion model (Hansen & Martins 1996). Brownian motion can be described with two parameters: the starting value, often denoted z_0, and the rate parameter, σ^2. The latter parameter, σ^2, predicts the differences that one will find among pairs of species after a certain amount of time has elapsed, and it is this parameter that is often described as the rate of evolution. It is worth noting, though,

that the Brownian motion rate is not strictly comparable to haldanes and darwins. For example, because both darwins and haldanes are constant only under a linear trend, when these metrics are calculated from data evolved under a Brownian motion model, both units show a negative dependence on the time interval over which they are measured (Gingerich 1983, Hendry & Kinnison 1999). Likewise, if species change following a constant linear trend, rates of evolution estimated assuming a Brownian motion model will have rate-scaling properties (Gingerich 2019).

2.3. Estimating Rates of Diversification

The simplest intuitive way to measure rates of speciation and extinction through time is to count the number of such events over some interval and divide by the length of time. Although such an approach does have some precedence in paleontology (Foote 2003), simply counting events ignores the fact that the number of speciation and extinction events in an interval almost certainly depends on the number of lineages alive in that interval (Raup 1985). Thus, most modern estimates of speciation and extinction rates are measured on a per-lineage basis and are based on birth–death models of diversification (Foote 2003, Raup 1985).

Under a birth–death model, each lineage has an origination rate λ and an extinction rate μ (and thus a collection of n such lineages has overall rates of $n\lambda$ and $n\mu$ (Kendall 1948). Speciation and extinction events are assumed to occur randomly in time (i.e., are the result of a Poisson process), such that the waiting times between events are exponentially distributed. If λ and μ are not equal, then birth–death models predict that species numbers will change exponentially through time. For this reason, some paleontological estimates are based on dividing the log-transformed number of events by the time interval, which then estimates λ and μ. Related metrics for phylogenetic trees involve a method of moments estimator (Magallon & Sanderson 2001) that accounts for the age and diversity of a clade. This estimator, in its simplest form for a pure-birth model with stem age t and extant diversity n, is $\lambda = \log(n)/t$. Modifications for including extinction, correcting for sampling, and accounting for crown age rather than stem age are also possible (Magallon & Sanderson 2001). The similar form of these equations from paleontology and phylogenetics emphasizes the shared assumption of a birth–death model.

In both paleobiology and phylogenetic biology, most current diversification rate estimation takes the form of estimating the parameters of such birth–death models using maximum likelihood or Bayesian inference (Harmon 2019, Nee 2006). This approach has the main advantage of using all the information about births and deaths stored in the phylogenetic tree. Maximum likelihood–based approaches for birth–death models have an additional complexity that the proper equation—and thus one's result—depends precisely on what one conditions: the age of the crown (the most recent common ancestor of extant lineages), the age of the stem (the most recent common ancestor of all lineages, extant and extinct, in a clade), or the survival (or not) of the stem or crown lineage(s) to the present day (MacPherson et al. 2020, Stadler 2013). This conditioning is essential because the process of extinction and the lineages that we sample for phylogenetic reconstruction are intimately linked (Höhna et al. 2011).

Many variations of birth–death models have been developed, including those that allow variation in rates through time (Etienne et al. 2012, Magee et al. 2020, Morlon et al. 2011, Rabosky 2014, Rabosky & Lovette 2008) and across clades (Alfaro et al. 2009, Höhna et al. 2019, Rabosky 2014) and those that allow rates to depend on organismal traits (FitzJohn 2010, Maddison et al. 2007) and/or the environment (Cantalapiedra et al. 2014, Condamine et al. 2019, Goldberg et al. 2011, Rolland et al. 2014). Each of these process-based models includes rate parameters that can potentially be estimated from the branching structure of phylogenetic trees.

3. TIME DEPENDENCE OF ALL EVOLUTIONARY RATES

All three rates of evolution, molecular, trait, and diversification, scale with time, with the most rapid rates observed over the shortest time periods (**Figure 1**). This pattern has generated discussion in the literature in all three cases, with controversy over both the existence of and the explanation for rate-scaling patterns. Despite the striking similarity of all three relationships, we are unaware of any paper giving a collective view of all three scaling patterns, so we present that here.

3.1. Time Dependence of Rates of Molecular Evolution

Rates of molecular evolution have been observed to scale with time. The earliest studies comparing divergence and time suggested complex patterns (reviewed in Gillespie 1986). The first suggestion of a consistent overall trend toward negative rate scaling (that we are aware of) traces to Wayne et al. (1991), who measured faster rates of molecular evolution when comparing mammal lineages with younger divergence times. Later, similar patterns were found in other organisms, such as birds (e.g., García-Moreno 2004) and mammals (Albà & Castresana 2005; see also Elhaik et al. 2005). The systematic pattern of molecular rates scaling with time was described by Ho et al. (2005, 2011). These studies show that estimates of rates are accelerated for comparisons of very recent divergence. In contrast to the other rate-scaling patterns discussed in Sections 3.2 and 3.3, rates of molecular evolution are thought to follow a J-shaped curve, flattening out after approximately two million years rather than continuing to decline over the longest timescales (Penny 2005).

This main result of rate scaling pointed out by Ho et al. (2005) has attracted controversy, with one review calling it a "tempest in a teapot" (Bandelt 2008, p. 1). Much of this argument has centered around whether the scaling patterns reflect statistics or interesting biological properties of the system. For example, Emerson (2007) argues that errors in estimating sequence divergence, biased sampling, and errors in calibration together explain the apparent scaling. In other words, Emerson attributes the entirety of the scaling shown by Ho et al. (2005) to various types of estimation errors, rendering the pattern ultimately uninteresting in terms of learning about biology. A follow-up from Ho et al. (2011) used substantially better data and improved analyses, again arguing in favor of rate scaling as evidence for interesting and general biological processes acting over long timescales, in other words, as a consequence of model misspecification. This approach was in turn disputed by Emerson & Hickerson (2015), again arguing for the primacy of estimation error. It is worth emphasizing, though, that all rebuttals do not deny the empirical pattern; instead, much of this discussion relates to whether this pattern is an artifact (i.e., driven by estimation error) or reflects some nonmodeled biological process of interest (i.e., driven by model misspecification). The bulk of the discussion of the latter argument centers around the distinction between mutation rates and long-term substitution rates and in turn relates to purifying selection and the time needed to rid populations of segregating neutral and deleterious mutations (Penny 2005). This debate, to our knowledge, remains unresolved.

3.2. Time Dependence of Rates of Trait Evolution

In a classic study of the rate scaling of trait evolution in macroevolution, Gingerich (1983) collated a large dataset on evolutionary rates spanning a wide range of time intervals, from the fossil record to contemporary rates of evolution in the laboratory. Plotting all those rate estimates together revealed a scaling pattern that spanned seven orders of magnitude of time. Gingerich showed that evolutionary rates correlated strongly with the time interval over which they were measured, with the highest rates estimated over the shortest intervals; later analyses have confirmed these results (Uyeda et al. 2011).

Since this key paper, several reviews and compilations of rates of evolution have been published in the literature. A large amount of overlap in the data is used for these reviews, as one original data compiled by Gingerich (1983) has been reused and added to over time. In general, follow-up studies confirm the basic pattern but differ in interpretation. For example, Kinnison & Hendry (2001) combined the Gingerich data with information about evolutionary rates gathered from more recent literature. They confirmed the negative scaling pattern and used a suite of randomization tests to suggest that the pattern was stronger than expected due to estimation error. Likewise, Uyeda et al. (2011) compiled data on evolutionary rates both from the Gingerich dataset and from phylogenetic data in the tree of life. Their paper showed clear patterns of rate scaling. Uyeda et al. (2011) emphasize an explanation based on model misspecification, demonstrating that trait divergence among species shows fairly consistent differences, regardless of time since divergence, for timescales less than one million years, accumulating only major differences beyond that.

3.3. Time Dependence of Diversification Rates

Finally, diversification rates scale with the time interval over which they are measured, in a way strikingly similar to rates of molecular and trait evolution. This pattern has been described in the literature several times (Linder 2008, Magallon & Sanderson 2001, McPeek & Brown 2007, Ricklefs 2006, Scholl & Wiens 2016). Two papers focused particular attention on the scaling pattern but came to opposing conclusions about the cause. Marin & Hedges (2018) identified an apparent acceleration in rates of diversification in young clades across a diversity of groups. They attributed this pattern to estimation error, in particular the undersampling of genomes and consequent underestimation of the ages of young nodes in the tree of life. Henao Diaz et al. (2019) showed a similar pattern of apparent rapid diversification among young clades. Their paper also showed that the pattern extended to the fossil record, with rates of formation and extinction of genera inversely related to the time interval over which they were measured (Foote 1994, 2005). This inclusion of fossils precludes any common explanation based on estimation error associated with inferring phylogenetic trees from gene sequence data.

4. RATE SCALING AS A CONSEQUENCE OF ESTIMATION ERROR

We consider estimation error to be statistical errors that arise during analysis even when the assumed biological process adequately describes nature. We focus in particular on three sources of estimation error: biased sampling, measurement and sampling error, and parameter nonidentifiability. With biased sampling, the entities that are compared to calculate rates are not a random subset of the total population. Measurement and sampling error, by contrast, can result from several types of errors, such as finite sample size or instrumental error, while measuring a desired quantity from observed data, whether species traits, species richness, or the timing of evolutionary events. Finally, parameter nonidentifiability occurs when a rate parameter is misestimated because it is not uniquely identified or only weakly estimable for a given dataset and model. All three phenomena can affect rate estimates, regardless of the type.

4.1. Estimation Error and Rates of Molecular Evolution

DNA sequences can be reconstructed with error due to artifacts from amplification and/or sequencing. Because sequencing errors should be random, they might not show any relationship to time since divergence, in turn leading to negative scaling of rates with time (Ho et al. 2011) (see

the sidebar titled Statistical Issues and Misconceptions Associated with Scaling of Rates). Such arguments are particularly compelling for ancient DNA, which might be especially susceptible to errors in sequence reconstruction (Ho et al. 2011). Likewise, divergence times are also estimated with error (Arbogast et al. 2002). As argued by Weir & Schluter (2008), this error tends to lead to scaling artifacts when errors in estimating divergence times (X in the sidebar) are greater than errors in estimating sequence divergence (Y in the sidebar).

4.2. Estimation Error and Rates of Trait Evolution

Both measurement error and biased sampling can lead to errors in estimating rates of trait evolution that can again, in some cases, result in scaling patterns with time. Ignoring measurement error will, on average, inflate differences among species, thereby elevating evolutionary rates. This effect is most pronounced over the shortest time intervals, resulting in highly inflated rates when age differences are small (Cooper et al. 2016, Silvestro et al. 2015). Thus, failing to account for measurement error can lead to rate-scaling patterns. This argument applies equally well to time-series and phylogenetic analyses, which frequently consider species means without considering measurement error (Kostikova et al. 2016, Silvestro et al. 2015). Measurement error can also affect our calculation of elapsed time and the timing of branching events in the phylogenetic tree. Similar to trait differences, ages are estimated with error, and that error is sometimes not accounted for in calculations of rates of trait evolution. Unlike the case of measurement error for species' traits, random error in branch lengths should not lead to negative scaling of rates through time. Rate scaling could result if ages were systematically biased (underestimated for younger comparisons, overestimated for older ones, or some combination thereof). Rate scaling can also arise when there is error in both the numerator and the denominator of a rate, but the error in the denominator is greater than that in the numerator (Long 1980). This seems highly likely for trait evolution and almost certainly makes some contribution to the observed scaling patterns.

Biased sampling can also have an impact on scaling for rates of trait evolution. We can see one example by Gould (1984, p. 994), who called the pattern of trait rates scaling with time a "psychological and mathematical artifact." Gould's main argument was that paleontologists were biased in what they measured when calculating rates. In the fossil record, two lineages might be connected in a time-series dataset only if they are within some threshold of similarity. Any lineages that are too different, then, are not used to calculate evolutionary rates. This difference could lead to rate scaling through systematically underestimated rates over long timescales, as old, highly divergent lineages are not likely to be included in rate calculations. Any pattern of biased sampling that affects the expected distribution of species' differences could lead to scaling, whether rates are calculated from phylogenetic trees or fossil data.

4.3. Estimation Error and Rates of Diversification

Finally, timescaling in diversification rates may be due to estimation error of either the rates themselves or the divergence times/clade ages. As mentioned in Section 2.2, one possible explanation for this scaling relationship between rates and time is estimation error—either of the rates themselves or of the divergence times/clade ages. For instance, it is well recognized that even with a lot of sequence data, there is still considerable uncertainty in divergence time estimates, particularly deep in the tree (Graur & Martin 2004, Revell et al. 2005). Henao Diaz et al. (2019) show in the supporting information of their article that such errors could lead to scaling patterns, echoing an early point by Ricklefs (2006). A similar argument applies to misestimation of time intervals in the fossil record. However, two considerations suggest that misestimated branch lengths are not the full explanation for the observed patterns. First, the pattern remains despite substantial improvements in

divergence time estimation, even as the size of analyzed clades has greatly increased (Henao Diaz et al. 2019). Second, simulations from Henao Diaz et al. (2019) suggest that the scaling patterns due to estimation error of branch lengths are not sufficient to explain the negative slope as observed. Fortunately, it appears that estimated rates of diversification are generally quite robust to misspecification of the clock model used to date nodes (Sarver et al. 2019, Wertheim & Sanderson 2011).

Diversification rate scaling can also be influenced by biased sampling. For phylogenetic trees, one must decide on the clades for analysis, breaking up a single tree of life into monophyletic clades. We do not yet know how this selection of clades affects diversification rates, although the scaling pattern seems to persist through alternative selection schemes in Henao Diaz et al. (2019). One clear effect is due to the exclusion of small clades: Trees are unlikely to be constructed or considered for particularly species-poor (or monotypic) clades. If these depauperate clades are not included in analyses, we should expect a biased view of rates across time. Other effects on scaling are possible based on how clades are defined by crown age or stem age (Stadler et al. 2014). These issues with sampling are clearly present with paleontological data (Hopkins et al. 2018). One expects these biases to lead to scaling to the extent that it breaks the association between the length of the time interval and the number of observed diversification events.

4.4. Model Adequacy, Estimation Error, and the Limits to Parameter Estimation

The most promising path forward is to improve our ability to both estimate measurement error and account for it in our modeling approaches (Houle et al. 2011). Significant progress has been made in nearly every modeling type to account for measurement error. For example, several methods for estimating evolutionary rates from Brownian motion models use external estimates of observational measurement error to provide unbiased estimates of rate (Felsenstein 2008, Ives et al. 2007); other approaches account for sequencing error in rates of molecular evolution (Ho et al. 2015, Johnson & Slatkin 2008) and unsampled lineages in rates of diversification (Höhna 2014). Although the statistical procedure can be straightforward, determining what values to use as measurement error can be confused when biological error and statistical sampling properties become conflated (Hansen & Bartoszek 2012). This issue is explored more fully in Section 5. Sampling bias can be more challenging, but some approaches account for sampling when fitting evolutionary models (FitzJohn 2010). While corrections for sampling error are still often omitted from comparative analyses, we believe there is growing and widespread recognition that such biases and errors must always be included in a comparative analysis (Silvestro et al. 2015), and in general, we think researchers are largely aware and cautious of this sort of model inadequacy.

Even when data are measured without error and models are adequate, there are limits to parameter estimation. For a simple example, consider a set of molecular sequences for which every site is completely saturated so as to remove all phylogenetic signal. This set of sequences, regardless of branch length or time interval, is consistent with any sufficiently high substitution rate (Mossel & Vigoda 2005). The resulting flat likelihood surface estimates rate parameters subject to the vagaries of a particular optimization algorithm and largely invariant to time, resulting in slower rates estimated over longer time intervals. Extending this problem to model selection, we would be remiss to not mention the dark corridors of uncertainty in model-fitting stemming from model nonidentifiability. Nonidentifiable models are defined as two or more models that produce observationally equivalent distributions of data and result in rate parameters that cannot be uniquely identified. For example, Louca & Pennell (2020) have challenged the entire enterprise of estimating rates of diversification from molecular phylogenies. They prove that for the case of any homogeneous birth–death process, in which the process is the same for all lineages at any given time point but can change through time, neither λ nor μ nor any combination thereof (e.g., net

diversification rate $r = \lambda - \mu$) can be uniquely defined or estimated (Louca & Pennell 2020). This is due to the inherent nonidentifiability of birth–death processes: For any time variable process, where λ and μ vary through time according to some mathematical functions, an infinite number of alternative functions (called a congruence class) generate an identical expected distribution of tree sizes and branching patterns that cannot be distinguished using maximum likelihood or any other method using phylogenetic trees (Louca & Pennell 2020). This result suggests that interpreting model fit should be done with caution. In particular, current methods select, somewhat arbitrarily, particular parameter values from the congruence class. Similar congruence classes exist for continuous trait models of ultrametric phylogenies [e.g., Brownian motion versus Brownian motion with a trend (Felsenstein 1988), Ornstein-Uhlenbeck models versus accelerating change models (Hansen & Martins 1996, Slater et al. 2012, Uyeda et al. 2015)]. It is likely that similar congruence classes exist across all macroevolutionary models, not only diversification models. These problems blend into Section 5, in which we discuss the consequences of such model misspecification.

5. RATE SCALING AS A CONSEQUENCE OF MODEL MISSPECIFICATION

There are countless ways that our macroevolutionary models might be incorrect, and many of these ways could lead to scaling patterns. Some of them are well summarized in the previous literature on this topic (Ho et al. 2011, Marin & Hedges 2018, Penny 2005). In this review, rather than comprehensively cover all the ways that our simple models may be wrong, we focus on four possible categories of model misspecification. We direct our focus on shared explanations that may explain all three rate-scaling patterns in a way that is consistent with contemporary evolutionary theory.

5.1. Evolution Is Accelerating Toward the Present Day

Perhaps evolution is speeding up, with all rates peaking at the present day. Such a pattern could explain why everything we observe over short timescales from our present-day point of view appears rapid, and why including longer timescales stretching back into the distant past results in slower rates. This explanation is potentially viable for any datasets that rely on the present day for sampling, such as analyses based on phylogenetic trees. Perhaps the most compelling argument in support of this view relates to potentially accelerated rates of evolution, speciation, and extinction in the Anthropocene as humans alter the world around them (Otto 2018). Additionally, some researchers have suggested that evolutionary rates might accelerate as the total biodiversity of the earth increases through time (Emerson & Kolm 2005). In our view, both of these hypotheses fail to explain the pattern discussed here on two counts. First, there is no expectation that evolution should speed up uniformly. One might anticipate that evolution is faster in places where diversity is higher (Emerson & Kolm 2005) or where human impacts are strongest (Otto 2018); one would not expect a consistent pattern in all places and stretching back over such long timescales. Second, any explanation based on acceleration toward the present day cannot account for the presence of similar scaling patterns for trait evolution and diversification using purely paleontological data (Henao Diaz et al. 2019, Uyeda et al. 2011).

5.2. Evolutionary Entities Are Highly Heterogeneous

We know that evolutionary lineages are highly heterogeneous in a number of ways. All the processes that we are concerned with in this review vary tremendously across clades, with some lineages evolving rapidly and others more slowly. This heterogeneity is why, in molecular evolution, the relaxed clock is often used, modeling rate variation across branches in the tree of

life; why the shape of that tree is unbalanced, reflecting variation in speciation and extinction rates across clades (Mooers & Heard 1997, Ng & Smith 2014); and why the tempo of trait evolution is highly variable among even closely related clades.

It might not be immediately obvious that rate heterogeneity leads to timescaling for the resulting rate estimates. Indeed, if we are able to accurately account for the heterogeneity, then rate estimates should vary but independently of the time interval. However, in practice, heterogeneity is a type of model misspecification by which we fit a sort of average rate across heterogeneity across clades (Yang 2006). In such a case, older clades lead to rate estimates that adequately sample across the heterogeneity, leading to rates that match the long-term average, but as clades get younger, we are more and more likely to sample a part of the tree of life with unusually high rates, unpolluted by the tempering influence of slow lineages. There will be, by chance, slow young clades as well, but the unusually fast young clades will dominate the pattern, leading to the classic negative relationship between rates and time.

5.3. Evolutionary Events Are Clustered in Time

All the models we use to estimate evolutionary rates assume that events occur independently in time. For example, for discrete-event models, such as sequence evolution or birth–death models, the waiting time to the next event is independent of whether an event has just occurred (Harmon 2019, Yang 2006). Likewise, Brownian motion models assume that variance accumulates through time independently of what changes have recently occurred in the population (Felsenstein 1973). Although these assumptions allow the use of Markov models and ease mathematical analysis, they may not be realistic or accurate for modeling evolution.

We suggest that macroevolution may commonly violate this independent events assumption and that macroevolutionary events in particular might be clustered in time, occurring in a pulsed pattern (Foote 2005, Landis & Schraiber 2017). For example, species might experience stasis over relatively long time intervals and then periods of rapid change followed by further stasis (Estes & Arnold 2007). Likewise, speciation and extinction events might be clustered in intervals, perhaps associated with rapid environmental change or major alterations in species interactions (Foote 2005). Clustered events lead to rate scaling because some short time intervals occur during these periods of concentrated change, leading to elevated rates (Sadler 1981). Consistent with this explanation, rate-scaling plots often show tremendous scatter over short time intervals, suggesting that evolution is often, but not always, rapid over short timescales (**Figure 1**). Finally, different patterns of clustering through time could lead to different scaling patterns and could be estimated from the shape of the rate-scaling pattern (Foote 2005). Nonhomogeneous event models (Goldberg & Foo 2020, Hawkes 2018) might be particularly useful in investigating the clustering of evolutionary events. Indeed, models of pulsed evolution show a better fit to some empirical datasets for both trait evolution (Landis & Schraiber 2017) and speciation and extinction rates (Foote 2005), and the success of relaxed molecular clock models (Drummond et al. 2006) suggests pulsed rates of molecular evolution.

5.4. Evolution Is Constrained

Constraints on evolutionary change set bounds that limit the long-term effects of rapid evolution over short timescales (Schwenk 1994). This logic can be applied to all three rates. For molecular evolution, sequences can evolve rapidly over short timescales by accumulating neutral or nearly neutral substitutions (Ohta 1992). However, there are a limited number of such mutations in the vicinity of the current sequence, resulting in rapid saturation of sequence space due to the constraints of natural selection (DeSalle et al. 1987). Some of the genes most commonly used

to estimate evolutionary rates are key metabolic components of cells, critical for function and likely under long-term constraints for functionality. Given this constrained state space, rate scaling emerges for rate parameters if the constraints are not adequately modeled or as a consequence of our inability to estimate rate once complete saturation occurs (Elhaik et al. 2005).

The argument for rate scaling driven by constraint is perhaps most developed for trait evolution. Here, a constrained model of evolution, called the Ornstein-Uhlenbeck (OU) model, is commonly used for analysis (Hansen 1997). This model is often described as Brownian motion, but with a central tendency to make evolution go toward some optimal value (Butler & King 2004). OU models, in general, have four parameters: a rate parameter (σ^2), the constraint parameter (α), an optimal value (Θ), and a starting value z_0. OU models are commonly selected as the best fitting for comparative data compared with various alternatives (Cooper et al. 2016, Harmon et al. 2010, Pennell et al. 2015). If we analyze data generated under a constrained OU model by assuming unconstrained Brownian motion, we will observe rapid change over short timescales, approaching the OU rate parameter, σ^2 (Uyeda et al. 2015). We will also see slower and slower apparent rates over longer times as constraints begin to dampen variance among species compared with the Brownian expectation (Estes & Arnold 2007, Uyeda et al. 2011).

Finally, constrained models are also commonly invoked for diversification. In this case, many authors have suggested that speciation and extinction are diversity dependent, with net diversification rates slowing as species richness nears some carrying capacity (Condamine et al. 2019, Etienne et al. 2012, Harmon & Harrison 2015, Harmon et al. 2003, McPeek 2008, Rabosky & Hurlbert 2015). This pattern is fundamentally different from the other two constraints in that such a process would leave a clear mark in phylogenetic data, with branching events concentrated at the base of the tree of life. However, it remains unclear whether such a pattern is common or rare in comparative data (Pannetier et al. 2021). One can construct scenarios in which a complex pattern of diversity dependence can lead to rate scaling. For example, imagine that all life was generated under a multi-regime diversity-dependent model, in which lineages were subject to carrying capacities based on their close relatives so that diversification rates slow as species accumulate. Imagine also that lineages sometimes escaped those constraints and entered into new regimes with independent carrying capacities (e.g., new adaptive zones). In that case, we would observe high rates of diversification only when considering clades that have just moved into new regimes—and these clades would be disproportionately young (Schluter 2000). Thus, we would see the expected negative scaling of rates with clade age. In any case, there is some debate and uncertainty in the literature about the amount of support for diversity dependence of speciation and extinction.

In all cases, failure to account for, or accurately estimate, the constraints on a process can result in rate scaling for model parameters (**Figure 2**). Models that account for the long-term dynamics of these constraints might lead the way to deeper understanding of the state space of evolutionary change and how the evolutionary process leads to convergence and saturation (**Figure 2**). This evolutionary state space itself may change over time (e.g., Beaulieu et al. 2012, Butler & King 2004). Such models may also behave as the clustered models described in Section 5.3, with rapid evolution concentrated near peak shift events; our potential explanations for rate scaling need not be mutually exclusive.

5.5. Evolution Is Hierarchical

All three models follow macroevolutionary tradition in taking the species as the unit and modeling changes that accrue among species over deep time (Stanley 1979). This tradition ignores the hierarchical nature of evolution and, in particular, can conflate change happening among populations

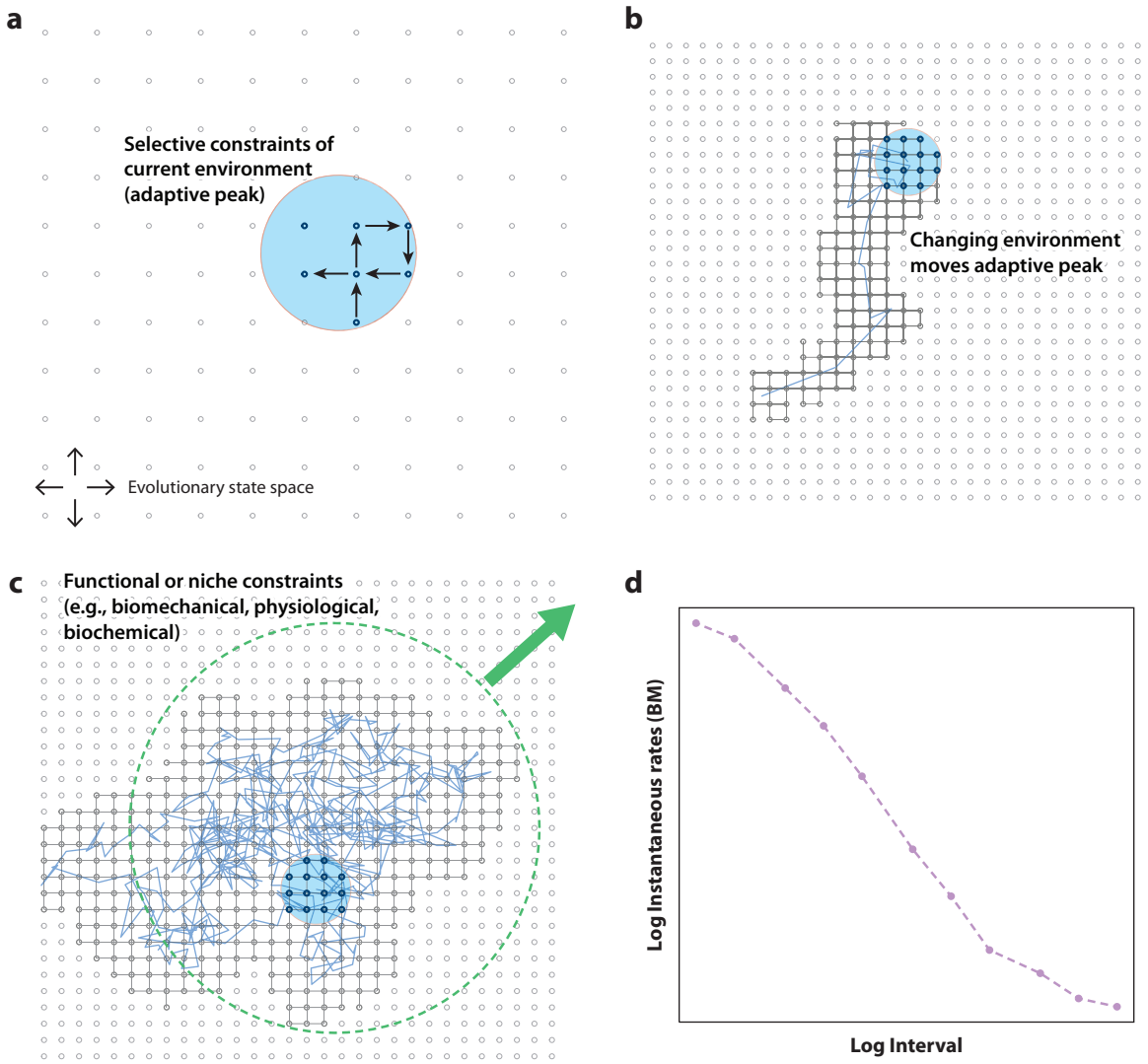

Figure 2

Hierarchical constraints on divergence at multiple scales. (*a*) The state space for evolution is illustrated as a grid, with points representing either phenotypic states or sequences. Neighboring points are one evolutionary step away from each other. The blue circle represents the set of phenotypic or sequence states that are adapted to a particular environment. Evolution occurs among these allowable states but is constrained by stabilizing selection. Evolution among these states becomes saturated, resulting in a loss of identifiability of evolutionary rate and rate scaling. (*b*) Over longer timescales, the set of evolutionary states that are adaptive may change (*gray lines*) in response to changing abiotic or biotic conditions. This change could represent, for example, change in the adaptive peak (*blue line*). (*c*) These adaptive peaks may themselves encounter further constraints (*dashed green circle*), for example, representing the limits of an ecological niche, an adaptive zone, or the biomechanical/physiological/biochemical limits of a particular organismal function. If the function or niche of an organism also changes, then the dashed green circle may also change on even longer timescales (*green arrow*), where they may eventually encounter still other constraints. These nested selective constraints that change on ever-increasing macroevolutionary timescales can ensure that evolutionary rates measured over shorter timescales are always faster than those measured over longer timescales and that model-based estimates of evolutionary rates that fail to account for these hierarchical constraints are incapable of accounting for saturation at all levels simultaneously. (*d*) Instantaneous rates of the process in panel *c* estimated with Brownian motion with increasingly large intervals result in ever-decreasing rate estimates.

within species with change occurring among species (Eldredge et al. 2005; Jablonski 2000, 2007). In each case of evolutionary rates, there is good reason to suspect that this failure of models to cross the species boundary is a problem for estimating evolutionary rates. In all three cases, there are also models and methods that can potentially directly address this limitation.

For molecular evolution, the core idea that we should expect substitutions to accumulate linearly with respect to time stems from the neutral theory of molecular evolution (Li 1997). However, this theory applies to the substitution rate and assumes that all measured differences are substitutions. If, instead, some of the differences we observe represent variation within species—that is, alleles that have not yet reached fixation—then we have misspecified our model (Penny 2005). Our model does not account for the dynamics among individuals within species and populations and considers only fixed differences between species. For recent divergence, differences that we see reflect both the substitution rate and the mutation rate, whereas over long timescales, most differences are substitutions. This situation leads directly to a predicted pattern of rates of molecular evolution scaling negatively with the time interval over which they are measured (Ho et al. 2011).

Under selection, scenarios are more complex. Selection by itself does not necessarily lead to rate scaling; in fact, under nearly neutral models of molecular evolution we can still expect substitutions to accumulate linearly through time (Ohta 1992). However, issues emerge again when we consider processes that occur among individuals within species. As most mutations are deleterious, selection filters ephemeral polymorphisms generated at present. This exacerbates the pattern described in Section 4.4 for neutral models. In simple terms, comparisons over short timescales include actual substitutions, neutral segregating variants, and deleterious mutations that have not been purged from the population. Over long timescales, differences eventually reflect mostly substitutions, both neutral and nonneutral (Penny 2005). Again, failure to account for the presence of both deleterious and neutral segregating variants within species populations—which can be considered as a model misspecification—leads to a pattern of negative rate scaling. This pattern leads to problems analogous to measurement error but cannot be corrected simply with estimates of instrument repeatability or sampling variance; rather, it is biological error that must be included in any adequate process model of evolution (Felsenstein 1985, Hansen & Bartoszek 2012).

For trait evolution, a similar type of hierarchical explanation lies in Futuyma's (1987, 2010) ephemeral divergence model. Under this model, most evolutionary trait divergence takes place in small populations, for example, at the periphery of a species' range where selective optima may be different. This divergence is mostly ephemeral, as the ultimate fate of many such populations is extinction or re-assimilation (Eldredge et al. 2005, Futuyma 1987). According to Futuyma (2010), then, novel traits that diverge persist only rarely, when these otherwise ephemeral changes are locked in by speciation and the evolution of reproductive isolation or are spread among the whole species. Under the ephemeral divergence model, we expect to see rapid, but ephemeral, trait evolution over short timescales and seemingly much slower rates over long timescales (Eldredge et al. 2005).

A similar explanation for diversification comes in the form of ephemeral or incipient speciation models (Etienne & Rosindell 2012, Rosenblum et al. 2012, Rosindell et al. 2010). The idea here is that the initiation of speciation, where populations begin to become independent lineages, occurs at a high rate but that these ephemeral forms only rarely persist as full species over longer timescales (Dynesius & Jansson 2014, Rosenblum et al. 2012). Such a process results in faster rates of speciation initiation over short timescales and slower rates over longer timescales (Harvey et al. 2019, Rabosky 2016).

All three of the above hierarchical explanations predict that the rate-scaling pattern will be nonlinear, with an inflection arising from the transition from within-species to between-species

changes. Within-species, population-level comparisons include within-species constraints, along with standing or ephemeral variation, and between-species comparisons consider only such changes that persist and become fixed at the species level. This is a prediction that can be directly tested. Indeed, most empirical patterns of rate scaling in evolution appear to have the predicted nonlinear shape with an inflection between 1 and 10 million years ago (Ho et al. 2011, Uyeda et al. 2011). Whether rate scaling disappears asymptotically at this micro-to-macro inflection point, or continues with a different scaling factor, is variable among the three rates examined and the model under which they are estimated. Ultimately, however, the hierarchical explanation relies on saturation of variation and constraint at successive levels of hierarchy, causing the data to lose information about time—either because change has not yet accumulated or because the state space has become saturated (**Figure 2**). Notably, this is also a definition of phylogenetic signal that generalizes across the biological hierarchy (Hansen & Orzack 2005, Jablonski 2007). Because successive levels of biological organization, from species to genera and above, may be subject to their own constraints that may depend on factors that themselves change at rates ranging from generations to eons, patterns of rate scaling may be expected to similarly occur at successive levels of the biological hierarchy (**Figure 2**). Adequately describing such dynamics requires models that likewise can accommodate the hierarchical nature of evolution (Kostikova et al. 2016, Reitan & Liow 2019, Reitan et al. 2012).

6. CONSEQUENCES OF RATE SCALING

The relative importance of each of these explanations for scaling of evolutionary rates among our three identified scaling patterns remains unclear. However, regardless of the cause, the prevalence of rate scaling means that this pattern has effects on many of our analyses and should be considered when interpreting evolutionary rates.

The most obvious effect that pervasive rate-scaling patterns have is that they tend to cause us to confuse the young with the fast. For example, when comparing evolutionary rates—whether speciation, extinction, trait evolution, or molecular evolution—across time periods of different length, we tend to estimate the highest rates for the shortest interval. This can be expressed in different ways. Rates of trait evolution tend to seem fastest in younger clades, speciation and extinction rates tend to be estimated as higher in lineages from younger habitats, and rates of molecular evolution seem faster when comparing shallow with deeply divergent pairs of taxa. This is not to say that the younger or shorter comparisons will always be uniformly faster; in fact, in all known scaling relationships there is enough scatter around the regression line that we should expect variation. Still, the scaling pattern should be considered when comparisons are made across different timescales.

A more subtle problem that also stems immediately from rate scaling is that all parameters that scale with the same covariate tend to correlate with one another. In this context, then, we should expect to see correlations among evolutionary rates, as all are related to time. Indeed, extensive papers have documented putative relationships between all three possible pairs of rates: molecular evolution and speciation rates (Lanfear et al. 2010, Pagel et al. 2006), trait evolution and speciation rates (Cooney & Thomas 2021; Rabosky & Adams 2012; Rabosky et al. 2012, 2013), and molecular evolution and trait evolution rates (Berv & Field 2018, Omland 1997). Although we are not suggesting that any of these studies are incorrect, per se, we argue that future papers in this area should strive to account for the potential effects of timescaling in such analyses.

7. CONCLUSION

There is no lack of explanations for rate scaling, which range from purely statistical to biologically meaningful. We hope we have provided a clear guide to the ubiquity of these patterns, how such

patterns can emerge from statistical and biological sources, and how they affect our analyses. Many of the mechanisms we describe likely play a major role in rate scaling, yet there remains considerable mystery in the relative importance of each one as ultimate explanations for these patterns. Although estimation error certainly explains some of these patterns, substantial evidence suggests that at least a portion of the apparent pattern likely results from the inadequacy of our models in representing evolutionary processes. Ultimately, we think the ubiquitous nature of rate scaling in evolution reveals much about our ignorance of how microevolutionary change accumulates from micro- to macroevolutionary scales, or at least how to adequately model it (Houle et al. 2011). Furthermore, we lack a mechanistic understanding of how to interpret apparent constraints on macroevolution (Boucher et al. 2018, Houle et al. 2011). Indeed, much of our discussion and the relevant literature on rates are intimately tied to Gould & Eldredge's (1977) motivation in proclaiming that stasis is data and to the enduring paradox of stasis (Hansen & Houle 2004). Spirited debate has ensued, and continues to this day, about the causes of this pattern, and it is unquestionable that this reframing revealed the great biases that evolutionary biologists often have in the questions they ask and the methods they use. Indeed, many of the great mysteries in evolution are not why change occurs but why it fails to occur. Traditionally, evolutionary biologists have framed the earth's incredible biodiversity as an outstanding mystery to be solved. Perhaps, though, we should be surprised not by how much but rather by how little biodiversity there is, considering the tremendous potential for evolution to generate new genes, traits, and species, even over short timescales. Understanding the underlying mechanisms of these limits and how they evolve over scales of time and space holds the most promise for adequately describing evolutionary processes with meaningful parameters that characterize the tempo and mode of evolutionary change.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

M.W.P. was funded by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC). J.R. received funding from the European Union's Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie (grant no. 785910). This work also benefited from Investissement d'Avenir grants managed by the Agence Nationale de la Recherche [Centre d'Étude de la Biodiversité Amazonienne (CEBA): ANR-10-LABX-25–01; Toward a comprehensive theory of interactions between organisms from molecular biology to eco-evolutionary scale (TULIP): ANR-10-LABX-0041]. B.N.S. was funded by the National Science Foundation (NSF) Graduate Research Fellowship Program (grant no. 1832399). J.C.U. was funded by the NSF Division of Environmental Biology (grant no. 1942717) and L.J.H. was funded by the NSF Office of Advanced Cyberinfrastructure (grant no. 1835904). Some ideas in this review were discussed in a working group funded by the University of British Columbia (UBC) through the Grants for Catalyzing Research Clusters program.

LITERATURE CITED

Albà MM, Castresana J. 2005. Inverse relationship between evolutionary rate and age of mammalian genes. *Mol. Biol. Evol.* 22(3):598–606

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *PNAS* 106(32):13410–14

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu. Rev. Ecol. Syst.* 33:707–40

Bandelt H-J. 2008. Clock debate: when times are a-changin': time dependency of molecular rate estimates: tempest in a teacup. *Heredity* 100(1):1–2

Beaulieu JM, Jhwueng D-C, Boettiger C, Brian C, O'Meara BC. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* 66(8):2369–83

Berv JS, Field DJ. 2018. Genomic signature of an avian Lilliput effect across the K-Pg extinction. *Syst. Biol.* 67(1):1–13

Brown JM, Thomson RC. 2018. Evaluating model performance in evolutionary biology. *Annu. Rev. Ecol. Evol. Syst.* 49:95–114

Boucher FC, Démery V, Conti E, Harmon LJ, Uyeda J. 2018. A general model for estimating macroevolutionary landscapes. *Syst. Biol.* 67(2):304–19

Butler MA, King AA. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164(6):683–95

Cantalapiedra JL, FitzJohn RG, Kuhn TS, Fernández MH, DeMiguel D, et al. 2014. Dietary innovations spurred the diversification of ruminants during the Caenozoic. *Proc. R Soc. B* 281(1776):20132746

Charlesworth B, Lande R, Slatkin M. 1982. A neo-Darwinian commentary on macroevolution. *Evolution* 36(3):474–98

Condamine FL, Rolland J, Morlon H. 2019. Assessing the causes of diversification slowdowns: temperature-dependent and diversity-dependent models receive equivalent support. *Ecol. Lett.* 22(11):1900–12

Cooney CR, Thomas GH. 2021. Heterogeneous relationships between rates of speciation and body size evolution across vertebrate clades. *Nat. Ecol. Evol.* 5(1):101–10

Cooper N, Thomas GH, Venditti C, Meade A, Freckleton RP. 2016. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc. Lond.* 118(1):64–77

DeSalle R, Freedman T, Prager EM, Wilson AC. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.* 26(1–2):157–64

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLOS Biol.* 4(5):e88

Dynesius M, Jansson R. 2014. Persistence of within-species lineages: a neglected control of speciation rates. *Evolution* 68(4):923–34

Eldredge N, Thompson JN, Brakefield PM, Gavrilets S, Jablonski D, et al. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31(2):133–45

Elhaik E, Sabath N, Graur D. 2005. The "inverse relationship between evolutionary rate and age of mammalian genes" is an artifact of increased genetic distance with rate of evolution and time of divergence. *Mol. Biol. Evol.* 23(1):1–3

Emerson BC. 2007. Alarm bells for the molecular clock? No support for Ho et al.'s model of time-dependent molecular rate estimates. *Syst. Biol.* 56(2):337–45

Emerson BC, Hickerson MJ. 2015. Lack of support for the time-dependent molecular evolution hypothesis. *Mol. Ecol.* 24(4):702–9

Emerson BC, Kolm N. 2005. Species diversity can drive speciation. *Nature* 434(7036):1015–17

Erwin DH. 2000. Macroevolution is more than repeated rounds of microevolution. *Evol. Dev.* 2(2):78–84

Estes S, Arnold SJ. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* 169(2):227–44

Etienne RS, Haegeman B, Stadler T, Aze T, Pearson PN, et al. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R Soc. B* 279(1732):1300–9

Etienne RS, Rosindell J. 2012. Prolonging the past counteracts the pull of the present: Protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* 61(2):204–13

Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* 25(5):471–92

Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125(1):1–15

Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–71

Felsenstein J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *Am. Nat.* 171(6):713–25

FitzJohn RG. 2010. Quantitative traits and diversification. *Syst. Biol.* 59(6):619–33

Foote M. 1994. Temporal variation in extinction risk and temporal scaling of extinction metrics. *Paleobiology* 20(4):424–44

Foote M. 2003. Origination and extinction through the Phanerozoic: a new approach. *J. Geol.* 111:125–48

Foote M. 2005. Pulsed origination and extinction in the marine realm. *Paleobiology* 31:6–20

Futuyma DJ. 1987. On the role of species in anagenesis. *Am. Nat.* 130(3):465–73

Futuyma DJ. 2010. Evolutionary constraint and ecological consequences. *Evolution* 64(7):1865–84

García-Moreno J. 2004. Is there a universal mtDNA clock for birds? *J. Avian Biol.* 35(6):465–68

Gillespie JH. 1986. Rates of molecular evolution. *Annu. Rev. Ecol. Syst.* 17:637–65

Gingerich PD. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222(4620):159–61

Gingerich PD. 2009. Rates of evolution. *Annu. Rev. Ecol. Evol. Syst.* 40:657–75

Gingerich PD. 2019. *Rates of Evolution: A Quantitative Synthesis*. Cambridge, UK: Cambridge Univ. Press

Goldberg EE, Foo J. 2020. Memory in trait macroevolution. *Am. Nat.* 195(2):300–14

Goldberg EE, Lancaster LT, Ree RH. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Syst. Biol.* 60(4):451–65

Gould NE-SJ, Eldredge N. 1972. Punctuated equilibria: an alternative to phyletic gradualism. In *Models in Paleobiology*, ed. TJM Schopf, pp. 82–115. Baltimore, MD: Johns Hopkins Univ. Press

Gould SJ. 1984. Smooth curve of evolutionary rate: a psychological and mathematical artifact. *Science* 226(4677):994–96

Gould SJ. 1985. The paradox of the first tier: an agenda for paleobiology. *Paleobiology* 11(1):2–12

Gould SJ, Eldredge N. 1977. Punctuated equilibria: the tempo and mode of evolution reconsidered. *Paleobiology* 3(2):115–51

Graur D, Martin W. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. *Trends Genet.* 20(2):80–86

Haldane JBS. 1949. Suggestions as to quantitative measurement of rates of evolution. *Evolution* 3(1):51–56

Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51(5):1341–51

Hansen TF, Bartoszek K. 2012. Interpreting the evolutionary regression: the interplay between observational and biological errors in phylogenetic comparative studies. *Syst. Biol.* 61(3):413–25

Hansen TF, Houle D. 2004. Evolvability, stabilizing selection, and the problem of stasis. In *Phenotypic Integration: Studying the Ecology and Evolution of Complex Phenotypes*, ed. M Pigliucci, K Preston, pp. 130–50. Oxford, UK: Oxford Univ. Press

Hansen TF, Martins EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50(4):1404–17

Hansen TF, Orzack SH. 2005. Assessing current adaptation and phylogenetic inertia as explanations of trait evolution: the need for controlled comparisons. *Evolution* 59(10):2063–72

Harmon L. 2019. *Phylogenetic Comparative Methods: Learning from Trees*. EcoEvoRxiv. **https://doi.org/10.32942/osf.io/e3xnr**

Harmon LJ. 2014. Macroevolutionary rates. In *The Princeton Guide to Evolution*, ed. A Baum, DJ Futuyma, HE Hoekstra, RE Lenski, AJ Moore, et al., pp. 567–72. Princeton, NJ: Princeton Univ. Press

Harmon LJ, Harrison S. 2015. Species diversity is dynamic and unbounded at local and continental scales. *Am. Nat.* 185(5):584–93

Harmon LJ, Losos JB, Davies TJ, Gillespie RG, Gittleman JL, et al. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64(8):2385–96

Harmon LJ, Schulte JA, Larson A, Losos JB. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301(5635):961–64

Harvey MG, Singhal S, Rabosky DL. 2019. Beyond reproductive isolation: demographic controls on the speciation process. *Annu. Rev. Ecol. Evol. Syst.* 50:75–95

Hawkes AG. 2018. Hawkes processes and their applications to finance: a review. *Quant. Finance.* 18(2):193–98

Henao Diaz LF, Harmon LJ, Sugawara MTC, Miller ET, Pennell MW. 2019. Macroevolutionary diversification rates show time dependency. *PNAS* 116(15):7403–8

Hendry AP, Kinnison MT. 1999. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* 53(6):1637–53

Hendry AP, Kinnison MT. 2001. An introduction to microevolution: rate, pattern, process. *Genetica* 112–113:1–8

Ho SYW, Duchêne S, Molak M, Shapiro B. 2015. Time-dependent estimates of molecular evolutionary rates: evidence and causes. *Mol. Ecol.* 24:6007–12

Ho SYW, Lanfear R, Bromham L, Phillips MJ, Soubrier J, et al. 2011. Time-dependent rates of molecular evolution. *Mol. Ecol.* 20(15):3087–101

Ho SYW, Phillips MJ, Cooper A, Drummond AJ. 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22(7):1561–68

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ. 2007. Evidence for time dependency of molecular rate estimates. *Syst. Biol.* 56(3):515–22

Höhna S. 2014. Likelihood inference of non-constant diversification rates with incomplete taxon sampling. *PLOS ONE* 9(1):e84184

Höhna S, Freyman WA, Nolen Z, Huelsenbeck JP. 2019. A Bayesian approach for estimating branch-specific speciation and extinction rates. bioRxiv 555805. **https://doi.org/10.1101/555805**

Höhna S, Stadler T, Ronquist F, Britton T. 2011. Inferring speciation and extinction rates under different sampling schemes. *Mol. Biol. Evol.* 28(9):2577–89

Holmes EC, Dudas G, Rambaut A, Andersen KG. 2016. The evolution of Ebola virus: insights from the 2013–2016 epidemic. *Nature* 538(7624):193–200

Hopkins MJ, Bapst DW, Simpson C, Warnock RCM. 2018. The inseparability of sampling and time and its influence on attempts to unify the molecular and fossil records. *Paleobiology* 44(4):561–74

Houle D, Pélabon C, Wagner GP, Hansen TF. 2011. Measurement and meaning in biology. *Q. Rev. Biol.* 86(1):3–34

Hunt G. 2007. The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *PNAS* 104(47):18404–8

Hunt G. 2012. Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology* 38(3):351–73

Ives AR, Midford PE, Garland T Jr. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Syst. Biol.* 56(2):252–70

Jablonski D. 2000. Micro- and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. *Paleobiology* 26(4):15–52

Jablonski D. 2007. Scale and hierarchy in macroevolution. *Palaeontology* 50(1):87–109

Jackson DA, Somers KM. 1991. The spectre of "spurious" correlations. *Oecologia* 86(1):147–51

Johnson PLF, Slatkin M. 2008. Accounting for bias from sequencing error in population genetic estimates. *Mol. Biol. Evol.* 25(1):199–206

Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In *Mammalian Protein Metabolism*, ed. HN Munro, 3:21–132. New York: Academic

Kendall MG. 1948. *Rank Correlation Methods*. New York: Griffin

Kenney BC. 1982. Beware of spurious self-correlations! *Water Resour. Res.* 18(4):1041–48

Kinnison MT, Hendry AP. 2001. The pace of modern life II: from rates of contemporary microevolution to pattern and process. *Genetica* 112–113:145–64

Kostikova A, Silvestro D, Pearman PB, Salamin N. 2016. Bridging inter- and intraspecific trait evolution with a hierarchical Bayesian approach. *Syst. Biol.* 65(3):417–31

Landis MJ, Schraiber JG. 2017. Pulsed evolution shaped modern vertebrate body sizes. *PNAS* 114(50):13224–29

Lanfear R, Ho SYW, Love D, Bromham L. 2010. Mutation rate is linked to diversification in birds. *PNAS* 107(47):20423–28

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* 50(6):913–25

Li W. 1997. *Molecular Evolution*. Sunderland, MA: Sinauer

Linder HP. 2008. Plant species radiations: where, when, why? *Philos. Trans. R. Soc. B* 363(1506):3097–105

Long SB. 1980. The continuing debate over the use of ratio variables: facts and fiction. In *Sociological Methodology*, ed. KF Schuessler, 11:37–67. San Francisco: Jossey-Bass

Louca S, Pennell MW. 2020. Extant timetrees are consistent with a myriad of diversification histories. *Nature* 580(7804):502–5

MacPherson A, Louca S, McLaughlin A, Joy JB, Pennell MW. 2020. *A general birth-death-sampling model for epidemiology and macroevolution*. bioRxiv 334383. **https://doi.org/10.1101/2020.10.10.334383**

Maddison WP, Midford PE, Otto SP, Oakley T. 2007. Estimating a binary character's effect on speciation and extinction. *Syst. Biol.* 56(5):701–10

Magallon S, Sanderson MJ. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55(9):1762–80

Magee AF, Höhna S, Vasylyeva TI, Leaché AD, Minin VN. 2020. Locally adaptive Bayesian birth-death model successfully detects slow and rapid rate shifts. *PLOS Comput. Biol.* 16(10):e1007999

Marin J, Hedges SB. 2018. Undersampling genomes has biased time and rate estimates throughout the tree of life. *Mol. Biol. Evol.* 35(8):2077–84. Erratum. *Mol. Biol. Evol.* 2018. 35(10):2595

McPeek MA. 2008. The ecological dynamics of clade diversification and community assembly. *Am. Nat.* 172(6):E270–84

McPeek MA, Brown JM. 2007. Clade age and not diversification rate explains species richness among animal taxa. *Am. Nat.* 169(4):E97–106

Mooers AO, Heard SB. 1997. Inferring evolutionary process from phylogenetic tree shape. *Q. Rev. Biol.* 72(1):31–54

Moreno LF, Stratton HH, Newell JC, Feustel PJ. 1986. Mathematical coupling of data: correction of a common error for linear calculations. *J. Appl. Physiol.* 60(1):335–43

Morlon H, Parsons TL, Plotkin JB. 2011. Reconciling molecular phylogenies with the fossil record. *PNAS* 108(39):16327–32

Mossel E, Vigoda E. 2005. Phylogenetic MCMC algorithms are misleading on mixtures of trees. *Science* 309(5744):2207–9

Nee S. 2006. Birth-death models in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* 37:1–17

Ng J, Smith SD. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *J. Evol. Biol.* 27(10):2035–45

Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23:263–86

Omland KE. 1997. Correlated rates of molecular and morphological evolution. *Evolution* 51(5):1381–93

Otto SP. 2018. Adaptation, speciation and extinction in the Anthropocene. *Proc. R Soc. B* 285(1891):20182047

Pagel M, Venditti C, Meade A. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314(5796):119–21

Pannetier T, Martinez C, Bunnefeld L, Etienne RS. 2021. Branching patterns in phylogenies cannot distinguish diversity-dependent diversification from time-dependent diversification. *Evolution* 75(1):25–38

Pennell MW, FitzJohn RG, Cornwell WK, Harmon LJ. 2015. Model adequacy and the macroevolution of angiosperm functional traits. *Am. Nat.* 186(2):E33–50

Penny D. 2005. Relativity for molecular clocks. *Nature* 436(7048):183–84

Prairie YT, Bird DF. 1989. Some misconceptions about the spurious correlation problem in the ecological literature. *Oecologia* 81(2):285–88

Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS ONE* 9(2):e89543

Rabosky DL. 2016. Reproductive isolation and the causes of speciation rate variation in nature. *Biol. J. Linn. Soc. Lond.* 118(1):13–25

Rabosky DL, Adams DC. 2012. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* 66(6):1807–18

Rabosky DL, Hurlbert AH. 2015. Species richness at continental scales is dominated by ecological limits. *Am. Nat.* 185(5):572–83

Rabosky DL, Lovette IJ. 2008. Density-dependent diversification in North American wood warblers. *Proc. R Soc. B* 275(1649):2363–71

Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, et al. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* 4:1958

Rabosky DL, Slater GJ, Alfaro ME. 2012. Clade age and species richness are decoupled across the eukaryotic tree of life. *PLOS Biol.* 10(8):e1001381

Raup DM. 1985. Mathematical models of cladogenesis. *Paleobiology* 11:42–52
Reitan T, Liow LH. 2019. layeranalyzer: Inferring correlative and causal connections from time series data in R. *Methods Ecol. Evol.* 10(12):2183–88
Reitan T, Schweder T, Henderiks J. 2012. Phenotypic evolution studied by layered stochastic differential equations. *Ann. Appl. Stat.* 6(4):1531–51
Revell LJ, Harmon LJ, Glor RE. 2005. Under-parameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.* 54(6):973–83
Reznick DN, Ricklefs RE. 2009. Darwin's bridge between microevolution and macroevolution. *Nature* 457(7231):837–42
Ricklefs RE. 2006. Global variation in the diversification rate of passerine birds. *Ecology* 87(10):2468–78
Rolland J, Condamine FL, Jiguet F, Morlon H. 2014. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLOS Biol.* 12(1):e1001775
Rolland J, Silvestro D, Litsios G, Faye L, Salamin N. 2018. Clownfishes evolution below and above the species level. *Proc. R Soc. B* 285(1873):20171796
Rosenblum EB, Sarver BAJ, Brown JW, Des Roches S, Hardwick KM, et al. 2012. Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evol. Biol.* 39(2):255–61
Rosindell J, Cornell SJ, Hubbell SP, Etienne RS. 2010. Protracted speciation revitalizes the neutral theory of biodiversity. *Ecol. Lett.* 13(6):716–27
Sadler PM. 1981. Sediment accumulation rates and the completeness of stratigraphic sections. *J. Geol.* 89(5):569–84
Sarver BA, Pennell MW, Brown JW, Keeble S, Hardwick KM, et al. 2019. The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. *PeerJ* 7:e6334
Schluter D. 2000. *The Ecology of Adaptive Radiation*. Oxford, UK: Oxford Univ. Press
Scholl JP, Wiens JJ. 2016. Diversification rates and species richness across the tree of life. *Proc. R Soc. B* 283(1838):20161334
Schwenk K. 1994. A utilitarian approach to evolutionary constraint. *Zoology* 98:251–62
Sheets HD, Mitchell CE. 2001. Uncorrelated change produces the apparent dependence of evolutionary rate on interval. *Paleobiology* 27(3):429–45
Silvestro D, Kostikova A, Litsios G, Pearman PB, Salamin N. 2015. Measurement errors should always be incorporated in phylogenetic comparative analysis. *Methods Ecol. Evol.* 6:340–46
Slater GJ, Harmon LJ, Alfaro ME. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. *Evolution* 66(12):3931–44
Stadler T. 2013. How can we improve accuracy of macroevolutionary rate estimates? *Syst. Biol.* 62(2):321–29
Stadler T, Rabosky DL, Ricklefs RE, Bokma F. 2014. On age and species richness of higher taxa. *Am. Nat.*184(4):447–55
Stanley SM. 1979. *Macroevolution: Pattern and Process*. San Francisco: WH Freeman & Co.
Uyeda JC, Caetano DS, Pennell MW. 2015. Comparative analysis of principal components can be misleading. *Syst. Biol.* 64(4):677–89
Uyeda JC, Hansen TF, Arnold SJ, Pienaar J. 2011. The million-year wait for macroevolutionary bursts. *PNAS* 108(38):15908–13
Wayne RK, Van Valkenburgh B, O'Brien SJ. 1991. Molecular distance and divergence time in carnivores and primates. *Mol. Biol. Evol.* 8(3):297–319
Weir JT, Schluter D. 2008. Calibrating the avian molecular clock. *Mol. Ecol.* 17(10):2321–28
Wertheim JO, Sanderson MJ. 2011. Estimating diversification rates: How useful are divergence times? *Evolution* 65(2):309–20
Wiens JJ. 2011. The causes of species richness patterns across space, time, and clades and the role of "ecological limits." *Q. Rev. Biol.* 86(2):75–96
Yang Z. 2006. *Computational Molecular Evolution*. Oxford, UK: Oxford Univ. Press