Norms of evidence in the classification of living fossils

Beckett Sterner*

School of Life Sciences, Arizona State University, Tempe, AZ, United States

Some species have held fast for millions of years as constants in a changing world. Often called "living fossils," these species capture scientific and public interest by showing us the vestiges of an earlier world. If living fossils are defined by a holistic pattern of low evolutionary rates or stasis, however, then classifying a species as a living fossil involves the application of sophisticated norms of scientific evidence. Using examples from *Crocodilia* and the tuatara (*Sphenodon punctatus*), I show how scientists' evidential criteria for classifying living fossils are contentious and underspecified in many cases, threatening the concept's explanatory interest and its adequacy for sustaining a collective problem agenda as proposed by Scott Lidgard and Alan Love. While debates over the definition of the living fossil concept may appear fruitless, I suggest they can be productive insofar as the debate leads to clarified and improved evidential standards for classification. To this end, I formulate a view of the living fossil concept as an investigative kind, and compare two theoretical frameworks as a basis for shared evidential norms: the Zero Force Evolutionary Law framework, introduced by Daniel McShea and Robert Brandon, and the statistical model selection framework first developed by Gene Hunt in the 2000s.

KEYWORDS

stasis, natural kinds, tempo and mode of evolution, random walk, trait evolution, evolutionary rate

1. Introduction

Some taxonomic groups seem to have persisted, unchanged, for millions of years against a world in flux. Often called "living fossils," these taxa capture the interest of scientists and the public alike by showing us the vestiges of an earlier world. Horseshoe crabs, for instance, have changed so little in their visual appearance that one can immediately recognize fossils from millions of years ago (Avise et al., 1994; Lamsdell, 2019). Superficial appearances, though, may obscure nuanced patterns of evolution. The outwardly ancient appearance of crocodiles, for example, hides rapid evolution within a limited set of options (Felice et al., 2021). Exactly what makes something a living fossil has been an ongoing subject of debate among biologists and philosophers (Werth and Shear, 2014; Carnall, 2016; Bennett et al., 2018; Lidgard and Love, 2018; Turner, 2019; Lidgard and Love, 2021; Watkins, 2021). One can view living fossils, for example, as an extreme class of exceptions to the more commonly observed pattern of mosaic evolution, in which the traits of lineages evolve at different rates over time (DeSilva, 2018; Parravicini and Pievani, 2019). This locates living fossils in a wide-ranging research program that aims to analyze the relationship of population-level processes to taxonomic diversification and divergence (Lidgard and Love, 2018; Lidgard and Kitchen, 2023).

To sustain an interdisciplinary research agenda against this backdrop of conceptual disagreement, though, researchers need a shared set of evaluative criteria to at least agree on what counts as a productive contribution. I focus in particular on the idea that living fossils in

some sense represent extreme cases of slow evolutionary change, which connects them to broader debates in evolutionary biology about the nature and frequency of different evolutionary rates. Any rate or pattern of change that biologists attribute to a lineage or taxonomic group as a whole, though, is a theoretical construct that is not observable directly through measuring the properties of individual organisms. As a result, difficult conceptual and practical questions arise about whether and how one can summarize the history of an evolutionary lineage in a single rate. In fact, this problem applies more generally to any attempt to relate the evolutionary rates of parts and wholes, e.g., to comparing the evolutionary rates of genetic loci, whole genomes, features of a morphological module, and body size. As Lidgard and Love point out, "when focused on either molecular or morphological characters that serve as proxies for species or lineages, there are rampant part–whole ambiguities in evaluating evolutionary stasis and change, many of which bear directly on controversies about categorizing living fossils" (Lidgard and Love, 2018, p. 762).

I take the purpose of having a shared evidential framework, then, to be guiding productive, intersubjectively reliable research on living fossils when so little about their defining features and proper methods of study are settled as common knowledge. While debates over the definition of the living fossil concept may appear fruitless, I suggest they can be productive insofar as the debate leads to clarified and improved evidential standards for classification. To this end, I formulate a view of the living fossil concept as an investigative kind, and I analyze the suitability of two theoretical frameworks for setting shared evidential norms: the first is based on the Zero Force Evolution Law (ZFEL; McShea and Brandon, 2010; McShea et al., 2019; Brandon and McShea, 2020); the second is based on Statistical Model Selection (SMS) methods for linear Gaussian time series (Hunt, 2006, 2008b; Hopkins and Lidgard, 2012; Reitan et al., 2012; Hunt et al., 2015; Voje, 2020). I argue that the ZFEL framework's reliance on null hypotheses is ill-suited to contemporary research practices and that the ZFEL is not metaphysically more fundamental when we consider statistical expectations for whole lineages. I also argue that the SMS framework provides a better approach to data analysis, but that major gaps remain in our ability to formulate multi-variate, multi-level models relevant to defining living fossils and distinguishing them from other types of evolutionary patterns.

To begin, I survey different stances about the living fossil concept in section 2 in order to raise several conceptual and methodological concerns about the concept's adequacy as the basis for a shared problem agenda. Section 3 illustrates these points using recent controversies over classifying lineages as living fossils, and section 4 identifies some key methodological questions that follow as a result. Sections 5 and 6 then analyze how the frameworks based on the ZFEL and SMS address these questions.

2. Shared evidential criteria for living fossils are missing

Academic debates about the definitions of concepts range from being a waste of time to catalysts for major breakthroughs. Telling the difference is a fruitful area of overlap between philosophy and science, and both fields have contributed to analyzing the conditions under which conceptual debates lead to meaningful collective progress (e.g., Chang, 2004; Brigandt and Love, 2012; Brigandt, 2020; Pradeu et al.,

2021; Sterner, 2022) and references therein. Understanding how conceptual change happens in science, more generally, and assessing its rationality are also central goals for philosophy of science (Nersessian, 2017; Nickles, 2017, 2021). In early work, philosophers often characterized progress as a process of replacement, such as when scientists discard an established definition for a term because its meaning is different in a better-supported theory. Extensive case studies in many scientific fields show that simple abandonment of old meanings is historically unusual or rare, however (Kellert et al., 2006; Brigandt, 2020; Ludwig and Ruphy, 2021), including for living fossils (Lidgard and Kitchen, 2023). Instead, research communities often sustain varying forms and degrees of pluralism by ensuring participants have adequate context and training to disambiguate between meanings as needed. This may apply at varying social scales, such as for the dominant meanings of terms such as function or gene across disciplines or species names among taxonomists. Nonetheless, ambiguous concepts are often controversial among scientists because, for example, they see ambiguity as likely to cause confusion or misinterpretation, have a preference for simplicity or context-independence, or view ambiguity as an opportunity to champion a particular viewpoint (Sterner, 2022).

There are a variety of philosophical stances one can take toward conceptual debates, many of which researchers have already adopted for living fossils. An eliminativist stance would have the research community abandon the use of the living fossil concept as too ambiguous or arbitrary to be productive. Some argue the name "living fossil" is just a catchy metaphor that scientists have used to bundle together a heterogeneous bunch of stuff that lacks any generalizable properties or explanations (Carnall, 2016). A different deflationary option proposes that there is no special explanation to be found for living fossils either individually or as a category; the striking similarities we observe between past and present-day organisms, for example, may simply be unusual, chance outcomes of otherwise typical evolutionary processes (Werth and Shear, 2014). Alternatively, a unificationist stance would seek to find one definition that could command consensus in the field and either subsume or separate off other views and purposes. Some authors have argued for keeping the living fossil category by refining its key defining properties (Herrera-Flores et al., 2017; Bennett et al., 2018), for example, or restricting its scope to a particular application, e.g., conservation (Turner, 2019). All of these stances seek to eliminate the "mess" of existing confusions in one way or another.

Alternatively, there are options that embrace the mess as a positive state of affairs. Lidgard and Love (2018, 2021) argue we should avoid definitional debates and focus instead on the interdisciplinary agenda of research questions raised by living fossils. "The role of the living fossil concept can be understood as setting an integrated agenda for research—interrelated suites of questions about patterns in need of explanation and processes relevant to specific character constellations and wholes—that advances our understanding of evolutionary stasis across hierarchical levels of organization" (Lidgard and Love, 2018, p. 766). On their view, the living fossils literature has the coherence of a cluster of interlinked problems: a shared topic connected by multiple concepts and practices that coordinate research interests and activities without presupposing an underlying unity.

An important concern for this approach is that controversies over the classification of taxa as living fossils may threaten to overwhelm the perception of a shared problem cluster. Historically, people have

often categorized taxa as living fossils on the basis of qualitative impressions, sometimes using on only one or a few actual specimens. This can make for lively and popular debate, but without a shared framework of investigation, such debates often generate more heat than light. Attempts to formulate more exact criteria have identified a number of types of evolutionary patterns and relations—such as stasis, low rates of change, and deep phylogenetic divergence close to the common ancestor of a group—that are promising candidates for quantitative analysis (Lidgard and Love, 2018).

Operationalizing these ideas has proven to be an enduring and productive challenge for evolutionary biology, however. Pragmatically, a scientist interested in roughly the sort of phenomenon people label as living fossils has to be able to trust there is something worthwhile to learn from a new or updated example. If researchers apply the term too imprecisely, the purpose of the concept deflates to being a public-facing rhetorical device for getting attention. Moreover, even if scientists have a precise meaning in mind, when they apply the concept with insufficient evidence or rigor, it undercuts the value of a shared problem agenda. Living fossils seem to be at risk of this happening: scientists disagree not only about which criteria to use but also about how these criteria apply to particular cases, as we'll see in several examples below.

Responding to this concern, Lidgard and Love point to advances in quantitative models of trait evolution. "Empirical advances in quantitatively evaluating evolutionary modes among different morphological characters in fossil lineages open a path for investigating those membership criteria by generating rigorous measurements of stability, which then can be used to explore stasis in bundles of molecular and morphological characters" (Lidgard and Love, 2018, p. 768). However, they also note the related challenge of clarifying the relationships scientists assume, often tacitly, between the specific histories of individual traits and the holistic pattern of evolution shown by a lineage. For example, one can use existing single trait models to operationalize the "stability" of a bundle of characters in several ways that may lead to different empirical outcomes: (1) we might require each trait to individually show stability as a necessary and sufficient condition for the bundle to be stable; (2) we might define the stability of a bundle as a statistical function of

the traits composing it, e.g., by taking an average of the rates shown by each trait; (3) we might define a holistic model of stability for the bundle, e.g., invariance of morphological shape or ecological function, that makes statistical predictions about how individual traits should evolve. The modeling framework they cite (Hopkins and Lidgard, 2012; Hunt et al., 2015) applies primarily to univariate time series of fossil lineages, and on its own does not provide a sufficient basis for addressing multi-trait patterns of evolution, especially across part–whole levels.

3. Some examples of part–whole ambiguities in classifying living fossils

Identifying the correct whole–part relationships to use for studying living fossils is neither a purely empirical nor theoretical problem, as its answer depends on how one defines evolutionary rates and operationalizes their measurement, among other issues. Table 1 lists a series of examples at different compositional levels and shows how they are connected to commonly used diagnostic criteria for living fossils. For example, which set of lower-level parts are sufficient to establish a low rate for the lineage overall, and what should scientists conclude when a lineage exhibits traits changing at different rates? In this section, I highlight some examples drawn from quantitative research on fossil lineages in general as well as on two disputed living fossil groups in particular: *Crocodilia* and the reptile species tuatara (*Sphenodon punctatus*).

In a general study of 635 traits measured across 153 fossil species lineages, Hopkins and Lidgard (2012) found the majority of species showed conflicting patterns of evolutionary change in their traits. In particular, they examined the frequency and distribution of three types of patterns individuals traits may show: stasis, modeled as Gaussian fluctuations around a fixed mean; Brownian motion (random walk), represented as a stochastic diffusion process; and a directional trend, modeled as Brownian motion with a linear tendency to increase or decrease with time. Looking at the set of traits individually, they found Brownian motion was most common (53.5% of traits), then stasis (41.9%), and directional trend (5.7%). They also found that the proportion of traits in each lineage was roughly similar,

TABLE 1 Example properties of cellular organisms that are relevant to classifying lineages as living fossils according to the compositional level on which they occur.

Compositional level	Example properties	Relevance to living fossil criteria
Clade	Taxonomic richness, ancestral vs. derived traits, ancient divergent lineages	Comparing past and present taxonomic richness
		Identifying plesiomorphic characters
		Inferring a lineage's (relative) time of divergence
Species or population lineage	Ecological niche, allometry, developmental modularity, geographic range, means and variances of (sub)organismal properties	Comparing similarity of living individual to ancestral fossil
		Comparing past and present geographic ranges
		Prolonged geological duration relative to similar entities
		Slow evolutionary change relative to similar entities
Individual unit	Body size, shape, mass, age, sex, genotype, behaviors, metabolic rate, skeletal structure	Slow evolutionary change relative to similar entities
Morphological/ macromolecular	Size, shape, color, arrangement, and number of structural parts, biomechanical properties	Slow evolutionary change relative to similar entities
Molecular	Genetic sequence, chemical composition, gene regulation and expression	Slow evolutionary change relative to similar entities

so that the median proportions of traits in lineages showing each pattern were about 50%, 40%, and 0%, respectively. While some lineages did show stasis or Brownian motion across all their traits, at those that scientists measured, most species present a more complicated holistic picture.

To press this point, consider whether one could just appeal to collecting a "representative sample" of traits from a lineage. In practical terms, this is generally infeasible as an ideal for data collection. In paleontology, for example, complete specimens are rarely available, so that studies of morphological variation typically focus on whatever parts are most frequently preserved, e.g., jaw bones or leaf imprints. On a conceptual level, furthermore, it is not clear what a "representative sample" of traits would even be, given the variety of compositional levels and types of properties one could study (see Table 1). Different traits are also known to reflect evolutionary dynamics at different scales. Neutral DNA sites evolve faster than sites under negative selection, for example, and so erase their phylogenetic legacy more quickly. Similarly, some morphological traits evolve as transient local adaptations to environmental circumstances while others cause irreversible genetic isolation between populations. Similar questions have proven to be major obstacles for species delimitation, classification, and phylogenetics (Sterner and Lidgard, 2014; Haber and Velasco, 2021).

The first example uses recent studies of crocodilian snouts to highlight the ambiguity of whether more traits are better or simply different. Or to put it another way, why assume a lineage exhibits a single, comprehensive rate? *Crocodilia* (also called *Crocodylia*) is a taxonomic order of predatory, semi-aquatic reptiles that contains 26 living species, including crocodile and alligator species, and a rich fossil history (Stubbs et al., 2021). Crocodilians in combination with a large number of extinct fossil taxa comprise the clade *Pseudosuchia* that forms the closest sibling branch to all birds.

Biologists have historically used 2-dimensional lateral (side) or dorsal (top-down) profiles of crocodilian skulls to measure morphological evolution, partly to increase taxonomic sampling because more complete fossils are rare. Based on these profile views, biologists have found that crocodilian morphology is highly conserved, reflecting their shared semi-aquatic, predatory lifestyle, with most variation occurring in the length of the snout. A recent paper, however, applies new 3D imaging techniques to incorporate other potential sources of variation, such as in the shapes of the front and back of the skulls and internal chambers and bone structures (Felice et al., 2021). In addition to the known elongation of the snout, they find that "the cross-sectional shape of the snout is also a key feature separating aquatic and piscivorous taxa from terrestrial and omnivorous/herbivorous taxa" (Felice et al., 2021, p. 6). Crucially, incorporating a new dimension of morphological variation brought into focus a new ecological source for divergence among lineages that contrasts with the convergent pattern of elongation that responds to the clade-wide shift to a semi-aquatic lifestyle. This illustrates how adding more traits may change estimates of evolutionary rates simply because the additional traits may reflect new historical events or relationships. The meaning of an "overall average" rate across all traits remains ambiguous.

Crocodilians also provide a second example of part–whole ambiguity: one can quantify rates in terms of the dynamics of both population means and variances (Hunt, 2012). In the same study by Felice et al., they conclude that crocodilians are not aptly called living

fossils, because "despite having low overall disparity, modern crocodyloids are not experiencing evolutionary stasis. Instead, they are rapidly and repeatedly exploring a limited range of phenotypes" (Felice et al., 2021, p. 6). This distinction is fuzzier than it appears, though, when we shift from qualitative theorizing to statistical modeling. The standard model for a trait exhibiting stasis in fossil lineages describes the population average as fluctuating according to a Gaussian distribution around a fixed mean. While the model is stationary with respect to time, evolution does occur as the population average moves up and down. If the variance of the Gaussian distribution is large (relative to other traits, let us assume), then the lineage actually evolves quite rapidly as it travels above and below the mean, but the trait fails to accumulate any net divergence. Alternatively, the qualitative idea of rapidly exploring a limited range of phenotypes is also consistent with a constrained random walk model, where the population average can accumulate net displacement from its original starting position but cannot go above or below certain boundaries. Both models can show rapid change in population averages within a fixed phenotypic space, but only the latter model can be interpreted as having a rate of cumulative divergence over time.

An ongoing debate over the relation between genomic and morphological rates in the tuatara provides a third example. While scientists have historically diagnosed lineages as living fossils using morphological evidence, genome sequencing provides an interesting way to compare genetic to phenotypic rates (Janecka et al., 2012). Hay et al. (2008) presented the first evidence addressing mutation rates in the tuatara, analyzing a sequence of ancient DNA samples from the species covering about 600 to 8,800 years ago. They found the surprising result that tuatara showed a 50% higher mutation rate than other vertebrates known at the time, even though the tuatara had the slowest rates of morphological evolution as well as slow organismal growth and metabolism, which together would predict a slow mutation rate. They concluded that "rates of neutral molecular and phenotypic evolution are decoupled" in the species (Hay et al., 2008, p. 106).

Since then, however, research has illustrated the multiplicity of ways to understand and measure molecular and phenotypic rates, and conclusions about the tuatara have seesawed back and forth based in part on the traits and taxa that scientists have sampled. On the molecular side, Miller et al. (2009) [see also response by Subramanian et al. (2009)] disputed Hay et al.'s conclusion based on several issues including effects from sampling mitochondrial vs. microsatellite DNA and different geographic populations. On the morphological side, Meloro and Jones (2012) tried to undercut the tuatara's living fossil status by using increased sampling of extinct lineages to argue *Sphenodon* has undergone substantial morphological evolution in the last 220 million years and is not the product of stasis. Herrera-Flores et al. (2017) for the first time posed two quantitative criteria for the living fossil status of tuatara. They found that the tuatara lineage has significantly lower phenotypic rates than other lineages and is positioned in a morphologically conservative position relative to its clade near the average trait values. However, they had to analyze data from a smaller subset of skull morphology, the lower jaw, to acquire a larger taxonomic sample of fossil and extant relatives. Most recently, Gemmell et al. (2020) report a full draft genome sequence and compare it to other reptiles, mammals, and birds. Parallel to Herrera-Flores et al.'s finding about faster rates in basal taxa, Gemmell et al. find evidence for punctuated rates in amniotes more broadly and a

slow rate for the tuatara: 7% lower than any rate observed in the sample of other reptile species. Nonetheless, they document a wide range of genomic innovations and compositional differences since the tuatara's common ancestor with other vertebrate lineages. As a result, substantial ambiguity remains about how summary genetic and phenotypic rates for the tuatara should be operationalized. It remains unclear, for instance, whether higher mutation rates are supposed to be correlated with phenotypic rates in the sense of increased stationary fluctuations or accumulated change.

While Lidgard and Love have suggested that adequate evidential practices and criteria are already available for classifying lineages as living fossils, the examples I presented in this section indicate the situation is not so straightforward. There is clear potential for the living fossil problem agenda to get stuck in the swamp of endless classificatory disputes as different groups operationalize criteria for living fossils in incompatible ways for the same lineages. These difficulties are not necessarily fatal, though.

4. Evidential norms for investigating living fossils

From a positive viewpoint, definitional debates can be productive insofar as they clarify and raise evidential standards for attributing living fossil status to a lineage. To develop this point further, I will add one more stance about the living fossils concept to our list: that living fossils are investigative kinds. Philosopher of biology Ingo Brigandt originally introduced the idea of investigative kinds to capture how the concept of biological species has developed over time through a combination of empirical research and conceptual debate and refinement (Brigandt, 2003). As he writes,

> "An investigative kind is a group of things that are presumed to belong together due to some underlying mechanism or a structural property… An investigative-kind concept thus originates when a certain pattern among a class of objects is observed and it is assumed to be founded on some theoretically important, but yet unknown relevant mechanism that generates this pattern. An investigative-kind concept is associated with a search for the basis of this kind… An investigative-kind concept may change its reference throughout scientific investigation… Objects originally assumed to belong to the extension may prove not to be members of the kind. If it becomes clear that there are several relevant mechanisms that account to some extent for the observed pattern that was important for the introduction of the term, the concept may split" (Brigandt, 2003, p. 1,309).

As I understand it, treating living fossils as an investigative kind involves taking a dialectical stance about our understanding of the concept's meaning. As it stands now, the living fossil concept is inadequate for the purposes of classifying and explaining the evolutionary behaviors of lineages, but the best path forward is an iterative process of theoretical and empirical investigation to align the contents and boundaries of the concept with the epistemic roles it plays in associated research problems. Similarly, even though biologists continue to differ widely in their preferred meaning of the species concept across research fields (Stankowski and Ravinet, 2021), they have developed shared practices for generating and evaluating

empirical evidence that can decisively show whether a particular group of organisms should be classified as a species in a given sense.

In the rest of this section, I characterize two types of norms relevant to scientists' practices of data collection and analysis and illustrate how these norms may influence the classification of living fossils. The first type of norm concerns what trait and taxon sampling scientists should use to provide a necessary or sufficient evidential basis for classifying living fossils (or evolutionary patterns more broadly). A study's trait sampling is the set of phenotypic or genetic characters the authors measured and included in their dataset. Taxon sampling is the set of taxonomic groups, e.g., species or lineages, for which the authors have trait data. The second type of norm concerns which analytical methods scientists take to be most appropriate for determining the evidence supporting a lineage's classification. I take analytical methods here to include methods for statistical estimation or prediction, e.g., of rates of change, and for classifying lineages as fitting a particular type of evolutionary pattern, e.g., using null hypothesis testing or other statistical model selection techniques.

The multi-variate and multi-level nature of living fossils has important methodological implications for data collection and analysis. Statistically speaking, it is possible that a lineage as a whole may exhibit a dominant mode of evolution, such as stability or a linear trend in size, even as a majority of trait measurements individually best fit to other patterns. A similar phenomenon is familiar to phylogeneticists who must take into account incomplete lineage sorting when using gene trees to infer species trees (Haber and Velasco, 2021).

In a study, Hopkins and Lidgard (2012) used simulations to explore the prevalence of mosaic evolution in the morphology of a trilobite lineage for which a previous empirical study had found an increasing trend in body size. Figure 1 shows their main simulation results. Even though almost half of the observed morphological variation over time is explained by a linear directional trend (Principal Component 1 in Figure 1B), they find that more traits fit strongly to stasis or random walk models (Figure 1C, bottom right subpanel). Moreover, the chance of finding a pair of traits that show the same pattern of evolution is less than half across the simulation runs they conducted, and it decreases rapidly to zero as one considers more traits (Figure 1D). Although these simulated data represent a bundle of characters evolving according to a linear trend, rather than the stability characteristic of living fossils, the basic point holds that a clear evolutionary pattern at one level may not be reflected in the majority or average pattern shown by traits measured at a lower level. Since most studies of living fossils either rely on a qualitative analysis of morphology or quantitative analyses of a single trait or averaged rate across traits, these results could be seriously misleading if such mismatches between patterns at different levels are common.

One's conception of living fossils is relevant to data collection in other ways by influencing which traits or other lineages are relevant. If one conceives of living fossils as showing morphological and genetic stability because they occupy stable ecological niches, then there is no reason to expect slow rates in ecologically irrelevant traits (except if they are linked or dependent on traits that are ecologically important, of course). As we saw with the crocodilian example above, sampling skull morphology from dorsal vs. lateral perspectives can change which ecologically or phylogenetically important traits appear in the dataset. Such an ecological view of living fossils would therefore treat some traits as having high evidential relevance for classifying lineages
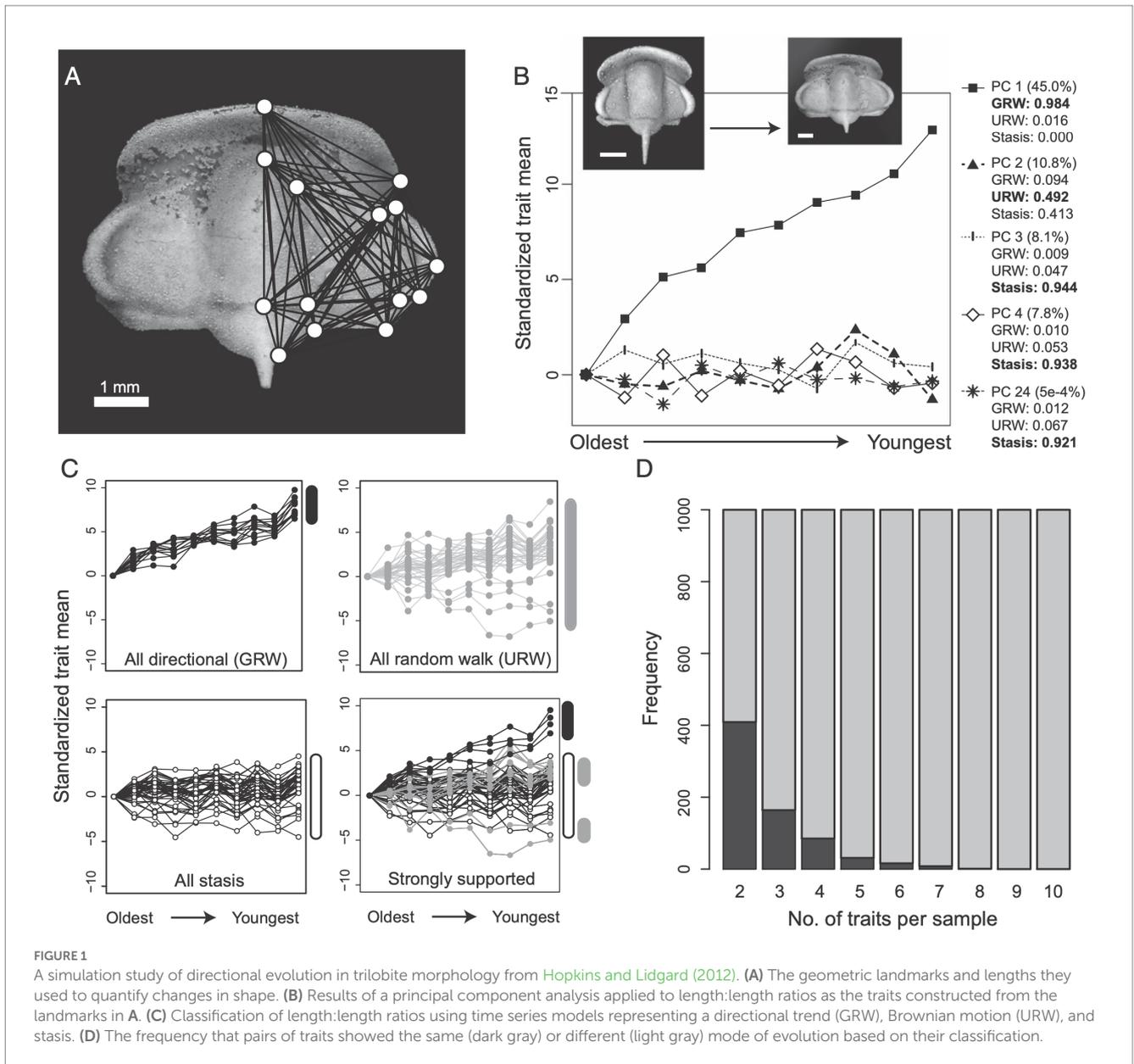
FIGURE 1
A simulation study of directional evolution in trilobite morphology from Hopkins and Lidgard (2012). **(A)** The geometric landmarks and lengths they used to quantify changes in shape. **(B)** Results of a principal component analysis applied to length:length ratios as the traits constructed from the landmarks in **A**. **(C)** Classification of length:length ratios using time series models representing a directional trend (GRW), Brownian motion (URW), and stasis. **(D)** The frequency that pairs of traits showed the same (dark gray) or different (light gray) mode of evolution based on their classification.

and others as low relevance. On the taxon sampling side, one could classify lineages as living fossils by comparing their evolutionary rates to other lineages (e.g., in the tuatara genome example) or by evolutionary mode, regardless of rate. Early studies of the tuatara exaggerated its morphological stability by not incorporating extinct fossil lineages that would have provided phylogenetic context for the evolution of ancestral states (Meloro and Jones, 2012). Alternatively, one could treat a lineage as a living fossil if the relevant set of selected traits fit best to the stasis model, regardless of whether the traits fluctuate around their stable means more or less than other lineages.

An additional concern for norms of data analysis comes from the possibility that stochastic processes may give rise to distinctive patterns purely by chance. An atom, for instance, may decay very slowly compared to its expected half live while still obeying the same underlying physical process as those that decayed earlier. Similarly, a random walk in one or two dimensions will always return to its initial state by chance given enough time, even though an ensemble of random walks will on average grow further from their initial state. We can therefore make a distinction between lineages that accidentally vs. non-accidentally meet the identifying characteristics of living fossils. A non-accidental living fossil lineage evolves according to a process whose expected statistical behaviors match the classificatory criteria, while an accidental living fossil lineage meets the criteria because chance events caused it deviate from the expected behavior. If traits evolve according to processes, such as random walks, that purely by chance show little net change later in time, then lineages fitting commonly used criteria for living fossils can occur by accident.

This distinction matters for the explanatory interest of living fossils because no special explanation is required for lineages that meet the criteria for being living fossils purely by chance. By analogy, getting 20 heads in a row when flipping a truly fair coin demands no special explanation beyond chance. While extreme or atypical outcomes are in theory possible no matter how much data we collect, their practical relevance generally becomes increasingly rare as sample

size increases. This indicates the need for an operational, statistical way to determine when there is strong evidence. Evolutionary biologists disagree, however, on the relative suitability of different statistical techniques such as null hypothesis tests vs. AIC-type model selection approaches. The two theoretical frameworks I consider in the following sections will exemplify this debate and the other evidential norms.

5. Comparing the ZFEL and statistical model selection frameworks

A shared evidential framework can help researchers meet several prerequisites for dialectical investigation of the living fossil concept. One prerequisite is enabling comparative analysis: there has to be a basis for making like-to-like comparisons of data, analyses, and conclusions across cases and researchers. Even if it proves impossible for researchers to eliminate sources of empirical disagreement due to divergent background assumptions (Sterner and Lidgard, 2021), progress is still possible if researchers can readily translate results between alternative viewpoints (Rescher, 2000; Sterner et al., 2022). Closely related is the need to clarify theoretical assumptions or principles. Methodology rests on theory in order to justify the correctness of a particular way of collecting and analyzing information to answer a question. Comparative analysis can be positively misleading in this respect if the cases we are comparing are based on faulty analyses. A third requirement is to clarify the shared or divergent aims that researchers bring to evaluating definitions of the living fossil category. This is critical for seeing which categories need to be kept because they are epistemically (or otherwise) valuable, even if terminology needs to expand or be revised.

The basic point of a shared evidential framework, then, is to guide methodology across different labs and research projects. While it would also be illuminating to survey further examples of statistical methods used in living fossils research studies, these are likely to prove fragmented and partial. Instead, I now turn to compare two potential evidential frameworks that are especially relevant for living fossils research.

5.1. ZFEL framework

Several features make the ZFEL framework potentially highly suitable for conceptualizing evidential criteria for living fossils as well as their explanatory importance. In its general formulation, the ZFEL states: "In any evolutionary system in which there is variation and heredity, there is a tendency for diversity and complexity to increase, one that is always present but may be opposed or augmented by natural selection, other forces, or constraints acting on diversity or complexity" (McShea and Brandon, 2010, p. 4). Because all units capable of evolving must possess heritable variation, McShea and Brandon argue that the ZFEL is best understood as a fundamental law of biology that describes an always acting, if not always dominant, tendency toward increased variation in properties of the parts within systems (complexity) and properties of systems in a population (diversity). The ZFEL is therefore not straightforwardly an empirical generalization like Cope's law but rather operates similarly to Newton's first law in classical mechanics: it provides the baseline or null pattern

that should result if no other processes of interest, such as selection or constraints, are acting. The status of the ZFEL as a law and its epistemic role in evolutionary biology have been hotly debated by other philosophers (Barrett et al., 2012; Brandon and McShea, 2012; Gouvêa, 2015), but some biologists have adopted it in their empirical analyses of macro-evolutionary trends (Smith and Donoghue, 2022), and McShea and Brandon have continued to develop it into a more quantitative, statistical formulation (McShea et al., 2019; Brandon and McShea, 2020; but see Gingerich, 2020).

McShea and Brandon's perspective on trait evolution is multi-level by conception: they define complexity as within-system variation in the properties of its parts, and diversity as variation in the properties among a population of systems. The ZFEL is also not restricted to a particular level of evolutionary units in the biological hierarchy, making it suitable for addressing the spectrum of units currently recognized as living fossils, including genes and species.

The primary message McShea and Brandon draw from the ZFEL is that change is the default or null expectation for all heritable traits, so biologists should treat stasis as a core phenomenon in need of explanation (McShea and Brandon, 2010, p. 113). They argue this is a radical change in how biologists view biologically significant patterns and therefore what demands special explanation. "Our goal is to create a framework that better enables us to empirically investigate when and where and how natural selection acts and interacts with other evolutionary forces… What the zero-force condition does is give us a neutral background against which to see selection in action" (McShea and Brandon, 2010, p. 104). More specifically, they argue that the ZFEL for diversity should be "the standard explanation for rising diversity in macroevolution" and that "from the ZFEL point of view, it is the long periods of stasis that are remarkable" in the fossil record (McShea and Brandon, 2010, p. 113).

While McShea and Brandon discuss statistical trends in the diversity or complexity of single traits, they have not yet applied the ZFEL to multivariate trends. The standard way that McShea and Brandon understand complexity, for example, is that it applies to multiple instances of a part within a system, e.g., multiple cells in a body or fenceposts in a fence. For example, we can apply ZFEL to predict that the complexity of the 32 types of teeth in humans, i.e., the variation we observe in the set of 32 types, should increase in time if unopposed by other forces. However, not all sets of traits can be readily interpreted this way—the length ratios in the trilobite example are a good example, since these represent different measurements of the properties of a single homologous structure. The ZFEL therefore does not provide a baseline expectation for how the complexity of all traits will evolve within a lineage, since Brandon and McShea's definition of complexity does not apply universally. However, the ZFEL does predict increasing diversity for all traits among lineages in the absence of other forces.

A simple extension of the ZFEL framework could therefore be to say that a bundle of traits is consistent with ZFEL if a majority of the traits individually shows the predicted growth of complexity or diversity, depending on the particular comparison we want to make. In contrast, we can say the bundle exhibits stability if a majority of traits show less growth in complexity or diversity than predicted by ZFEL. Methodologically we determine this fact by testing whether ZFEL is rejected as a null hypothesis independently for each trait. In-between cases, i.e., where the traits show mosaic evolution from

5.2. SMS framework

A second evidential framework developed within evolutionary biology uses information-theoretic model selection methods to analyze trait evolution in fossil lineages (Hunt, 2006, 2008b; Hopkins and Lidgard, 2012; Reitan et al., 2012; Hunt et al., 2015; Voje, 2018, 2020). While this framework does not have a single, shared name, I will refer to it as the statistical model selection framework for convenience. The immediate origins of the framework is Hunt's work in the 2000s that showed how information-theoretic methods of model selection, such as the Akaike Information Criterion (AIC), provide a general and powerful approach to detecting patterns of phenotypic change in fossil lineages such as stasis, Brownian motion, and a directional trend. Hunt (2006) originally called it the likelihood framework, and Voje (2020) refers to it as Hunt's framework, but neither is fully apt to the range of contributors and techniques used. The roots of the framework, though, extend further back to the landmark of Simpson (1944) book on evolutionary tempo and mode and early applications of time series models to fossil lineage data by Raup (1977).

For our purposes here, we can treat the SMS framework as having two main components: a set of candidate models for analyzing data collected from fossil trait series, and a set of statistical methods and criteria for identifying the best-supported model among the candidate set. The core three models are stasis, a random walk, and a linear directional trend, as described above. Other models in the literature include the Ornstein-Uhlenbeck (Hunt, 2008a) and decelerated evolution models (Voje, 2020), but I omit these for space reasons. One can also combine these models in a piecewise fashion to fit different dynamics in multiple periods using changepoint analysis (Hunt et al., 2015), and Voje (2023) has recently published a software package for estimating multivariate models.

Crucially, the set of candidate models is motivated by the goals of distinguishing general classes (modes) of evolutionary dynamics and estimating biological meaningful parameters, which can be interpreted as evolutionary rates (Hunt, 2012; Voje, 2016). The SMS framework therefore exemplifies model-based science rather than the Newtonian laws of nature approach that inspired the ZFEL. The SMS framework, moreover, treats the candidate models equally when applying them to data rather than identifying one or more models as more fundamental or epistemically prior. This was a key innovation introduced by Hunt (2006) through adopting information-theoretic model selection methods, e.g., using the AIC and related criteria. Instead of comparing one or more alternative models to a single null model, AIC-type methods use penalized likelihoods to estimate the distance of each fitted model to the true, data-generating distribution. The model with the lowest AIC score is the best-supported, and one can calculate the degree of support by comparing the difference or normalized ratio of model scores. This shift in reference point for calculating statistical evidence allows the AIC-type approach to compare any number of models simultaneously without requiring them to follow a nested structure or sequence. A Bayesian approach is also possible but has not been widely adopted in the literature (see Hannisdal, 2006).

5.3. Evidential criteria in the ostracod example

McShea's recent work developing a quantitative formulation of the ZFEL (McShea et al., 2019) allows us to compare it to the SMS framework on a shared dataset of body size evolution in 11 ostracod lineages. Ostracods are a group of about 13,000 species of small crustaceans in the taxonomic class Ostracoda that are abundant in fresh water and marine environments. Biologists do not generally treat copepods as living fossils, though, and the dataset consists of univariate body size measurements for each lineage, so this example serves mainly to illustrate how the ZFEL and SMS frameworks differ in their use of model selection methods, with the ZFEL relying on null hypothesis testing and SMS on AIC-type methods.

In fact, the dataset was first collected and analyzed by Hunt et al. (2010) using the SMS framework to study climate-driven trends in body-size: in particular, Bergmann's Rule that species tend to have larger bodies in cooler climates and smaller bodies in warmer climates. According to Hunt et al. (2010, p. 1,256), "we are in the unusual position of having good reason to believe that the pattern is adaptive… but lacking a clear understanding of what, exactly, selection is acting upon." Their study aimed to explore whether Bergmann's Rule could be observed operating as evolution within fossil lineages at a particular site over time during periods of environmental warming and cooling. On this evolutionary interpretation, Bergmann's Rule would be an example of natural selection acting in a correlated way on body size. This would also entail statistical deviations from the pattern of trait diversification predicted by ZFEL, which expects each lineage's body size to evolve independently as a random walk uncorrelated with environmental temperature.

Hunt et al. collected trait measurements from 19 ostracod lineages they found in a deep sea drilling core sampled from a site in the Indian Ocean. They determined that the core sampling spans a time period of roughly 40 to 0.2 million years ago (Ma), and they used isotope and other environmental data sources to partition the core into three segments: an early period of cooling (40–30 Ma), a middle period of little net climate change (30–14 Ma), and a late period of cooling (14–0 Ma; Hunt et al., 2010, p. 1,259). Bergmann's Rule would therefore predict a trend of increasing body size in the early and late periods but not the middle. Figure 2 shows their main results, with the 19 lineages split across four panels to make for easier reading.

Their main finding is evidence for body size increasing as a function of temperature during the early and late periods of cooling, supporting an evolutionary mechanism for Bergmann's Rule in this case. They also find no evidence of a positive directional trend in body size during the middle period. As they summarize, "We suggest that our results support a view in which all of the directionality in body-size evolution stems from trends in environmental conditions (temperature or correlated variables in this case)" (Hunt et al., 2010, p. 1,265).

Using this example, we can compare the SMS and ZFEL frameworks along three main dimensions: (1) which data are analyzed, (2) what models are considered, and (3) how statistical evidence is calculated. In terms of data, Hunt et al. (2010) split the observations into three periods as we noted, but in their analysis McShea et al. (2019) only consider the middle period with low net climate change. McShea et al. (2019) do not state why, but given Hunt et al's (2010) findings that other evolutionary forces are acting during the early and late periods, the ZFEL is perhaps more interesting to test in the middle period where body size has weak correlation with environmental
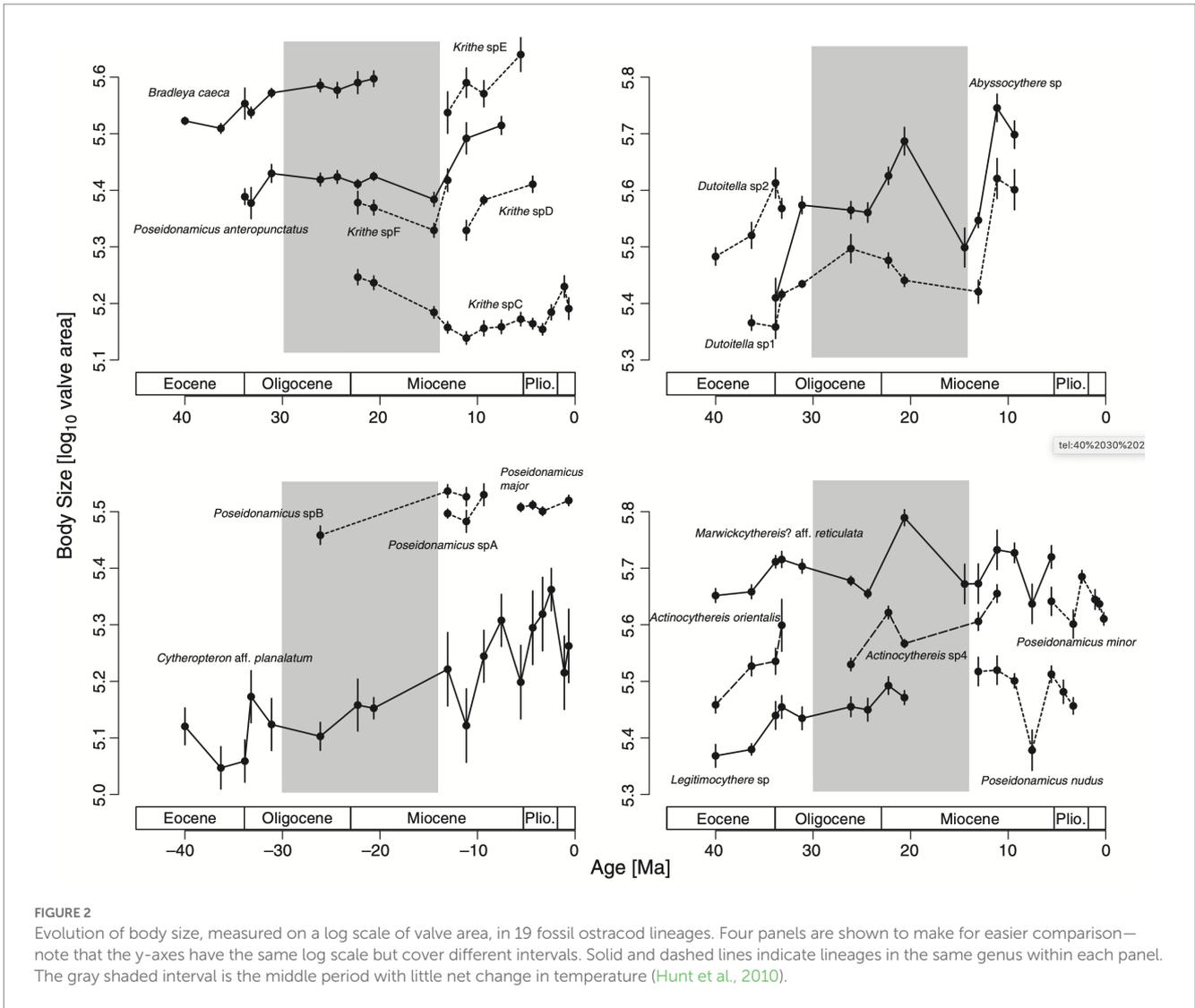
FIGURE 2
Evolution of body size, measured on a log scale of valve area, in 19 fossil ostracod lineages. Four panels are shown to make for easier comparison—note that the y-axes have the same log scale but cover different intervals. Solid and dashed lines indicate lineages in the same genus within each panel. The gray shaded interval is the middle period with little net change in temperature (Hunt et al., 2010).

change. For the models considered, Hunt et al. (2010) applied three basic types of model: a random walk model; a model with a linear trend, independent of the environment; and a model with a linear dependence on temperature change. McShea et al. (2019), in contrast, use a single model that describes the behavior of pairs of uncorrelated random walks. To calculate statistical evidence, Hunt et al. (2010) used the corrected Akaike Information Criterion and associated Akaike weights to compare the relative support for each of the models. This does not require a null model and compares the difference of the best and second-best model scores. In contrast, the ZFEL framework uses a value of p test to see if the uncorrelated random walk model can be rejected as statistically unlikely given the data. Agreeing with Hunt et al's (2010) results, McShea et al. (2019) find that the *p*-values for the null hypothesis test are larger than 0.05. The uncorrelated random walk model therefore is not rejected by the test, indicating the data are consistent with ZFEL's predictions.

5.4. Critical comparison of frameworks

The ZFEL framework has some key limitations relative to the SMS framework that make it ill-suited as a foundation for theorizing about multi-variate trait evolution in lineages. In particular, the SMS framework can incorporate the ZFEL without assigning it a privileged epistemic status in data analysis, avoiding some key problems. In short, there is no need for a Newtonian first law in the time series modeling of fossil lineages. The basic point of the ZFEL is that anytime a population of evolving systems experiences stochastic, heritable fluctuations in their properties, the population will tend to diverge unless other factors intervene. In practice, the ZFEL framework treats this view as entailing a special epistemic status for a particular model, the uncorrelated random walk, such that we must first reject this model as unlikely before considering alternatives. Within the SMS framework, however, one can interpret the ZFEL as a warrant for always including a random walk model in the set of candidate models. Using the AIC-type approach, the random walk model is then evaluated symmetrically with respect to other candidate models.

There is no requirement in the SMS framework, then, for the model representing the ZFEL to be distinguished as more fundamental or epistemically prior to other models in the candidate set. Instead, the ZFEL serves to justify a methodological norm that scientists should always include a model representing stochastic diffusion as a pattern of trait evolution when analyzing their datasets. Moreover, if ZFEL is presented as a null to be rejected, this will generally

be statistically underpowered when it comes to determining the evidence for slow rates or stasis against multiple other, non-nested models.

The ZFEL is also less suited to setting statistical expectations for mosaic evolution at the level of a whole lineage. The ZFEL is not more fundamental than negative selection when we consider expected patterns of change among all traits of a lineage: lineages that actually experienced simultaneous, unconstrained diffusion across all their traits would rapidly become unviable and go extinct. Moreover, some traits must stay constant in order for us to even observe variation between the states of homologous characters. We must therefore assume some constancy to observe increasing diversity and complexity.

6. Addressing validity challenges to the living fossil concept

Adopting a statistical frame of mind (Hagen, 2003) for classifying living fossils poses a risk as well as an opportunity for the concept's future. The model of "umbrella constructs" developed by sociologists Hirsch and Levin (1999) provides a useful way of contextualizing the importance of an evidential framework for living fossils. While Hirsch and Levin's original work analyzed the concept of organizational effectiveness in management research, they proposed a general "lifecycle" for interdisciplinary research organized around a broad, ambiguous concept that can apply to living fossils as well (Sterner, 2022; Lidgard and Kitchen, 2023). In particular, Hirsch and Levin identify four life-cycle stages: emerging excitement, the validity challenge, "tidying up with typologies," and construct collapse (Hirsch and Levin, 1999, p. 199). As Lidgard and Kitchen (2023) show in their systematic historical review of studies proposing living fossil taxa, the topic continues to attract growing interest that is also diversifying across biological fields and types of entities, e.g., genes as well as species or higher taxa. However, there is now an established literature challenging the scientific validity and utility of the living fossil concept, reflecting a potential crisis around the concept's future legitimacy. Hirsch and Levin distinguish three basic types of outcomes for efforts to "tidy up" a messy concept in response to validity challenges: overcoming the challenges to preserve the umbrella construct as a whole, restricting future research to only part of the original construct, or a general collapse of interest.

Critically, current discourse about living fossils lacks an expanded terminology that allows scientists to designate more specific meanings while still explicitly invoking the core phrase. Biologists, for example, have developed an expanded vocabulary modifying the core term of biological function, such as "molecular function," "conserved function," "developmental function," "gene function," and so on. These more precise terms are constructed by combining a modifier, e.g., "molecular," with the core term, "function," in a way that semantically clarifies the intended meaning while also constructing a lexical network among related terms. Without this terminological scaffolding, debate about living fossil classifications risk devolving into defending binary yes or no positions about a case, obscuring how participants are referencing different meanings for the concept.

Thinking with evidential frameworks such as ZFEL and SMS helps identify relevant factors for how the living fossil concept will develop in response to recent validity challenges. Both frameworks

rely on sets of candidate models to represent and discriminate between types of evolutionary phenomena, e.g., stasis, random walks, and directional trends. These models correspond to some of the major criteria that figure in definitions of living fossils, but biologists do not currently name or understand any of these models as primarily representing a type of living fossil. A living fossil lineage may be defined by showing morphological stasis throughout its existence, for example, but biologists think of the stasis model for fossil trait series in much more general terms. The concept of punctuated equilibrium has faced similar criticism in response to improved statistical methods in macro-evolutionary biology (Pennell et al., 2014). As it stands, then, the quantitative tools that biologists have developed to classify evolutionary modes are sufficient to address validity challenges for only part of the living fossil concept.

Treating living fossils as an investigative kind suggests a potential way to fill this gap in candidate models that specifically represent types of living fossils. As I've suggested, proposing new definitions can be fruitful if it enhances evidential standards while accommodating multiple understandings. As an illustration, let us return to the Stubbs et al.'s (2021) analysis of crocodilians, where they find no evidence of a clade-wide slowdown of evolutionary rates yet observe that skull morphology remains constrained compared to closely related extinct taxa. "A so-called 'living fossil' clade such as the crocodylomorphs may show slow rates in some sub-clades at certain times, but it had, and has, the potential for fast rates and rapid morphological diversification" (Stubbs et al., 2021, p. 8). This conclusion seems initially to be a rejection of the living fossil label, but it can be recast in positive terms as a novel type of living fossil. We could formulate the type as a "competition-constrained living fossil" that is defined by the following criteria:

1. The taxon shows no prolonged fast rate excursions or long intervals with rapid rates.
2. However, the taxon also does not show a substantial slowdown in evolutionary rates representing stasis or stabilizing selection.
3. The taxon accumulates morphological disparity steadily.
4. However, the taxon occupies a restricted morphological space relative to other related groups.
5. This restricted range of variation is due to ecological competition with other groups.

Rhetorically, this definition does not apply to the living fossil concept as a whole, but rather fits within the aim of identifying more precise types of living fossils that can be subjected to rigorous statistical analysis. Proposing an explicit definition in this fashion puts us in a better position to examine the adequacy of the data collection and analysis norms that Stubbs et al. followed in their study for application to other possible living fossils. For example, their analysis relied on a principal component analysis of selected traits from the skull and jaw that they expected to be highly ecologically meaningful, but their analytical methods may not generalize to cases featuring a more heterogenous set of traits under varying degrees and types of selection. Their use of Brownian motion models also conflates stasis with a low evolutionary rate, even though evolutionary tempo does not correlate with mode in other cases (Voje, 2016). Future work will need to determine whether "tidied-up" definitions such as this one can be buttressed

by general evidential norms and practices, or it may turn out that the parts of the definition will have more persistence than the whole.

7. Conclusion

The case of living fossils suggests that shared evidential norms are an important but overlooked complement to shared explanatory aims and criteria in interdisciplinary research (Love, 2008). Shared evidential norms for classifying cases of living fossils are critically important for productive interdisciplinary research. If researchers embrace the mess of sustaining multiple definitions for living fossils, as seems necessary given the diversity of its uses (Lidgard and Kitchen, 2023), then the same examples will be classified differently depending on definition. This is challenging enough, but it is manageable if researchers can readily translate examples and conclusions among definitions. If researchers cannot agree about how the same definition applies to potential cases, however, the concept's overall utility threatens to fall apart.

In response, I argued that definitional debates can be productive if they drive improved evidential norms, and I proposed treating the living fossil concept as an investigative kind as a useful way to address this need. The evidential criteria for classifying living fossils based on multi-trait and multi-level relationships are contentious and underspecified in many cases. Two issues in particular stood out as rooted in part–whole ambiguities. First, how do the properties of a system's parts "add up" to provide evidence for categorizing its behavior as a whole? Second, how should biologists distinguish chance vs. genuine patterns of stability or persistence in multi-trait datasets? This poses a threat to the explanatory interest of the living fossil concept, which depends on identifying biologically meaningful bundles of characters that show unusual stability or persistence. I analyzed how two different evidential frameworks draw on biological and statistical theory to justify classificatory practices and how they might apply to living fossils and described how. I found that the frameworks disagree about the appropriate statistical methods researchers should use, and both frameworks have important gaps remaining for the multi-trait, multi-level properties of living fossils. Nonetheless, I argued that the SMS framework is better suited to guide future research because it avoids the biologically contentious and statistically inefficient privileging of random walks as a null or default pattern to reject.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Avise, J. C., Nelson, W. S., and Sugita, H. (1994). A speciational history of "living fossils": molecular evolutionary patterns in horseshoe crabs. *Evolution* 48, 1986–2001. doi: 10.1111/j.1558-5646.1994.tb02228.x

Barrett, M., Clatterbuck, H., Goldsby, M., Helgeson, C., McLoone, B., Pearce, T., et al. (2012). Puzzles for ZFEL, McShea and Brandon's zero force evolutionary law. *Biol Philos* 27, 723–735. doi: 10.1007/s10539-012-9321-7

Bennett, D., Sutton, J. M. D., and Turvey, S. T. (2018). Quantifying the living fossil concept. *Palaeontol Electron* 21:750. doi: 10.26879/750

Brandon, R. N., and McShea, D. W. (2012). Four solutions for four puzzles. *Biol Philos* 27, 737–744. doi: 10.1007/s10539-012-9330-6

Brandon, R. N., and McShea, D. W. (2020). *The missing two-thirds of evolutionary theory*. 1st edn. Cambridge, UK: Cambridge University Press.

Brigandt, I. (2003). Species pluralism does not imply species eliminativism. *Philos. Sci.* 70, 1305–1316. doi: 10.1086/377409

Brigandt, I. (2020). "How are biology concepts used and transformed?" in *Philosophy of science for biologists*. eds. K. Kampourakis and T. Uller. 1st edn. Cambridge, UK: Cambridge University Press, 79–101.

Brigandt, I., and Love, A. C. (2012). Conceptualizing evolutionary novelty: moving beyond definitional debates. *J. Exp. Zool. B Mol. Dev. Evol.* 318, 417–427. doi: 10.1002/jez.b.22461

Carnall, M. (2016). Let's make living fossils extinct. *Guardian Sci.* Available at: https://www.theguardian.com/science/2016/jul/06/why-its-time-to-make-living-fossils-extinct

Chang, H. (2004). *Inventing temperature: measurement and scientific progress*. Oxford, UK: Oxford University Press.

DeSilva, J. M. (2018). "Mosaic evolution" in *The international encyclopedia of biological anthropology*. ed. W. Trevathan (Chichester: John Wiley & Sons), 1–2. doi: 10.1002/9781118584538

Felice, R. N., Pol, D., and Goswami, A. (2021). Complex macroevolutionary dynamics underly the evolution of the crocodyliform skull. *Proc. R. Soc. B Biol. Sci.* 288:20210919. doi: 10.1098/rspb.2021.0919

Gemmell, N. J., Rutherford, K., Prost, S., Tollis, M., Winter, D., Macey, R., et al. (2020). The tuatara genome reveals ancient features of amniote evolution. *Nature* 584, 403–409. doi: 10.1038/s41586-020-2561-9

Gingerich, P. D. (2020). Empirical lognormality of biological variation: implications for the "zero-force evolutionary law." *bioRxiv* [Preprint]. Available at: biorxiv.org/content/10.1101/2020.04.28.066563v1

Gouvêa, D. Y. (2015). Explanation and the evolutionary first law(s). *Philos. Sci.* 82, 363–382. doi: 10.1086/681603

Haber, M., and Velasco, J. (2021). "Phylogenetic inference" in *The Stanford encyclopedia of philosophy*. ed. E. N. Zalta. Available at: https://plato.stanford.edu/archives/win2021/entries/phylogenetic-inference/

Hagen, J. (2003). The statistical frame of mind in systematic biology from quantitative zoology to biometry. *J. Hist. Biol.* 36, 353–384. doi: 10.1023/A:1024479322226

Hannisdal, B. (2006). Phenotypic evolution in the fossil record: numerical experiments. *J. Geol.* 114, 133–153. doi: 10.1086/499569

Hay, J. M., Subramanian, S., Millar, C. D., Mohandesan, E., and Lambert, D. M. (2008). Rapid molecular evolution in a living fossil. *Trends Genet.* 24, 106–109. doi: 10.1016/j.tig.2007.12.002

Herrera-Flores, J. A., Stubbs, T. L., and Benton, M. J. (2017). Macroevolutionary patterns in Rhynchocephalia: is the tuatara (*Sphenodon punctatus*) a living fossil? *Palaeontology* 60, 319–328. doi: 10.1111/pala.12284

Hirsch, P. M., and Levin, D. Z. (1999). Umbrella advocates versus validity police: a life-cycle model. *Organ. Sci.* 10, 199–212. doi: 10.1287/orsc.10.2.199

Hopkins, M. J., and Lidgard, S. (2012). Evolutionary mode routinely varies among morphological traits within fossil species lineages. *Proc. Natl. Acad. Sci.* 109, 20520–20525. doi: 10.1073/pnas.1209901109

Hunt, G. (2006). Fitting and comparing models of phyletic evolution: random walks and beyond. *Paleobiology* 32, 578–601. doi: 10.1666/05070.1

Hunt, G. (2008a). Gradual or pulsed evolution: when should punctuational explanations be preferred? *Paleobiology* 34, 360–377. doi: 10.1666/07073.1

Hunt, G. (2008b). Evolutionary patterns within fossil lineages: model-based assessment of modes, rates, punctuations and process. *Paleontol Soc Papers* 14, 117–131. doi: 10.1017/S1089332600001649

Hunt, G. (2012). Measuring rates of phenotypic evolution and the inseparability of tempo and mode. *Paleobiology* 38, 351–373. doi: 10.1666/11047.1

Hunt, G., Hopkins, M. J., and Lidgard, S. (2015). Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proc. Natl. Acad. Sci.* 112, 4885–4890. doi: 10.1073/pnas.1403662111

Hunt, G., Wicaksono, S. A., Brown, J. E., and Macleod, K. G. (2010). Climate-driven body-size trends in the ostracod fauna of the deep Indian Ocean. *Palaeontology* 53, 1255–1268. doi: 10.1111/j.1475-4983.2010.01007.x

Janecka, J., Chowdhary, B., and Murphy, W. (2012). Exploring the correlations between sequence evolution rate and phenotypic divergence across the mammalian tree provides insights into adaptive evolution. *J. Biosci.* 37, 897–909. doi: 10.1007/s12038-012-9254-y

Kellert, S. H., Longino, H. E., and Waters, C. K., (2006). *Scientific pluralism. Minnesota studies in the philosophy of science*, 19. Minneapolis, MN: University of Minnesota Press.

Lamsdell, J. C. (2019). Evolutionary history of the dynamic horseshoe crab. *Int Wader Stud* 21:173. doi: 10.18194/db.00173

Lidgard, S., and Kitchen, E. (2023). Revealing the rise of a living fossil menagerie. *Front. Ecol. Evol.* 11:764. doi: 10.3389/fevo.2023.1112764

Lidgard, S., and Love, A. C. (2018). Rethinking living fossils. *Bioscience* 68, 760–770. doi: 10.1093/biosci/biy084

Lidgard, S., and Love, A. C. (2021). The living fossil concept: reply to Turner. *Biol Philos* 36:13. doi: 10.1007/s10539-021-09789-z

Love, A. C. (2008). Explaining evolutionary innovations and novelties: criteria of explanatory adequacy and epistemological prerequisites. *Philos. Sci.* 75, 874–886. doi: 10.1086/594531

Ludwig, D., and Ruphy, S. (2021). "Scientific pluralism" in *The Stanford encyclopedia of philosophy*. ed. E. N. Zalta. Available at: https://plato.stanford.edu/archives/win2021/entries/scientific-pluralism/

McShea, D. W., and Brandon, R. N.. (2010). *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*. Chicago, IL: University of Chicago Press.

McShea, D. W., Wang, S. C., and Brandon, R. N. (2019). A quantitative formulation of biology's first law. *Evolution* 73, 1101–1115. doi: 10.1111/evo.13735

Meloro, C., and Jones, M. E. H. (2012). Tooth and cranial disparity in the fossil relatives of *Sphenodon* (Rhynchocephalia) dispute the persistent "living fossil" label. *J. Evol. Biol.* 25, 2194–2209. doi: 10.1111/j.1420-9101.2012.02595.x

Miller, H. C., Moore, J. A., Allendorf, F. W., and Daugherty, C. H. (2009). The evolutionary rate of tuatara revisited. *Trends Genet.* 25, 13–15. doi: 10.1016/j.tig.2008.09.007

Nersessian, N. J. (2017). "Conceptual change" in *A companion to cognitive science*. ed. W. Bechtel and G. Graham (Malden, MA: John Wiley & Sons), 157–166. doi: 10.1002/9781405164535.ch7

Nickles, T. (2017). "Scientific revolutions" in *The Stanford encyclopedia of philosophy*. ed. E. N. Zalta. Available at: https://plato.stanford.edu/archives/win2021/entries/scientific-revolutions/

Nickles, T. (2021). "Historicist theories of scientific rationality" in *The Stanford encyclopedia of philosophy*. ed. E. N. Zalta. Available at: https://plato.stanford.edu/archives/win2021/entries/rationality-historicist/

Parravicini, A., and Pievani, T. (2019). Mosaic evolution in hominin phylogeny: meanings, implications, and explanations. *J. Anthropol. Sci.* 96, 45–68. doi: 10.4436/JASS.97001

Pennell, M. W., Harmon, L. J., and Uyeda, J. C. (2014). Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol.* 29, 23–32. doi: 10.1016/j.tree.2013.07.004

Pradeu, T., Lemoine, M., Khelfaoui, M., and Gingras, Y. (2021). Philosophy in science: can philosophers of science permeate through science and produce scientific knowledge? *Br. J. Philos. Sci.* doi: 10.1086/715518

Raup, D. M. (1977). "Stochastic models in evolutionary paleontology" in *Developments in palaeontology and stratigraphy*. ed. A. Hallam, vol. 5 (Amsterdam, Netherlands: Elsevier), 59–78.

Reitan, T., Schweder, T., and Henderiks, J. (2012). Phenotypic evolution studied by layered stochastic differential equations. *Annals Appl Stat* 6, 1531–1551. doi: 10.1214/12-AOAS559

Rescher, N. (2000). *Pluralism: against the demand for consensus. Clarendon library of logic and philosophy*. Oxford: Clarendon Press.

Simpson, G. G. (1944). *Tempo and mode in evolution*. 15. New York: Columbia University Press.

Smith, T. J., and Donoghue, P. C. J. (2022). Evolution of fungal phenotypic disparity. *Nat Ecol Evol* 6, 1489–1500. doi: 10.1038/s41559-022-01844-6

Stankowski, S., and Ravinet, M. (2021). Quantifying the use of species concepts. *Curr. Biol.* 31, R428–R429. doi: 10.1016/j.cub.2021.03.060

Sterner, B. (2022). Explaining ambiguity in scientific language. *Synthese* 200:354. doi: 10.1007/s11229-022-03792-x

Sterner, B., and Lidgard, S. (2014). The normative structure of mathematization in systematic biology. *Stud Hist Philos Biol Biomed Sci* 46, 44–54. doi: 10.1016/j.shpsc.2014.03.001

Sterner, B., and Lidgard, S. (2021). Objectivity and underdetermination in statistical model selection. *Br. J. Philos. Sci.* doi: 10.1086/716243

Sterner, B., Sen, A., and Witteveen, J. (2022). Consensus and scientific classification. *Knowl. Organ.* 49, 236–256. doi: 10.5771/0943-7444-2022-4-236

Stubbs, T. L., Pierce, S. E., Elsler, A., Anderson, P. S. L., Rayfield, E. J., and Benton, M. J. (2021). Ecological opportunity and the rise and fall of crocodylomorph evolutionary innovation. *Proc. R. Soc. B Biol. Sci.* 288:20210069. doi: 10.1098/rspb.2021.0069

Subramanian, S., Hay, J. M., Mohandesan, E., Millar, C. D., and Lambert, D. M. (2009). Molecular and morphological evolution in tuatara are decoupled. *Trends Genet.* 25, 16–18. doi: 10.1016/j.tig.2008.11.001

Turner, D. D. (2019). In defense of living fossils. *Biol Philos* 34, 1–22. doi: 10.1007/s10539-019-9678-y

Voje, K. L. (2016). Tempo does not correlate with mode in the fossil record. *Evolution* 70, 2678–2689. doi: 10.1111/evo.13090

Voje, K. L. (2018). Assessing adequacy of models of phyletic evolution in the fossil record. *Methods Ecol. Evol.* 9, 2402–2413. doi: 10.1111/2041-210X.13083

Voje, K. L. (2020). Testing eco-evolutionary predictions using fossil data: phyletic evolution following ecological opportunity. *Evolution* 74, 188–200. doi: 10.1111/evo.13838

Voje, K. L. (2023). Fitting and evaluating univariate and multivariate models of within-lineage evolution. *Paleobiology* April, 1–18. doi: 10.1017/pab.2023.10

Watkins, A. (2021). The epistemic value of the living fossils concept. *Philos. Sci.* 88, 1221–1233. doi: 10.1086/714875

Werth, A., and Shear, W. (2014). The evolutionary truth about living fossils. *Am. Sci.* 102, 434–443. doi: 10.1511/2014.111.434