Systematic Biology (2025), **0**, 0, *pp.* 1–**??**
doi:10.1093/sysbio/output

Evolutionary Tempo, Supertaxa and Living Fossils

Graham E. Budd[1,*] and Richard P. Mann[2]

[1] *Department of Earth Sciences, Palaeobiology, Uppsala University, Uppsala, SE 752 36, Sweden*
[2] *Department of Statistics, School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK*

**Corresponding author: email graham.budd@pal.uu.se*

Abstract

A relationship between the rate of molecular change and diversification has long been discussed, on both theoretical and empirical grounds. However, the effect on our understanding of evolutionary patterns is yet to be fully explored. Here we develop a new model, the Covariant Evolutionary Tempo (CET) model, with the aim of integrating patterns of diversification and molecular evolution within a framework of a continuously changing 'tempo' variable that acts as a master control for molecular, morphological and diversification rates. Importantly, tempo itself is treated as being variable at a rate proportional to its own value. This model predicts that diversity is dominated by a small number of extremely large clades at any historical epoch including the present; that these large clades are expected to be characterised by explosive early radiations accompanied by elevated rates of molecular evolution; and that extant organisms are likely to have evolved from species with unusually fast evolutionary rates. Under such a model, the amount of molecular change along a particular lineage is essentially independent of its height, which weakens the molecular clock hypothesis. Finally, our model explains the existence of 'living fossil' sister groups to large clades that are species poor and exhibit slow rates of morphological and molecular change. Our results demonstrate that the observed historical patterns of evolution can be modelled without invoking special evolutionary mechanisms or innovations that are unique to specific times or taxa, even when they are highly

non-uniform.

Key words: Patterns of diversification, Molecular clocks, living fossils

Introduction

The relationship between micro- and macroevolution has long been debated (Rolland et al., 2023; Jablonski, 2000; Erwin, 2000). A central question is the extent to which large-scale evolutionary patterns—observed in the fossil record and inferred from phylogenies—are shaped by the processes operating at the population level. Regardless of the outcome of this debate, however, there is often a *methodological* assumption of independence between microevolutionary changes (e.g., shifts in gene frequencies due to selection) and macroevolutionary patterns (e.g., diversification trends within a clade). Contemporary models of evolutionary history conceptualise the overall process as being governed by three independent components: the model of molecular substitution, the rate at which substitutions occur, and the nature of the branching process (Warnock and Wright, 2021). The simplest approach would be to employ a strict molecular clock with a Jukes-Cantor substitution model (Jukes and Cantor, 1969) on a known phylogeny, and assuming a fixed rate of branching—often represented by a homogeneous birth-death process (BDP) (Nee, 2006). Methodological advances, such as the development of relaxed clocks, now allow substitution rates to vary across the tree (see Dos Reis et al. (2016) for a review). Additionally, increasingly sophisticated models of molecular evolution have been introduced (Arenas, 2015). More recently, models have also emerged that incorporate variable diversification rates (see below), allowing for more complex representations of evolutionary trees, although the broad-scale patterns resulting from such models remain relatively unexplored.

Increasing sophistication in modelling ability has naturally also fuelled attempts to

understand the causes behind the variation being captured. To take molecular substitution rate variation first: two broad hypotheses exist about its causes. The first encompasses a range from mutational effects to features of the entire organism (such as body size or generation time); and the second is a 'speciation rate hypothesis' that links molecular change to speciation (Jobson and Albert, 2002). There are sound empirical and conceptual reasons for thinking that speciation and molecular change may well be intimately related (Hua and Bromham, 2017), and attempts have sometimes been made to consider them jointly (e.g. Sarver et al. (2019); Ritchie et al. (2022b)). Indeed, Eo and DeWoody go so far as to claim that "One of the most basic predictions in evolutionary biology is that the rate of diversification along a particular branch of the tree of life is some function of the rate of genome evolution on that branch." (Eo and DeWoody (2010), p. 3587). Provocative evidence for a close correlation of the two processes is seen for example in the early history of arthropods (Lee et al., 2013), where early branches of the clade contain just as much molecular change as later branches despite being far shorter in duration (Budd and Mann, 2020b), at least when the tree height is constrained by the fossil record. However, this is just one of several studies that over the last few decades have debated a potential link between both morphological and molecular rates of change and rates of speciation (e.g., Bromham (2024); Rabosky et al. (2013); Xiang et al. (2004); Webster et al. (2003); Venditti and Pagel (2010); Lanfear et al. (2010); Berv and Field (2018); Barraclough and Savolainen (2001)), although it should be noted that not all studies have found clear evidence of this link (e.g., Goldie et al. (2011)). There are at least two factors that might cloud the relationship between diversification and molecular change through time. The first is the so-called 'node density' effect, wherein in clades with more terminals, a resulting greater number of internal nodes will recover more molecular change and thus generate a spurious relationship between clade size and amount of molecular change (Hugall and Lee, 2007). The second is that if a relaxed clock methodology is employed to ascertain the time of origin of a clade, then any early burst of molecular (or morphological (Beck and Lee,

2014)) change or indeed diversification is likely to be smoothed out by pushing the age of the root deeper (Budd and Mann, 2020b; Bromham, 2020, 2003; Beaulieu et al., 2015; Shafir et al., 2020). If one were simply to accept the result of the molecular clock, then the apparent elevated early rates could theoretically be explained as an artefact caused by "bunching up" the early lineages to artificially squeeze the clade into a too-narrow time interval (c.f. Bromham and Hendy (2000)). However, we have previously marshalled strong reasons for thinking that the fossil record in such instances is often reliable, in which case early bursts of diversification should be taken seriously and not dismissed as dating artefacts (Budd and Mann, 2020a,b, 2024; Holmes and Budd, 2022). As a result, the well-known mismatch between the explicit fossil record and molecular clock origination estimates for many major clades such as animals (Budd and Mann, 2020b), birds (Berv and Field, 2018), placental mammals (Budd and Mann, 2024) and angiosperms (Smith and Beaulieu, 2024), (Coiro et al., 2019) itself points to cryptic excess molecular change at the base of trees (Beaulieu et al., 2015; Berv and Field, 2018). Previous critiques of molecular clocks have focused on either inappropriate age priors (e.g. Budd and Mann (2024); Brown and Smith (2018)) or issues with rate heterogeneity (e.g. Bromham and Woolfit (2004); Berv and Field (2018)); below we will suggest these are effectively two sides of the same coin. Clearly, if the branching process and rate of molecular change really are correlated, then this would have a significant impact on our understanding of the patterns of evolutionary change through time (see Duchêne et al. (2017) for investigation and discussion of this point).

Causes of variation in diversification rates are likewise much debated (e.g. Moen and Morlon (2014)). It is clear that, similarly to the case of molecular evolution itself, rates of diversification must vary across the tree, as a single homogeneous BDP cannot possibly capture the true patterns of diversification reflected in evolutionary history (c.f. Benton and Emerson (2007)). Notwithstanding this, the homogeneous birth death process (BDP) (Nee, 2006) (in which rates of speciation and extinction are fixed) is still commonly

employed in molecular analysis, especially for dating purposes, although its inadequacies are increasingly being recognised (e.g. Khurana et al. (2024)).

Any attempt to investigate a link between rates of genetic/morphological evolution and speciation must reckon with the heterogeneous nature of all of these variables. Historically, rate heterogeneity has largely been addressed in one of two ways: either by assuming rate shifts occur at significant points (e.g., Soltis and Soltis (2016)), or by assuming broad secular variation, e.g. with declining rates through time across the entire tree (Nee et al., 1994b; Strathmann and Slatkin, 1983); or some combination of both (e.g. in BAMM (Rabosky et al., 2014)). More recent models have moved away from considering isolated rate shifts to allow rates to vary either in small frequent increments associated with speciations (Maliet et al., 2019; Shafir et al., 2020), or continuously through anagenetic diffusion (Quintero et al., 2024) (for other non-continuous models, see the review in the supplementary information of Maliet et al. (2019)). The primary goal of these models has been the *inference* of rates through time, based on molecular data from extant taxa (Barido-Sottani and Morlon, 2023) which has now been implemented in BEAST2 (Bouckaert et al., 2019), clearly a substantial step forward from homogeneous models. However, some forward simulation has also revealed that these models can generate clades that match empirical observation; in particular simulated clades are often imbalanced and 'stemmy' (Maliet et al., 2019). This suggests that diversification rate heterogeneity may be one key to understanding the patterns of modern diversity. This is largely because the distribution of modern diversity predicted by homogeneous or epochally time-varying BDPs is geometric (Nee et al., 1994b; Kendall, 1948), and this remains the case even when non-selective mass extinctions are considered (Budd and Mann, 2020a). However, a certain amount of evidence suggests that extant sizes are in fact over-dispersed relative to this expectation (Blum and François, 2006; Stadler et al., 2016). Consider, for example, the crown group animal phyla, which for the sake of argument we can assume all emerged around 500 Ma (Budd and Mann, 2024). Estimating total species

diversity in the phyla is fraught with difficulty, but even so the species count differs widely. For example, the phyla have an average diversity of c. 50,000 species, but the arthropods have a diversity of well over one million species, thus being over twenty times larger than expected. Under a geometric distribution this is essentially impossible ($p \sim 10^{-7}$). This pattern is seen repeated hierarchically: e.g. most arthropods are insects, and most insects appear to be hymenopterans (Forbes et al., 2018). Similarly, the angiosperms are much more diverse than any other plant clades (e.g. c. 300000 versus 1000 gymnosperms) and birds much more so than crocodiles in the archosaurs (c.10000 versus c. 85). In other words, the existence of Stanley's "supertaxa" (Stanley, 1998) does not seem compatible with a purely geometric distribution of clade sizes as predicted by the homogeneous BDP. In addition, clade sizes show a complex relationship with age that is not easily explained by homogeneous diversification (Rabosky, 2010; Magallon and Sanderson, 2001; McPeek and Brown, 2007), and indeed attempts to estimate absolute diversification rates within a clade suggest several orders of magnitude variation (Magallon and Sanderson, 2001). It thus seems that clade sizes do often appear overdispersed relative to any expected geometric distribution (Khurana et al., 2024).

Taking these empirical findings together, and noting the apparent importance of rate heterogeneity across both microscopic and macroscopic evolutionary scales (Henao-Diaz and Pennell, 2023), it seems that a need exists for a synthesis that unites molecular evolution and species diversification, in which both vary through time. In this paper, then, we develop a model of diversification and molecular change in which all evolutionary rates covary, being controlled by a single variable evolutionary *tempo* that differs both between species, and within a species over time. Although our model does not depend on a particular instantiation of tempo, we nevertheless offer some suggestions about how it might be encoded in a realistic way in the genome below (see schematic for genetic encoding of tempo in Appendix 1). Our analysis of this model will show that it is consistent with the concentration of species into relatively few 'supertaxa' (Stanley, 1998);

that it offers a resolution to conflict between the fossil record and molecular clocks; and that it makes new predictions about the early history of major clades and the fate of the smaller clades that constitute the remaining part of modern diversity. Because of the way we formulate the model, it is amenable to numerical solution that allows us to investigate its general features, as opposed to simulations that would show the outcomes of rates over specific trees.

Methods and Materials

Model outline

As indicated above, heterogeneity in rates of speciation and extinction are key to explaining important empirical features of diversification. We here extend earlier approaches to model such heterogeneity (Rabosky et al., 2014; Maliet et al., 2019; Ritchie et al., 2022a; Quintero et al., 2024), and create a BDP model in which rates of speciation and extinction vary continuously and *covariantly* through anagenetic diffusion. We call this model the Covariant Evolutionary Tempo (CET) model. Under CET, all evolutionary rates are specific to a given taxon at a specific moment in time. Our model is close in formulation to that of Quintero et al. (2024). However, whereas they model this variation in speciation and extinction rates as geometric Brownian motion with an overall drift, and treat speciation and extinction independently, we instead posit that there exist baseline rates of speciation (λ) and extinction (μ) that are linearly modulated by a new variable we label as *tempo*, τ, which controls the relative rates of all evolutionary processes. At any given time a taxon with tempo τ has a speciation rate $\tau\lambda$ and an extinction rate $\tau\mu$.

This model is fully covariant, in that all rates are linked directly to τ; in effect the tempo represents a local speeding-up or slowing-down of evolutionary time, such that all processes happen faster or slower. In particular, we posit that tempo *itself* varies through time, and because we posit that tempo is in some way genetically encoded, this implies

that the evolution of τ itself proceeds at a rate proportional to τ, since the effect on molecular rates of mutation will obtain upon whichever part of the genome is responsible for this encoding. Specifically we model the log-tempo ($x = \log \tau$) as evolving according to a modified Ornstein–Uhlenbeck (OU) process that incorporates the effect of the tempo itself on all rates:

$$dx = -\theta e^x x dt + \sqrt{2\theta s^2 e^x} dW \tag{1}$$

where dW represents an incremental change from a Wiener process (popularly known as Brownian motion). We impose this model for the evolution of the log-tempo x since the tempo itself is constrained to be positive. The parameters of this stochastic differential equation are the mean reversion rate θ and the stationary variance of the process, s^2. The e^x terms in this equation come from the self-interaction of the tempo, which as well as multiplying the rate of all other processes also determines the rate at which it evolves itself, such that the effective increment of time is $\tau dt = e^x dt$. Our use of an OU process is motivated by two considerations. First, as we shall show, a Wiener process without a restoring force would lead to a runaway effect, where tempos increase without limit. Secondly, in Appendix 1 we describe a plausible schematic for how tempo is inherited that produces an inherent reversion to a mean value via entropic forces.

As we show in Appendix 1, this results in a drift-diffusion partial differential equation for the generating function of the resulting birth-death process:

$$\frac{\partial G_x}{\partial t} = e^x \left((\lambda G_x - \mu)(G_x - 1) - \theta x \frac{\partial G_x}{\partial x} + \theta s^2 \frac{\partial^2 G_x}{\partial x^2} \right) \tag{2}$$

where $G_x(t, z) = \sum_{n=0}^{\infty} P_n(t, x) z^n$, with $P_n(t, x)$ being the probability of generating n species over time t in a process starting with log-tempo x. Solving this equation for an initial condition $G_x(t = 0, z) = z$ provides the value of the generating function $G_x(t, z)$.

Equation 2 does not appear to permit solution in closed form, except for the long-term extinction probability $G_x(t, z = 0)$ for $t \to \infty$, which is $\frac{\mu}{\lambda}$ for all x, and is therefore tempo invariant. More generally, equation 2 can be straightforwardly solved numerically. The values of $P_n(t, x)$ can be retrieved from this generating function by

Fourier inversion (see Appendix 1).

We can derive further equations specifying the evolution of the mean number of species generated by the process over time, the expected number of lineages (species that will have modern descendants) and the distribution of tempos over time. Derivation of these equations is described in Appendix 1. The most important of these equations specifies the evolution of the mean number of species through time. Given a generating function G_x, the mean of the distribution, $N_x(t) \equiv \mathbb{E}(n \mid t, x)$ is given by:

$$N_x(t) = \left.\frac{\partial G_x(t,z)}{\partial z}\right|_{z=1} \tag{3}$$

Using this relation, equation 2 can be transformed into a simpler, linear form to represent the dynamics of the mean:

$$\frac{\partial N_x}{\partial t} = e^x \left(r N_x - \theta x \frac{\partial N_x}{\partial x} + s^2 \theta \frac{\partial^2 N_x}{\partial x^2} \right), \tag{4}$$

where $r = \lambda - \mu$ is the baseline net diversification rate. This equation reveals the key dynamics of the process: the expected number of species with log-tempo x locally increases exponentially at the rate r modulated by $\tau = e^x$. At the same time a drift-diffusion process modifies the tempo of each species, such that species tend to move towards a log-tempo of 0 (i.e. $\tau = 1$).

Justification for a Covariant Theory

Why should all evolutionary rates be covariant? As we have discussed above, previous birth-death models have allowed for independent variation in speciation and extinction (while in practice sometimes holding one of these constant), while the rates of molecular evolution have been assumed (generally implicitly) to be completely independent of diversification rates. In one sense our choice is pragmatic: we seek to explore the consequences of linking changing rates of molecular evolution to diversification rates, and the most parsimonious way to do this is to impose a perfect correlation between the two. Allowing for speciation and extinction rates to vary independently (or with some

non-unitary correlation) would greatly complicate the mathematical formulation of the birth-death model and its analysis, and cloud its implications. Empirically we are also strongly motivated by the apparently close (inverse) correlation between rates of molecular evolution and branch durations in for example Lee et al. (2013) and other studies, as noted in the introduction. Finally our choice is also theoretically informed. It is clear that as speciation and extinction vary, they must remain close to one another over time; a sustained period of much higher speciation will quickly produce an unrealistically large number of species, while a period of greater extinction than speciation will almost certainly drive the clade to extinction. Indeed, the linkage between the two has been formulated by Marshall as the third of his five "paleobiological laws" (see Marshall (2017) for discussion and justification of this point). Moreover, we expect that rates of speciation and extinction may largely be driven by the same causal factors, e.g., generation times and population size (for a classical discussion of the various links between speciation and extinction rates, see Stanley (1990), and more recently Greenberg and Mooers (2017)). Therefore, while we anticipate significant deviations from covariance between these processes at sufficiently short time scales, we expect it to be a realistic first-order approximation when considering rates on the scale of millions of years. We also note that although most discussions of molecular evolution have considered a link with speciation, we consider that in practice this implies a link with extinction too, for the reasons given above.

As far as our model is concerned, we note that many of the factors operating on speciation rates are also likely to affect molecular rates of change. For example, Bromham has stressed the need to consider the genome itself as a life-history trait ((Bromham, 2003, 2009, 2020), and thus open to the same influences (population size, generation time, etc) as other traits. Thus, under such a view of evolution, small body size or small populations might both influence speciation rate (Martin, 2017; Cooney and Thomas, 2021) and molecular evolution rates (Bromham, 2020) together, thus uniting the two broad ways of considering the causes of molecular change (Jobson and Albert, 2002). Naturally, such a

linkage between the two might itself vary, but in order to investigate its general effects, and certainly to greatly simplify the analysis, we have chosen a model with complete linkage.

Few studies have shown a convincing direct link between molecular substitution rates and phenotypic change (Bromham and Woolfit, 2004). Nevertheless, the two may be indirectly linked by other factors such as speciation rate, as both phenotypic and molecular are plausibly linked to speciation (for discussion of this point with some examples such as placental mammals and lungfish, see Budd and Mann (2018)). As we suggest below, some empirical evidence points to this being true, at least in some clades.

Results

We analysed our model by solving the probabilistic equations given above to obtain distributions at different time epochs, rather than by direct simulation of the tree evolution. Notably, our analysis does not provide a probability distribution over specific trees, but over coarser-grained variables such as diversity. It is not our goal to quantitatively fit our model to the modern diversity or evolutionary history of any specific clade, but rather to reveal the qualitative features the model predicts. Throughout we use a core set of parameters $\lambda = 0.51$ per species per myrs, $\mu = 0.5$ per species per myrs, $\theta = 0.01$/myrs, $s = 1$. These parameters are chosen to reflect reasonable expectations about the real evolutionary process: a baseline extinction rate of $\mu = 0.5$ per species per myrs comports with that chosen in previous analyses (e.g. Budd and Mann (2018)) and, combined with a speciation rate of $\lambda = 0.51$ per species per myrs is consistent with a typical species existing for c. 1 myrs, in broad agreement with the fossil record (see. e.g. Budd and Mann (2018)). The speciation rate is chosen to be of similar magnitude to the extinction rate, such that extinction plays a significant role in the evolutionary dynamics (Marshall, 2017) but is otherwise arbitrary. We choose a mean-reversion parameter $\theta = 0.01$/myrs to be equal to the net diversification rate as we will later show that if $r = \theta$ then the mean log-tempo converges to 0 (see Appendix 1, equation 40). Although this

choice is mathematically convenient, we do not expect that it represents any necessary feature of the evolutionary process, nor do the general features of our results depend on it. Finally the diffusion parameter $s = 1$ is chosen to be large enough to produce significant effects of the diffusive dynamics, and otherwise is simply a mathematically convenient choice.

Distribution of clade sizes

We solved equation 2 for times $0 \leqslant t \leqslant 500$ myrs and starting log-tempos $-10 < x < 10$ and performed a Fourier inversion (see Appendix 1) to retrieve the implied probability distribution $P_n(t = 500\text{myrs}, x)$. The distribution of clade sizes for a clade that starts with log-tempo $x = 0$, excluding clades of size zero, is shown in Figure 1A. The clade sizes follow a distribution that differs strongly from the geometric distribution expected under a typical BDP (indicated by the dashed line, assuming the same mean clade size). This distribution is characterised by most clades being small, but with a few extremely large clades. This means that clades that are many times greater than average (either mean or median) are much more probable than under a standard birth-death process. A corollary of this is that clade size a typical species 'experiences' (i.e. the expected clade size of a randomly selected species) is c. 8 times greater than the mean clade size. For clarity, we here define the experienced and mean clade sizes as the sizes of clades containing living organisms that have the same time of origin (for example, the sizes of parent clades that are all 500 myrs old).

In Figure 1A we indicate both the mean clade size and the mean experienced clade size for illustration. This result should be compared to the equivalent result from a standard BDP where the mean experienced clade is only two times greater than the mean (Budd and Mann, 2018). This implies that the large majority of species we might encounter and/or study are contained in extremely large clades. Since clades are hierarchically structured this also implies that the diversity of any clade is likely to be

dominated by its largest sub-clade. To illustrate this we consider the two sister-groups of a clade originating 500 Ma, and calculate the expected proportion of the total diversity that is contained in one sister-group chosen at random. As shown in Figure 1B, the probability that a given proportion of total diversity is contained in a given sister-group is peaked strongly close to zero and one, indicating that one sister-group or the other typically contains the large majority of species in the clade as a whole. For example, there is a c. 50% chance that the larger sister group is at least 20 times larger than the other. This can be compared to the equivalent result under a standard BDP, in which the proportion of diversity contained in one sister-group is uniformly distributed between zero and one indicated by the dashed line), and thus the probability of such an imbalance is only 10%. This implies that diversity among clades of the same age tends to follow the Single Big Jump principle (Vezzani et al., 2019), whereby sums of heavy-tailed random variables are dominated by their largest component.

Diversification through time

The above analysis reveals the expected pattern of diversity in clades of a fixed age (500 myrs) which all start from a common ancestor with a typical tempo ($x = 0$). How does this pattern change through time, and between clades with different initial tempos? To explore these questions, we focused on how the expected clade size varies through time for different initial values of x. We numerically solved equation 4 to obtain the expected clade size as a function of time values $0 < t < 500$ myrs, and for different initial values of $x_0 \in \{-2, 0, 2\}$. In Figure 2A we show how the mean clade size varies through time for different initial tempos including clades that have gone extinct before the time in question. In Figure 2B we show the variation in the mean number of species through time conditioned on knowing that the clade survives to the present day (solid lines), and also the expected number of *lineages* (dashed lines) through time – these are species that have at least one descendant in the present day, and form the 'reconstructed process' that can

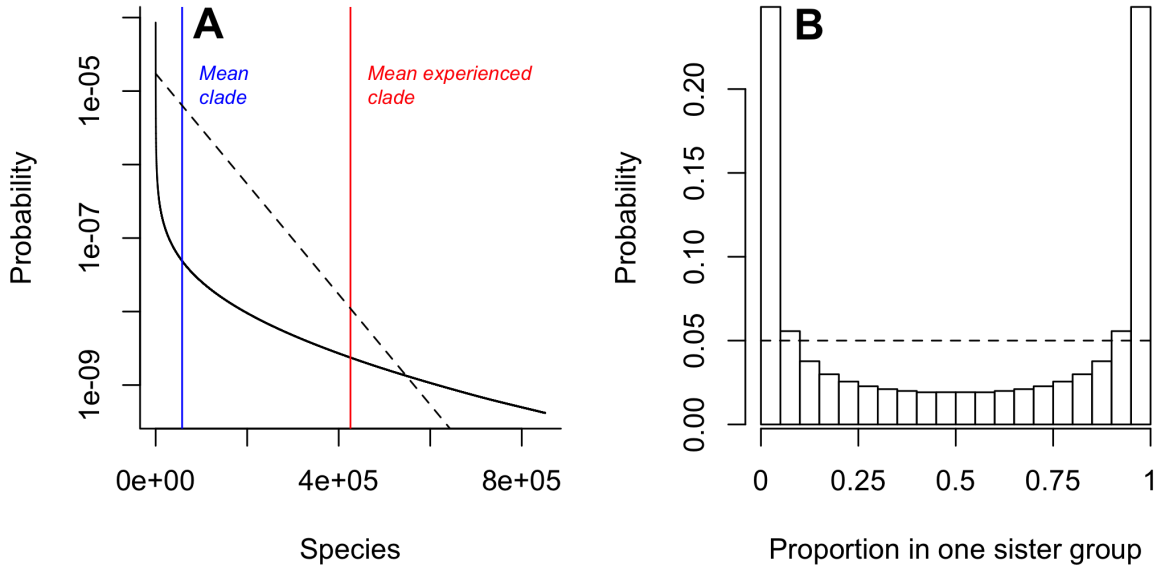

FIGURE 1. (A) Distribution of the number of species generated in clades that survive 500 myrs, with parameters $\lambda = 0.51$ per species per myrs, $\mu = 0.5$ per species per myrs, $\theta = 0.01$/myrs ,$s = 1$, and an initial log-tempo $x = 0$. Note the log scale on the y-axis. The distribution is long-tailed and is characterised by a high probability of few species ($P(n < 1000) \simeq 1/3$) and a long tail allowing some very large clades to be generated ($P(n > 50,000) \simeq 1/4$). The blue and red lines indicate the mean clade size (c. 60,000) and the mean experienced clade size of a randomly chosen taxon (c. 400,000) respectively, indicating that most taxa are found in very large clades. The dashed line shows the geometric distribution with the same mean expected under a standard BDP. (B) The probability distribution for the proportion of diversity contained within one randomly chosen sister group of a crown group, indicating that clades are typically highly imbalanced, with one sister group being much larger than the other. The dashed line shows the uniform distribution expected under a standard BDP

.

(in principle) be inferred from modern molecular data. Clades that survive to the present experience the 'Push of the Past' (Budd and Mann, 2018), an initial period of increased diversification when the clade is small. These results show that the initial tempo has a substantial impact on how the clade diversifies and its eventual expected size. As we would intuitively expect, clades with high tempos initially diversify more quickly, and conversely those with low tempos diversify slowly. However, after some period of time the rate of diversification becomes stable; initially high-tempo clades slow down and initially low-tempo clades speed up, such that all clades eventually diversify at the same fixed rate, as seen in emergence of parallel lines of growth from all three initial conditions.

The tempo of the root node of a clade therefore has transient effects that eventually

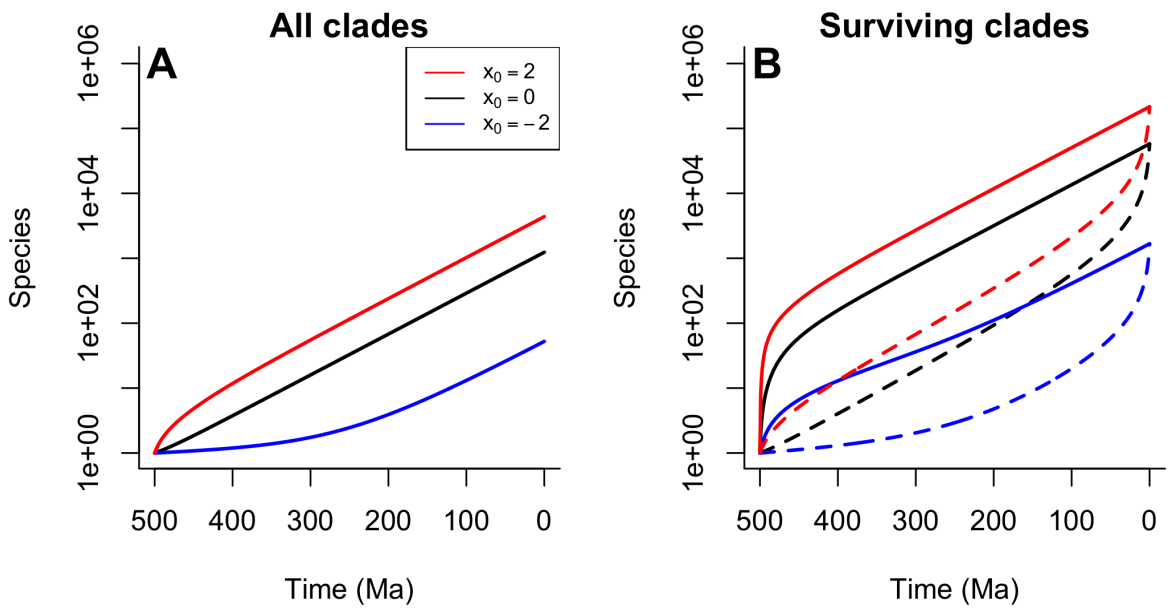

FIGURE 2. Diversification through time as a function of starting tempo. (A) The expected number of species through time for $x_0 = -2$ for a clade starting 500 Ma with different initial log-tempos: $x_0 = -2$ (blue line); $x_0 = 0$ (black line); $x_0 = 2$ (red line). These expectations include clades that are extinct. Clades with a higher starting tempo initially diversify more quickly (on average); eventually diversification stabilises to a fixed rate independent of the starting tempo. (B) Expected diversification profiles for clades that survive to the present day. Solid lines indicate the expected number of species through time; dashed lines indicate the expected number of lineages – species with surviving descendants. Surviving clades of all starting tempos experience the Push of the Past, mirrored by the Pull of the Present in the lineages. This effect is especially pronounced in the clades starting with the highest tempo.

decay as new species emerge whose own tempos diffuse away from the initial state. The duration of these transient effects is longer in clades that start with low tempos, since all processes including those that control the diffusion of tempos over time run slower. Although the effect of initial tempo is transient, it leaves an important signature in the eventual size of clades over the long term: because initially high tempo clades diversify more quickly in their early history, they reach a larger size before reverting to a constant diversification rate, meaning that they have a much greater expected diversity in the present. This intuitively suggests that the largest clades of a given age in the present are likely to be those that originated from a high-tempo common ancestor.

Distribution of tempos over time

As a clade diversifies, the various taxa will develop different tempos as they diverge independently from the initial starting tempo, leading to a time-dependent distribution of log-tempos $p(t, x)$. In the Appendix 1 we show that the evolution of this distribution obeys a replicator-mutation equation:

$$\frac{\partial p}{\partial t} = rp\left(e^x - \langle e^x \rangle\right) + \theta \frac{\partial x e^x p}{\partial x} + s^2 \theta \frac{\partial^2 e^x p}{\partial x^2} \tag{5}$$

where the term $\langle e^x \rangle = \int_{-\infty}^{\infty} e^x p(t, x) dx$ indicates the average value of e^x at a given time.

We numerically integrated this equation through times $0 < t < 500$ myrs for three initial starting log-tempos: $x_0 \in \{-2, 0, 2\}$ specified by initial conditions of the form $p(t = 0, x) = \delta(x - x_0)$, where $\delta(\cdot)$ is the Dirac delta function (Shutovskyi, 2023). The resulting evolution of the log-tempo probability distributions is shown in Figure 3. These results show that regardless of the starting tempo of the process, our model converges over time to the same stable distribution of log-tempos that is approximately normally distributed. Using the core set of model parameters described earlier gives a mean log-tempo of zero. When the process is initiated with a high tempo ($x = 2$) the convergence to this stable distribution is very rapid (red line). This is because the initially high tempo forces all processes to run fast, so time is effectively compressed. Conversely when the process is initiated with a slow tempo $x = -2$, the convergence is much slower, potentially taking hundreds of millions of years. In practical terms, this predicts the existence of long-lived substructures of the evolutionary tree in which evolution is effectively 'running slow'. If other evolutionary processes such as molecular and morphological change are also covariant to the tempo this would imply the existence of lineages with low diversity and minimal morphological or molecular change over very long periods of time. Since such small clades are common (Figure 1A), we expect that these 'living fossils' will be ubiquitous, and in particular that they will often be the sister group to the few large clades that dominate total diversity.

Varying the parameters of our model produces changes in the stable distribution of

tempos. In particular the mean of this distribution increases with larger r and decreases with larger θ (Figure 3B); in the limiting case where $r = 0$, the mean log-tempo can be shown to converge to -1 in closed form (see Appendix 1, equation 40). The dynamics of diversification tend to elevate the mean tempo, since higher tempo lineages produce more descendants on average per unit time, which inherit the same high tempo from their parents nodes. An interesting corollary to this point is that without any sort of mean reversion process, tempos (and thus diversification rates) would simply tend to rapidly increase without limit. As this is not observed empirically, the suggestion must be that something tends to draw log-tempos towards a characteristic mean value (c.f. Maliet et al. (2019); Aris-Brosou and Yang (2003); Lepage et al. (2006)). In Appendix 1 we show that such a mean reversion can arise without implying any necessary ecological mechanism: if tempo is encoded genetically then intermediate tempos are consistent with a greater number of possible genetic configurations, such that random mutations tend to cause a drift towards these values.

Patterns of historical tempo

So far we have considered what happens to various features of the evolutionary process as it is run forward from a particular initial condition. However, evolutionary analysis can be considered to be retrospective as well: one attempts to identify and explain patterns of evolution looking back in time from a vantage point in the present. As discussed by Budd and Mann (2018) this perspective necessarily distorts the patterns we are likely to observe, especially if one also chooses to analyse clades that have unusual modern-day properties. Such choices are commonplace: the most studied clades are often unusually diverse relative to clades of similar age; since most species are contained in these large clades they are often taken to be particularly representative of a particular epoch, despite in fact being a highly unrepresentative sample of clades in general.

To investigate the role that contingencies of clade selection have on the observed

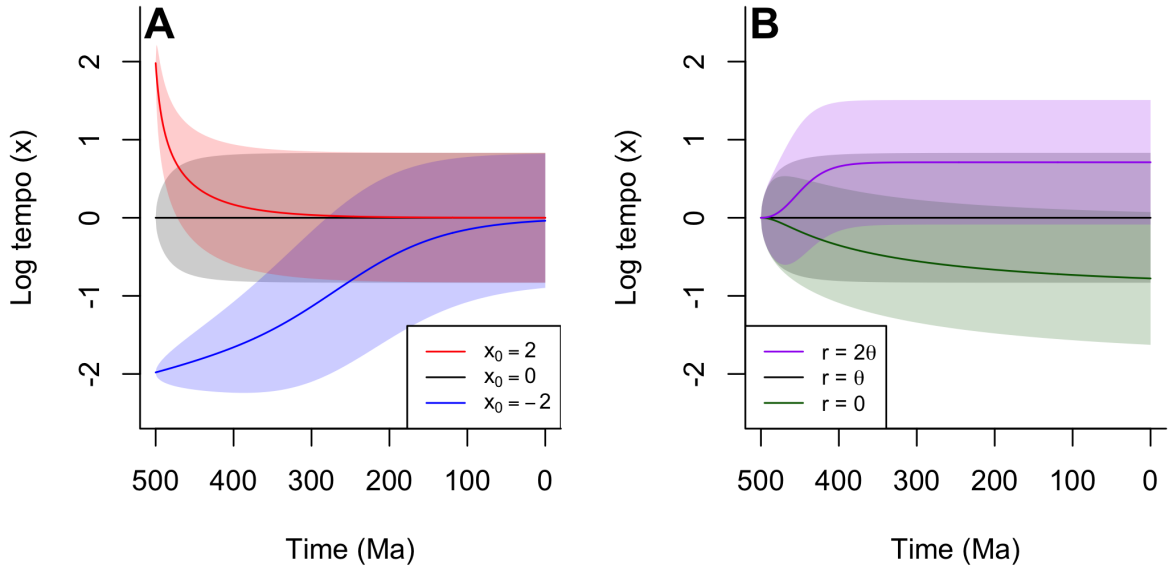

FIGURE 3. (A) The evolution of the distribution of log-tempos through time for clades starting from different initial log-tempos: $x_0 = -2$ (blue line); $x_0 = 0$ (black line); $x_0 = 2$ (red line). Lines indicate the expected log-tempo of a randomly chosen species, and shaded areas represent the standard deviation. Regardless of starting tempo, clades converge to the same equilibrium distribution of log-tempos. This convergence is fast in clades that start with high tempos. (B) Evolution of the log-tempo distribution for clades with different values of the diversification parameter r, with a fixed value of $\theta = 0.01$/myrs. Starting from the same tempo ($x_0 = 0$), clades reach different equilibrium log-tempo distributions depending on the value of r; higher values of r produce higher average tempos.

patterns of evolution, we considered two questions. First, if one randomly selects a species in the present and traces its lineage back in time, what expectations should we have about the evolutionary tempo of those ancestors? Second, what expectations should we have about the average tempo of earlier members of that clade overall? These are different questions since most historical taxa, even those with modern descendants (the lineages) will contribute little to modern diversity, owing to the Single Big Jump Principle (Vezzani et al., 2019) identified earlier (cf. Figure 1B). That is, the ancestors of most modern taxa constitute a very small subset of historical diversity.

First we consider how likely it is that a species alive today at time T originated from an ancestor at time t with log-tempo x. Since we assumed that the clade originates with an ancestor drawn from the equilibrium distribution $p(x)$, the prior probability that a species alive at time t has log-tempo x remains $p(x)$ by definition of the equilibrium. We can

determine a posterior estimate for the ancestor's log-tempo by application of Bayes rule:

$$\begin{aligned} &p(x \mid \text{ancestor of random modern taxon}) \\ &\quad = \frac{p(\text{ancestor of random modern taxon} \mid x)p(x)}{p(\text{ancestor of random modern taxon})} \end{aligned} \tag{6}$$

The likelihood term in this equation, $p(\text{ancestor of random modern taxon} \mid x)$ is proportional to the expected number of modern species that an ancestor at time t will generate, $N_x(T-)$. This means we can rewrite the above as:

$$p(x \mid \text{ancestor of random modern taxon}) = \frac{p(x)N_x(T-t)}{\int_{-\infty}^{\infty} p(x')N_{x'}(T-t)dx'} \tag{7}$$

The equation above estimates the log-tempo of a direct ancestor of a modern taxon. We can also ask what the tempo of a randomly chosen member of the clade in the past is. To estimate this we consider the probability of generating n_T species at time T from any starting log-tempo (based on solution of the generating function G_x) and the probability that a randomly chosen species at time t has log-tempo x_t if the process starts at x_0, $p(x_t \mid x_0)$. From these probabilities we can infer the probability of a historical log-tempo x_t conditioned on the current diversity n_T, using Bayes formula and marginalising over the unknown starting log-tempo x_0:

$$\begin{aligned} p(x_t \mid n_T) &= \int_{-\infty}^{\infty} p(x_t \mid x_0, n_T)p(x_0|n_T)dx_0 \\ &= \frac{\int_{-\infty}^{\infty} p(x_t \mid x_0)P(n_T \mid x_0, x_t)p(x_0)dx_0}{P(n_T)} \\ &\simeq \frac{\int_{-\infty}^{\infty} p(x_t \mid x_0)P(n_T \mid x_0)p(x_0)dx_0}{P(n_T)} \end{aligned} \tag{8}$$

where the final approximation assumes that $P(n_T \mid x_0, x_t) \simeq P(n_T \mid x_0)$. In general this approximation will be reasonable, because of the earlier result that modern diversity arises from a small subset of historical taxa. If the historical number of species at time t is high, a randomly chosen taxon is unlikely to contribute significantly to modern diversity and we can therefore treat n_T as being independent of this species and its tempo. Because of the Push of the Past (Budd and Mann, 2018), surviving clades will rapidly reach this state, and in the special case where $t = 0$ (i.e. the origin of the clade) the approximation holds

exactly.

Figure 4 illustrates our expectations about the historical patterns of tempo. Figure 4A shows the distribution of log-tempos for ancestors of a randomly chosen modern taxon, conditioned on our standard set of parameters ($\lambda = 0.51$ per species per myrs, $\mu = 0.5$ per species per myrs, $\theta = 0.01$/myrs, $s = 1$). In the present these are centered around $x = 0$, which is the stable overall distribution of log-tempos shown in Figure 3A. As we look backwards in time the expected log-tempo of the ancestor rises sharply, before plateauing at $x \simeq 0.6$ at c. 100 Ma. While the uncertainty represented by the standard deviation in grey permits a wide variety of ancestral tempos, beyond 100 Ma these ancestors will have elevated tempos with very high probability. Conversely, the tempo of the clade as a whole tends to peak at its origin, as shown in Figure 4B. This illustrates the overall expected log-tempo of historical species within a clade inhabited by a typical modern taxon (i.e. one with a diversity equal to the mean experienced clade size). That is, the clades that contain most modern taxa are defined by a high early rate of evolution, which then undergoes a consistent secular decline to the present, while the direct ancestors of most modern taxa have uniformly elevated rates of evolution across the history of the clade until close to the present. A consequence of this result is that most modern taxa share relatively recent common ancestors (c. 100-150Ma), as they overwhelmingly tend to originate via a small subset of lineages that maintain high tempos until this point. This is despite the most recent common ancestor of *all* species being close to the origin of the clade (in other words; the crown group is expected to emerge soon after the total group — for analysis see for example see Budd and Mann (2018)).

Effect of tempo variation on branch lengths and duration

We have now considered the effect of tempo variation on the dynamics of the birth-death process, and by extension on diversification. We motivated our approach by noting that rates of molecular evolution are commonly assumed to vary in modern relaxed

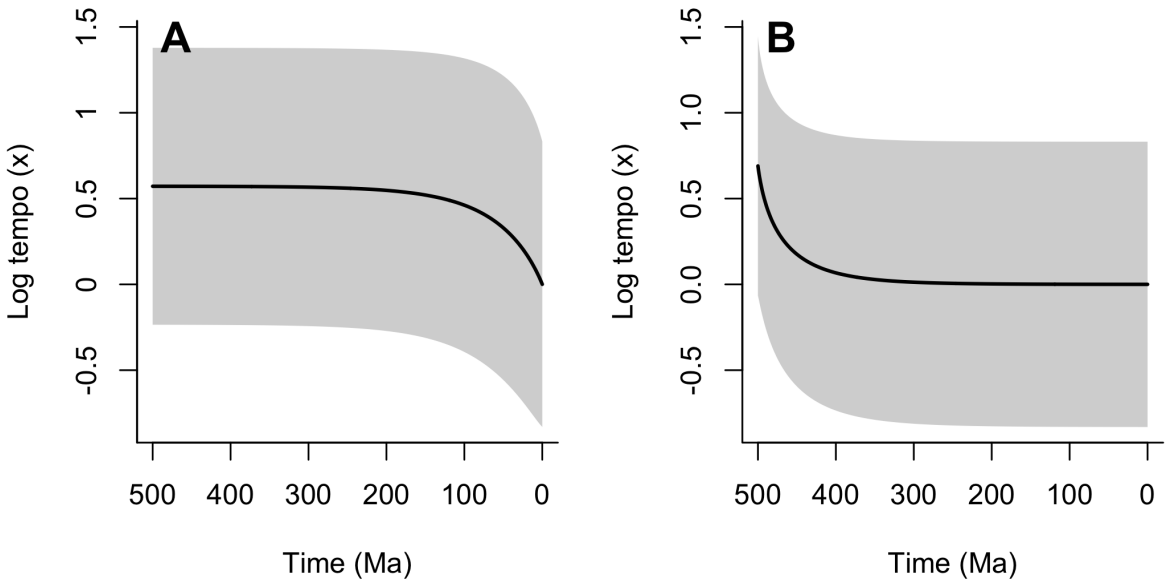

FIGURE 4. Expected patterns of historical tempo evolution. Lines indicate the expected log-tempo and shaded areas represent the standard deviation. (A) The expected historical log-tempo of ancestors of a randomly-chosen modern taxon, following its lineage back to the origin of the clade. Throughout this lineage, expected log-tempos are elevated relative to the present day, declining rapidly shortly before the present. (B) Expected historical log-tempo of species in the clade as a whole. This is shown for a clade of the mean experienced clade size (the typical clade size of a randomly chosen modern species). Expected tempos are highest at the origin of the clade and decline through time as the clade diversifies. In both panels, the distribution of tempos at time = 0 Ma represents the equilibrium distribution derived as the stable solution to equation 36

.

molecular clock analyses, and now we turn our attention to the interaction of molecular evolution and diversification. Specifically, we consider the expected duration (in real time, equivalent to branch height) and amount of molecular change along branches (= branch length) with differing initial log-tempo values. In our model, tempo can vary within a branch, so the duration of branches is not necessarily exponentially distributed, in contrast to standard BDP models. Instead, the probability that a branch terminates (either by speciation or extinction) in a small interval of time Δt depends on its current log-tempo and is given by $e^x(\lambda + \mu)\Delta t$.

As shown in Appendix 1, this implies that the probability density $f_x(t)$ that a branch originating with log-tempo x terminates at time t obeys a partial differential

equation of the form:

$$\frac{\partial f_x}{\partial t} = e^x \left(-(\lambda + \mu) f_x - \theta x \frac{\partial f_x}{\partial x} + \theta s^2 \frac{\partial^2 f_x}{\partial x^2} \right), \tag{9}$$

with initial condition $f_x(t = 0) = -e^x(\lambda + \mu)$. Figure 5A shows the solution to this equation for three different values of $x \in \{-2, 0, 2\}$, illustrating the intuitive result that branches with lower initial tempos tend to have a greater duration – that is they exist for a longer time before either speciating or going extinct.

How does this effect of the initial tempo translate into the amount of molecular change that occurs within a branch? This is an important question, because the relationship between branch duration and molecular change is fundamental to the practice of molecular dating and potentially more broadly to the inference of phylogenetic relationships based on the molecular genetic data from modern taxa because of the problems caused by long branch attraction (Kapli et al., 2021; Shafir et al., 2020).

If we assume that rates of molecular change co-vary with tempo alongside all other rates then the amount of molecular change Δw that occurs in some small unit of time Δt is given by:

$$\Delta w = e^x \Delta t \Rightarrow \frac{dw}{dt} = e^x \tag{10}$$

Applying a change of variables to express Equation 9 in terms of the molecular change w gives an equation obeyed by the probability density of molecular change $f_x(w)$ in a branch that starts with log-tempo x:

$$\frac{\partial f_x}{\partial w} = \left(-(\lambda + \mu) f_x - \theta x \frac{\partial f_x}{\partial x} + \theta s^2 \frac{\partial^2 f_x}{\partial x^2} \right), \tag{11}$$

with initial condition $f_x(w = 0) = \lambda + \mu$. Noticing that the partial derivatives in this equation will remain zero for all values of w, this simplifies to a standard exponential distribution:

$$f_x(w) = (\lambda + \mu) e^{-(\lambda+\mu)w}. \tag{12}$$

That is, the amount of molecular change contained in a branch is *independent* of the value of the tempo. This is illustrated in Figure 5B.

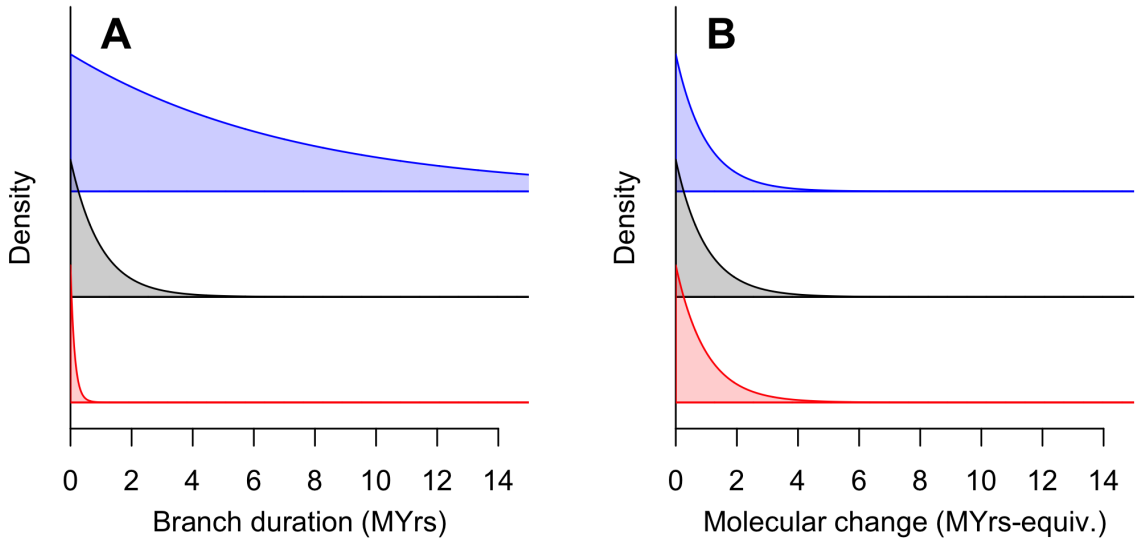

FIGURE 5. The distribution of branch durations (A) and amounts of molecular change along branches (B) for branches starting with different log-tempos: $x_0 = -2$ (blue); $x_0 = 0$ (black); $x_0 = 2$ (red), assuming that molecular evolution is covariant with tempo. Branches that start with lower tempos are much longer on average in real time than those with high tempos. However, the expected amount of molecular change is independent of the starting tempo.

The key result then is that branches that start at higher tempos are typically shorter, but contain just as much molecular change, as longer branches that originate from lower tempos. This implies that a clade that starts with a high tempo is likely to be characterised in its early stages by short-duration branches that nonetheless contain just as much molecular change as later branches that are longer in duration. Since we have shown above that early high tempos are expected especially in clades that are particularly large, we can expect this pattern to be commonly observed. As a corollary, if we further assume that morphological change also co-varies with tempo (c.f. Omland (1997); Lee et al. (2013)) then the same pattern of rapid change along short early branches would be observed morphologically by an analogous argument.

Discussion

We have described the CET model of macroevolution that allows the rates of speciation, extinction and molecular/morphological evolution to co-evolve through a variable evolutionary tempo parameter. This model provides a resolution to several outstanding difficulties in reconciling classical birth death models with empirical data. Allowing for tempo variation produces much greater variation in clade sizes over a given time horizon than under homogeneous models, consistent with the fact that modern diversity is dominated by a relatively small number of very large clades across different taxonomic levels. An underappreciated consequence of this distribution is that if we wish to understand how modern patterns of diversity arose, it is important to study the characteristic behaviour of such large clades, which, as we have shown here, differs markedly from that of clades as a whole. In other words, large and arguably charismatic clades such as arthropods, birds and angiosperms that are the subject of understandable interest have quite different patterns of evolution than what an 'average' clade might be inferred to have.

Our anaylsis predicts that these clades containing the bulk of modern diversity are likely to result from very high early evolutionary tempos, leading to short early branches (measured in real time). Because we conjecture that evolutionary tempo affects all rates in a covariant fashion, these short early branches are nonetheless expected to contain as much molecular and morphological change as later, longer branches, because the rates of molecular and morphological change are elevated in direct proportion to speciation and extinction.

This offers an explanation for the observation of, for example, such elevated rates coupled in short early branches found in molecular studies that take the fossil record as a reliable guide to the age of the clade (e.g. Lee et al. (2013)). In that example, early rates seem to be approximately 10 times higher than later ones, which would give a $\log(x)$ value of 2.3, in a clade that is at least 20 times larger than average. Our initial value of $\log(x)$ of

c. 0.7 for a clade c. 8 times larger than average in Figure 4B seems to be broadly compatible with this. We note that such studies tend to indicate a much older origin of the clade when the firm calibration based on the fossil record is removed; this emerges because of the use of a model that assumes a homogeneous birth death process as the underlying description of diversification (Budd and Mann, 2024), and because of a questionable assumption that the processes of diversification and molecular evolution are independent. Modern molecular clock analyses typically employ a 'relaxed-clock' methodology that permits substantial changes in the rate of molecular evolution across time and between lineages, but these rates are decoupled from the rates of speciation, extinction and lineage creation (e.g. Aris-Brosou and Yang (2003)). Such a rigorous decoupling between evolutionary processes seems intuitively unrealistic, and indeed elevated rates of molecular evolution have been posited as a *cause* of radiations (Lancaster, 2010), while in the fossil record morphological change is (necessarily) the key signature of diversification. As such, we argue that recognising the likely covariance between these rates is key to understanding apparent discrepancies between molecular signatures of diversification and the fossil record. Nevertheless, our model does not rely on any particular causal relationship between molecular change and diversification, and indeed these variables may be linked by underlying factors such as body size (Berv and Field, 2018).

A covariant process that extends to rates of molecular evolution will produce similar amounts of molecular change on all branches of the tree, regardless of their duration in time. This suggests that from a molecular standpoint there will be little or no difference between an older tree whose branch rates exhibit no secular trend, and a younger tree that experiences rapid early evolution and diversification followed by a slowdown (or indeed an even older tree that experienced very slow early evolution, although these will typically represent only a small proportion of modern diversity. As such, molecular data from modern taxa are unlikely to be able to discern which of these scenarios led to the molecular and species diversity we observe today. Precise and reliable

fossil calibrations, in combination with molecular data can potentially reveal the typical distribution of rates within the time scope of those calibrations. However, extrapolation of younger rates into deeper time is problematic, as we have shown that these are likely to be higher in the past, beyond the deepest precise calibrations (c.f. Budd and Mann (2020b)). This imposes a currently-insurmountable barrier to the use of the molecular clock for providing reliable clade age estimates, unless one can argue that rates of speciation and extinction are substantially decoupled from the process of molecular change. As noted earlier, making such an argument would preclude many putative explanations for observed rapid radiations, as well as being counter-intuitive. Although we have analysed a model in which there is a perfect correlation between all evolutionary rates, in practice we expect that any significant coupling will severely hamper the use of current clock methodologies. We suggest therefore that the use of molecular clocks for making extrapolative deep-time age estimates is fundamentally unreliable (interpolations within a tree, between nodes of known age are likely to be more constrained, but here we expect that molecular data will add little to dates derived directly from fossils (e.g. Brown and Smith (2018)).

As well as revealing the broad outlines of the dynamics of a varying tempo model of evolution, our analysis of this model also provides several empirical predictions:

1. Analysis of clades which are known to originate at similar times will show that the large majority of modern diversity is contained in a small subset of these clades. Most concretely, we anticipate that in pairs of sister groups, one group is likely to greatly dominate the diversity of the total (cf. Aldous (2001)).

2. The smaller sister group in a clade will be that which also experiences lower aggregate molecular and morphological change over its history. As such, the species in this group will tend to retain more plesiomorphic features relative to those in the larger sister group. Potential examples of such a phenomenon include the onychophorans relative to arthropods, cyclostomes relative to gnathostomes (Yu et al., 2024), or priapulids relative to other ecdysozoans (e.g. Webster et al. (2006)).

This prediction gives some succour to the popular notion of 'living fossil' that are slow-evolving, have few species, and which to some extent resemble ancestral taxa (c.f. Crisp and Cook (2005) for the traditional view that 'basal', species-poor groups should not be regarded as ancestral or 'primitive'; and Jenner (2022) for a more general discussion of the issue).

3. The direct ancestors of most modern species will show elevated rates of evolution (diversification, molecular and morphological) throughout their history. Those lineages that gave rise to a majority of modern species will therefore show consistent rates of molecular evolution until close to the present, when they fall. However, if one analyses all historical taxa in a large clade (which is where most modern taxa reside) we expect to see very high rates of molecular change concentrated at the origin of the clade, declining consistently to the present. Nevertheless, both of these expected patterns take place within a wider context in which rates of evolution remain consistent *overall* – that is, measured over all species in all clades at a given time.

4. If we further assume that rates of evolution are associated with body size and generation time (e.g. high rates being linked to small bodies and short generation times), we expect that a randomly chosen modern species will have experienced an increase in body size and generation time in the recent past, having probably originated from ancestors with smaller body size and shorter generation time (c.f. Berv and Field (2018)).

Each of these predictions already enjoys some degree of empirical support in the existing literature, as indicated above. However, further research is needed to test each systematically to the extent that these predictions could be judged to be successful or falsified.

In conclusion, our analysis suggests that a strong correlation between rates of molecular evolution and diversification would explain several empirical features of the natural world, unify two key areas of statistical modelling within a common framework,

and point towards necessary developments in phylogenetic inference and molecular dating in which this link is made explicit, such as an extension of the CET model to permit direct inference of actual historical rates from molecular data.

Online Supplementary Material

The R code for generating the figures is available from the Dryad Repository at `https://doi.org/10.5061/dryad.q573n5ts2`

Conflict of Interest

None declared.

Acknowledgements

This work was supported by Vetenskapsrådet (VR) grant 2022-03522 and UK Research and Innovation Future Leaders Fellowship MR/S032525/1 & MR/X036863/1. We are grateful to Tobias Uller, Ivan Prates, Dan Rabosky, Ben Slater and Jonathan Ward for discussions and help with literature for various aspects of this work. Hélène Morlon and her group kindly made comments on a previous draft of the ms. We are particularly grateful to Antonio Segalini, who was generous with his time and expertise in discussing and coding the numerical solution of PDEs, and the careful and constructive comments of David Černý and two other reviewers.

Appendix 1

Generating functions

We will make extensive use of probability generating functions. A quick review of their important properties follows. A probability generating function, $G(z)$ for the random

variable X is defined as:

$$G(z) = \sum_k P(X = k) z^k \tag{13}$$

The probability generating function has several important properties that will be useful in the subsequent exposition. In particular:

1. Normalisation: $G(z = 1) = \sum_k P(X = k) = 1$ (in cases where $P(X = k)$ represents a full probability distribution)
2. Extinction probability: $G(z = 0) = P(X = 0)$
3. Expectation: $M(X) = \sum_k kP(X = k) = \frac{\partial G}{\partial z}|_{z=1}$
4. Sum of random variables: If $W = X + Y$, then $G_W(z) = G_X(z)G_Y(z)$
5. Retrieval of probabilities: $P(X = k) = \frac{1}{k!} \frac{d^k G(z)}{dz^k} \Big|_{z=0}$

In respect of point (5) above, the values of $P(X = k)$ can be retrieved efficiently by Fourier inversion:

$$\begin{aligned} P(X = k) &= \frac{1}{k!} \frac{d^k G(z)}{dz^k} \Big|_{z=0} \\ &= \frac{1}{2} \int_{-\pi}^{\pi} G(\exp(i\theta)) \exp(-ik\theta) d\theta \end{aligned} \tag{14}$$

Where the integral expression makes use of the Cauchy integral formula. This expression can be efficiently solved numerically using Fast Fourier Transform methods (Gleeson et al., 2014)

Derivation of equation specifying evolution of the generating function

Define $G_x(t, z) = \sum_n P_n(t, x) z^n$ as the generating function for the number of species alive at time t from a process that starts at log-tempo x at time $t = 0$. We indicate the x dependence by means of a subscript for reasons of notational clarity in later analysis.

Assume that we know the generating function for all x at some time t. How will the generating function change over a small increment of time Δt? Since the process is

fundamentally homogeneous in time (i.e., there are no 'special' times'), we can construct this by considering a process that starts incrementally earlier than the known generating function. Within this small interval of time the process will change log-tempo incrementally according to an OU process, and furthermore may either speciate (producing two new independent processes with identical starting tempos) or go extinct. Given a current tempo x, the probability of speciation is $e^x \Delta t \lambda$, and that of extinction is $e^x \Delta t \mu$. Based on these possible events, the new generating function is given by a mixture of generating functions at time t:

$$\begin{aligned} G_x(t + \Delta t, z) &= G_x(t, z) \\ &+ e^x \Delta t \int_{-\infty}^{\infty} \left(\lambda G_{x'}(t, z)^2 + \mu - (\lambda + \mu) G_{x'}(t, z) \right) p(x' \mid x) dx'. \end{aligned} \tag{15}$$

Here $p(x' \mid x)$ specifies the probability for the tempo to transition from x to x' over the time interval Δt. We take x to evolve via an OU process, with autocorrelation parameter θ and a stationary variance s^2, experiencing an effective time $e^x \Delta t$ within real time Δt. Given this specification we have:

$$x' \mid x \sim \mathcal{N}(x - \theta x e^x \Delta t, 2 s^2 \theta e^x \Delta t) \tag{16}$$

which yields: $\mathbb{E}[x' - x] = -\theta x e^x \Delta t$ and $\mathbb{E}[(x' - x)^2] = 2\theta s^2 e^x \Delta_t$ up to first order terms in Δt.

Taking a 2nd-order Taylor expansion of $G_{x'}(t, z))$ around x and retaining first-order terms in Δt gives:

$$\begin{aligned} G_{x'} - G_x \simeq e^x \Delta t \left((\lambda G_x - \mu)(G_x - 1) \right) \\ + \mathbb{E}[x' - x] \frac{\partial G_x}{\partial x} + \frac{1}{2} \mathbb{E}[(x' - x)^2] \frac{\partial^2 G_x}{\partial x^2}. \end{aligned} \tag{17}$$

Where we have dropped the explicit dependence of G_x on arguments t and z for concision. Substituting the above expressions for $\mathbb{E}[x' - x]$ and $\mathbb{E}[(x' - x)^2]$ and taking the limit as $\Delta t \to 0$ gives the fundamental PDE of diversity evolution as given in equation 2.

$$\frac{\partial G_x}{\partial t} = e^x \left((\lambda G_x - \mu)(G_x - 1) - \theta x \frac{\partial G_x}{\partial x} + \theta s^2 \frac{\partial^2 G_x}{\partial x^2} \right) \tag{18}$$

Initial and boundary conditions

The most obvious question one can ask of this equation is: what is the probability that a process starting at log-tempo x will generate n species over time t? To answer this question we must solve equation 2 for different values of z, and use the Fourier inversion formula to retrieve the probability distribution $P_n(x,t)$. Solving equation 2 requires both initial and boundary conditions. For the question posed above the appropriate initial condition is given by $G(t=0,x,z) = z \forall x$, since a process that does not evolve for any time must have one species. Choosing appropriate boundary conditions is more difficult; since we must solve equation 2 numerically we take 'no flow' boundary conditions ($\frac{\partial G}{\partial x} = 0$) at some finite bounds $x_{\min}$ and $x_{\max}$ (we will usually use $-10 < x < 10$).

We can also ask how many species of log-tempo y will be produced at time t by a process that starts with log-tempo x at time $t = 0$. Define the generating function of this distribution by $G_x^y(t,x,z)$. Some consideration will show that the time evolution of G_x^y obeys the same PDE as that of G_x, but with a different initial condition. Since a process that starts with log-tempo x cannot instantaneously evolve to one of $y \neq x$, we use the initial condition: $G_x^y(t=0,x,z) = \delta(x-y)z$, where $\delta(\cdot)$ is the Dirac delta function.

Evolution of the mean diversity

The mean of a distribution is straightforwardly recovered from its generating function via the relationship $\mathbb{E}(n) = \sum_n nP_n = \frac{\partial G}{\partial z}|_{z=1}$. Applying this to the equation derived above for the evolution of the generating function gives the evolution of the mean diversity for a process that starts with log-tempo x. Defining $N_x(t) \equiv \mathbb{E}(n \mid x,t)$ as the expected value of n at time t for a process starting with log-tempo x:

$$\begin{aligned} \frac{\partial N_x}{\partial t} &= \left.\frac{\partial^2 G_x}{\partial t \partial z}\right|_{z=1} \\ &= e^x \left(2\lambda G_x|_{z=1} N_x - (\lambda+\mu)N_x - \theta x \frac{\partial N_x}{\partial x} + s^2 \theta \frac{\partial^2 N_x}{\partial x^2} \right) \end{aligned} \tag{19}$$

Since $G|_{z=1} = 1 \forall\ x, t$ by definition, we can simplify this to the expression given in equation 4:

$$\frac{\partial N_x}{\partial t} = e^x \left(r N_x - \theta x \frac{\partial N_x}{\partial x} + s^2 \theta \frac{\partial^2 N_x}{\partial x^2} \right) \tag{20}$$

where $r = \lambda - \mu$.

By using initial conditions $N_x(t = 0) = 1 \forall\ x$, solving this equation gives the mean number of species generated by a process starting at time $t = 0$ and log-tempo x. As with the discussion of initial conditions above, we can also apply the same equation with different initial conditions to consider how many species with specific log-tempo y are generated by a process that starts at log-tempo x. Denoting the expected number of such species of this type as $N_x^y(t)$, in this case we use the initial condition $N_x^y(t = 0) = \delta(x - y)$, analogously to the case of solving for the generating function. By definition, the expected number of species in total will be the sum over all final log-tempos: $N_x(t) = \int_{-\infty}^{\infty} N_x^y(t) dy$. Furthermore, we can ask what the expected number of species with log-tempo y is at time t if the starting log-tempo is unknown but specified by a probability distribution $p(x)$. In this case we have:

$$N^y = \int_{-\infty}^{\infty} N_x^y p(x) dx \tag{21}$$

and the expected total number of species (considering all possible starting and current log-tempos) can be denoted simply as $N(t)$ and is given by:

$$N = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} N_x^y p(x) dx dy \tag{22}$$

Conditioning on survival

Equation 4 describes the evolution of the mean number of species through time, including all cases where the process goes extinct before the current time. If we want to ask how many species will be alive at time t, assuming that the process hasn't gone extinct, we can do so straightforwardly by excluding the extinct cases:

$$\mathbb{E}(n_t \mid n_t > 0) = \frac{\mathbb{E}(n_t)}{P(n_t > 0)} = \frac{N_x(t)}{S_x(t)} \tag{23}$$

where $S_x(t) = 1 - G_x(t, z = 0)$ is the survival probability for a process starting at log-tempo x, determined from solving equation 2 for $z = 0$. However, we may also want to know the expected number of species at some time t, conditioned on knowing that the process will survive to some future time T. In this case the conditioning is more complex. We make use of the identity:

$$P(n_t) = P(n_t, n_T > 0) + P(n_t, n_T = 0), \tag{24}$$

which leads to:

$$\begin{aligned} \sum_{n_t} n_t P(n_t \mid n_T = 0) &= \frac{\sum_{n_t} n_t P(n_t) - \sum_{n_t'} n_t P(n_t, n_T = 0)}{P(n_T > 0)} \\ \Rightarrow \mathbb{E}(n_t \mid n_T > 0) &= \frac{N_x(t) - C_x(t)}{S_x(T)}, \end{aligned} \tag{25}$$

where $C_x(t)$ is a correction term depending on x and t that we need to determine. Define a new generating function $H_x(t, z) = \sum_{n_t} P(n_t, n_T = 0) z^{n_t}$. Differentiating H_x with respect to z and evaluating at $z = 1$ gives the required correction term in the equation above. As with the generating function G, the evolution of H is governed by equation 2:

$$\frac{\partial H_x}{\partial t} = e^x \left((\lambda H_x - \mu)(H_x - 1) - \theta x \frac{\partial H_x}{\partial x} + \theta s^2 \frac{\partial^2 H_x}{\partial x^2} \right) \tag{26}$$

Differentiating with respect to z gives:

$$\frac{\partial C_x}{\partial t} = e^x \left((\lambda (2 H_x|_{z=1} - 1) - \mu) C_x - x\theta \frac{\partial C_x}{\partial x} + s^2 \theta \frac{\partial^2 C_x}{\partial x^2} \right). \tag{27}$$

Unlike in the case for G_x, $H_x|_{z=1}$ varies as a function of x and t, and so solution of this equation for C_x requires simultaneously solving this PDE and 2 with initial conditions: $H_x(t, z) = z G_x(T - t, z = 0)$ and $C_x(t) = G_x(T - t, z = 0)$.

Lineages

Lineages are species in the past that have descendants in the present. Since molecular studies are based on extant species, any phylogeny reconstructed from these must consist of lineages. The evolution of lineages has thus been dubbed the ‘reconstructed

process' (Nee et al., 1994b), since these constitute the phylogeny that can, in principle, be reconstructed from molecular or morphological analysis of modern taxa.

We are interested in the number of species alive at time t which will have descendants at some later time T. Recall $N_x^y(t)$ is the expected number of species of log-tempo y at time t in a process that starts at log-tempo x. The expected number of these that will have descendants at time T is $S_y(T-t)$ (the survival probability over time $T-t$ for a new process starting with log-tempo y). Thus the expected number of lineages of log-tempo y at time t is $S_y(T-t)N_x^y(t)$. Summing over values of y gives the total expected number of lineages, $M_x(t)$ at time t for a process starting with log-tempo x, viewed from the perspective of time T (we leave this dependence on the time of observation implicit in the notation, but note that lineages are only defined from the perspective of a specific point in time):

$$M_x(t) = \int_{-\infty}^{\infty} S_y(T-t)N_x^y(t,x)dy \tag{28}$$

This expectation includes the cases where the number of lineages is zero, i.e where there are no species at time T. If we wish to condition on the process surviving to the present we must remove these cases by dividing by $P(n_T > 0) = S_x(T)$

$$M_x(t) \mid [N(T,x) > 0] = \frac{\int_{-\infty}^{\infty} S_y(T-t)N^y(t,x)dy}{S_x(T)} \tag{29}$$

Evolution of tempo distribution

Assuming that we start a process with log-tempo x, over time species generated by that process will diverge in tempos. How does this distribution of tempos evolve?

Consider starting a process with log-tempo x, and then selecting a species at random at some time t. The probability that this species has log-tempo y is given by:

$$p(y \mid x,t) = \frac{N_x^y(t)}{\int_{-\infty}^{\infty} N_x^{y'}(t)dy'} \tag{30}$$

If the starting log-tempo is unknown, but drawn from a distribution $p(x)$, then we can

marginalise the above equation with respect to x to find the later distribution $p(y \mid t)$:

$$p(y \mid t) = \frac{\int_{-\infty}^{\infty} N_x^y(t,x)p(x)dx}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N_x^{y'}(t)p(x)dxdy'} \tag{31}$$

Taking the derivative with respect to time gives:

$$\frac{\partial p(y \mid t)}{\partial t} = p(y \mid t)\left(f(y) - \int_{-\infty}^{\infty} p(y' \mid t)f(y')dy'\right) \tag{32}$$

where $f(y) = \frac{1}{N^y}\frac{\partial N^y}{\partial t}$. That is, the distribution of log-tempos evolves according to a replicator equation, where the 'fitness' of a log-tempo y is given by the proportional increase in $N^y = \int_{-\infty}^{\infty} N_x^y p(x)dx$.

If we assume that at some point in time the distribution of log-tempos is given by $p(x)$, we can consider the instantaneous evolution of N^y from this time. Defining the current time to be $t = 0$, we have the initial condition:

$$N_x^y(t=0) = \delta(x-y) \tag{33}$$

From equation 4, this implies that:

$$\frac{\partial N_x^y}{\partial t}|_{t=0} = e^x\left(r\delta(x-y) - x\theta\frac{\partial\delta(x-y)}{\partial x} + s^2\theta\frac{\partial^2\delta(x-y)}{\partial x^2}\right) \tag{34}$$

Applying standard rules for the operation of derivatives of the Dirac delta function, we can marginalise the above equation with respect to the initial distribution $p(x)$ to give:

$$\begin{aligned}\frac{\partial N^y}{\partial t}|_{t=0} &= \int_{-\infty}^{\infty}\frac{\partial N_x^y}{\partial t}p(x)dx \\ &= re^y p(y) + \theta\frac{\partial e^y y p(y)}{\partial y} + s^2\theta\frac{\partial^2 e^y p(y)}{\partial y^2}\end{aligned} \tag{35}$$

Substituting this into equation 32, and noting that again that $N^y = \int_{-\infty}^{\infty} N_x^y p(x)dx$, we get:

$$\frac{\partial p(y)}{\partial t} = rp(y)\left(e^y - \langle e^y\rangle\right) + \theta\frac{\partial e^y y p(y)}{\partial y} + s^2\theta\frac{\partial^2 e^y p(y)}{\partial y^2} \tag{36}$$

Where $\langle e^y\rangle = -\int_{-\infty}^{\infty} e^{y'}p(y')dy'$ is the mean value of e^y.

This then provides a replicator-mutation equation for the evolution of the tempo distribution, with the 'fitness' of log-tempo y being re^y. In particular, it specifies that the

stable long term distribution of log-tempos is given by the solution to:

$$rp(y)\left(e^y - \langle e^y \rangle\right) + \theta \frac{\partial e^y y p(y)}{\partial y} + s^2 \theta \frac{\partial^2 e^y p(y)}{\partial y^2} = 0 \tag{37}$$

Notably, we can see that if $r = 0$, we recover the standard Fokker-Planck representation for the stationary OU process in the transformed distribution $e^y p(y)$:

$$\frac{\partial e^y y p(y)}{\partial y} + s^2 \frac{\partial^2 e^y p(y)}{\partial y^2} = 0 \tag{38}$$

with the stationary solution $p(y) = \frac{\exp(-s/2)}{\sqrt{2\pi s^2}} e^{-y} \exp\left(\frac{-y^2}{2s^2}\right)$, implying a mean log-tempo of $\langle y \rangle = -1$.

From the equilibrium equation we can also find another useful relationship on the mean value. Multiplying equation 37 by y and integrating gives:

$$\int_{-\infty}^{\infty} y \left[rp(y)\left(e^y - \langle e^y \rangle\right) + \theta \frac{\partial e^y y p(y)}{\partial y} + s^2 \theta \frac{\partial^2 e^y p(y)}{\partial y^2} \right] dy = 0 \tag{39}$$

Integrating the partial differential terms by parts yields:

$$(r - \theta)\langle y e^y \rangle - \langle y \rangle \langle e^y \rangle = 0 \tag{40}$$

from which we can see that if $r = \theta$ then $\langle y \rangle = 0$, i.e. the mean log-tempo will converge to zero when the diversification parameter is equal to the mean-reversion parameter.

Branch duration and expected molecular change

Considering a branch that begins with log-tempo x, what is the expected time until that branch terminates, either by speciation or extinction? For a branch to endure for time $t + \Delta t$ it must first fail to terminate in time Δt, and then survive for a further time t with some new log-tempo x'. Integrating over the possible values of x' we have:

$$1 - F_x(t + \Delta t) = (1 - e^x(\lambda + \mu)\Delta t) \int_{-\infty}^{\infty} (1 - F_{x'}(t)) p(x'|x) dx', \tag{41}$$

where $F_x(t)$ is the cumulative probability that the branch originating with log-tempo x has terminated by time t.

Taking a second-order Taylor expansion of $F_{x'}(t)$ around $x' = x$ and retaining first order terms in Δt we have:

$$\frac{\partial F_x}{\partial t} = e^x \left((\lambda + \mu)(1 - F_x) - \theta x \frac{\partial F_x}{\partial x} + \theta s^2 \frac{\partial^2 F_x}{\partial x^2} \right) \tag{42}$$

The probability density for the branch to terminate at time t is given by differentiation of the cumulative distribution: $f_x(t) = \frac{\partial F_x}{\partial t}$. Applying this transformation to the equation above yields:

$$\frac{\partial f_x}{\partial t} = e^x \left(-(\lambda + \mu) f_x - \theta x \frac{\partial f_x}{\partial x} + \theta s^2 \frac{\partial^2 f_x}{\partial x^2} \right) \tag{43}$$

The probability density for a branch to terminate at time t thus follows the same form of differential equation as that for the mean number of species (equation 4), but with $-(\lambda + \mu)$ taking the place of r. Solving this equation requires the initial condition $F_x(t = 0) = 0 \ \forall x$, which implies $f_x(t = 0) = e^x(\lambda + \mu)$.

Assuming that molecular rates of change are covariant to tempo (e^x), for every increment of time Δt the expected amount of molecular change Δw (in arbitrary units that we label as myrs-equivalent; 1 myrs-equivalent being the expected molecular change in 1 myrs at a fixed tempo of $\tau = 1$) is $\Delta w = e^x \Delta t$. We can transform the above equation for $F_x(t)$ (which is given in terms of real time t) into one that applies over w via a change of variables, to give the cumulative probability $F_x(w)$ that a branch terminates before accumulating w units of molecular change.

$$\frac{\partial F_x}{\partial w} = \left((\lambda + \mu)(1 - F_x) - \theta x \frac{\partial F_x}{\partial x} + \theta s^2 \frac{\partial^2 F_x}{\partial x^2} \right), \tag{44}$$

and as above we obtain the probability density to terminate at w, $f_x(w)$, by differentiation: $f_x(w) = \frac{\partial F}{\partial w}$:

$$\frac{\partial f_x}{\partial w} = \left(-(\lambda + \mu) f_x - \theta x \frac{\partial f_x}{\partial x} + \theta s^2 \frac{\partial^2 f_x}{\partial x^2} \right) \tag{45}$$

Here we have the initial condition $F_x(w = 0) = 0 \ \forall x$, which implies $f_x(w = 0) = \lambda + \mu$. Consideration of this equation will show that the partial derivatives in x are initially zero and will remain zero for all values of w. Thus we can simplify the equation to:

$$\frac{\partial f_x}{\partial w} = -(\lambda + \mu) f_x. \tag{46}$$

The solution to this equation is straightforward and shows that w follows an exponential distribution with rate $\lambda + \mu$:

$$f_x(w) = (\lambda + \mu) \exp(-(\lambda + \mu)w) \tag{47}$$

The notable feature of this density is that it does not depend on the starting log-tempo x, implying that the amount of molecular change in a branch is independent of tempo.

Schematic for genetic encoding of tempo

Here we describe a simple model for how a genetic encoding of tempo can lead to the modified OU process we take as the basis for tempo evolution. Consider a binary string of n bases represented as '1' or '0', and define ρ as the proportion of bases that are 'active' – that is, encoded as '1'. We assume that these bases mutate independently and neutrally, and with a rate that is covariant to the tempo e^x, such that the probability for each base to mutate in a small interval of time Δt is $qe^x\Delta t$.

If the number of active bases at time t is given by $n\rho_t$, then in the interval of time Δt the number of bases that mutate from '1' to '0' is binomially distributed as $B(qe^x\Delta t, n\rho_t)$, and similarly the number mutating from '0' to '1' is binomially distributed as $B(qe^x\Delta t, n(1-\rho_t))$. If we take n to be large and Δt to be small these binomial distributions can be approximated by normal distributions, such that the number of mutations from '1' to '0' is normally distributed with mean $qe^x\Delta t n\rho_t$ and variance $qe^x\Delta t(1 - qe^x\Delta t)n\rho_t$, and the number of mutations from '0' to '1' is normally distributed with mean $qe^x\Delta t n(1-\rho_t)$ and variance $qe^x\Delta t(1 - qe^x\Delta t)n(1-\rho_t)$.

The change in the number of active bases is given by the number mutating from '0' to '1', minus the number mutating from '1' to '0'. Given the results above, this change is also normally distributed. Taking the limit as Δt becomes infinitesimal (denoted dt) and retaining only terms first order in dt we have:

$$d\rho \sim \mathcal{N}(-e^x q(2\rho - 1)dt, qe^x dt/n) \tag{48}$$

This is equivalent to the following form of stochastic differential equation:

$$d\rho = -e^x q(2\rho - 1)dt + \sqrt{e^x(q/n)}dW \tag{49}$$

where dW is an increment from a standard Wiener process, with mean zero and variance dt.

To specify a genetic encoding of the log-tempo, let us now define $x = \alpha(2\rho - 1)$, where α is some arbitrary constant of proportionality, such that when half of bases are active this defines $x = 0$. We can then rewrite the above equation as:

$$dx = -2e^x qxdt + \sqrt{4\alpha^2 e^x(q/n)}dW \tag{50}$$

Defining new variables $\theta = 2q$ and $s^2 = \alpha^2/n$, we have:

$$dx = -e^x \theta x dt + \sqrt{e^x 2\theta s^2}dW \tag{51}$$

which is precisely the modified OU process specified in equation 1. By taking α to be sufficiently large we can extend the boundaries of minimum and maximum values of x such that arbitrarily high or low values of x are possible within this model. We have assumed that n is large, and this assumption means that boundary effects around $\rho = 1$ and $\rho = 0$ can be safely ignored as these states are highly unlikely to occur under a random mutation process.

This then provides a schematic representation of how tempo could be genetically encoded in a manner that naturally leads to the modified OU process description that we employ in this paper. The purpose of this schematic is not to argue that this represents the actual genetic encoding of tempo in any specific details, but instead to illustrate how such an encoding would naturally give rise to the mean-reversion properties of the OU process, via the action of entropic forces. That is, the log-tempo tends to revert to the mean not due to any ecological mechanism, but simply because there are more possible encodings with $x \simeq 0$ than those that encode more extreme values of x. One way in which tempo might influence rates in the way required by the CET model would be if it was encoded by

a multilocus set of genes that influence body size, as body size appears to be associated with a syndrome of other features such as generation time and mutation rate (Martin, 2017). This encoding would satisfy the requirements of the CET model, although we would stress again that we have no formal commitment to it.

References

Aldous, D. J. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today. Statistical Science 16:23–34.

Arenas, M. 2015. Trends in substitution models of molecular evolution. Frontiers in Genetics 6:319.

Aris-Brosou, S. and Z. Yang. 2003. Bayesian models of episodic evolution support a late Precambrian explosive diversification of the Metazoa. Molecular Biology and Evolution 20:1947–1954.

Barido-Sottani, J. and H. Morlon. 2023. The ClaDS rate-heterogeneous birth–death prior for full phylogenetic inference in BEAST2. Systematic Biology 72:1180–1187.

Barraclough, T. G. and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. Evolution 55:677–683.

Beaulieu, J. M., B. C. O'Meara, P. Crane, and M. J. Donoghue. 2015. Heterogeneous rates of molecular evolution and diversification could explain the Triassic age estimate for angiosperms. Systematic Biology 64:869–878.

Beck, R. M. and M. S. Lee. 2014. Ancient dates or accelerated rates? morphological clocks and the antiquity of placental mammals. Proceedings of the Royal Society B: Biological Sciences 281:20141278.

Benton, M. J. and B. C. Emerson. 2007. How did life become so diverse? the dynamics of diversification according to the fossil record and molecular phylogenetics. Palaeontology 50:23–40.

Berv, J. S. and D. J. Field. 2018. Genomic signature of an avian Lilliput effect across the K-Pg extinction. Systematic Biology 67:1–13.

Blum, M. G. and O. François. 2006. Which random processes describe the tree of life? A large-scale study of phylogenetic tree imbalance. Systematic Biology 55:685–691.

Bouckaert, R., T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. De Maio, et al. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology 15:e1006650.

Bromham, L. 2003. Molecular clocks and explosive radiations. Journal of Molecular Evolution 57:S13–S20.

Bromham, L. 2009. Why do species vary in their rate of molecular evolution? Biology letters 5:401–404.

Bromham, L. 2020. Causes of variation in the rate of molecular evolution. The Molecular Evolutionary Clock: Theory and Practice Pages 45–64.

Bromham, L. 2024. Combining Molecular, Macroevolutionary, and Macroecological Perspectives on the Generation of Diversity. Cold Spring Harbor Perspectives in Biology Page a041453.

Bromham, L. and M. Woolfit. 2004. Explosive radiations and the reliability of molecular clocks: island endemic radiations as a test case. Systematic Biology 53:758–766.

Bromham, L. D. and M. D. Hendy. 2000. Can fast early rates reconcile molecular dates with the Cambrian explosion? Proceedings of the Royal Society of London. Series B: Biological Sciences 267:1041–1047.

Brown, J. W. and S. A. Smith. 2018. The past sure is tense: on interpreting phylogenetic divergence time estimates. Systematic Biology 67:340–353.

Budd, G. E. and R. P. Mann. 2018. History is written by the victors: the effect of the push of the past on the fossil record. Evolution 72:2276–2291.

Budd, G. E. and R. P. Mann. 2020a. The dynamics of stem and crown groups. Science Advances 6:eaaz1626.

Budd, G. E. and R. P. Mann. 2020b. Survival and selection biases in early animal evolution and a source of systematic overestimation in molecular clocks. Interface Focus 10:20190110.

Budd, G. E. and R. P. Mann. 2024. Two notorious nodes: a critical examination of relaxed molecular clock age estimates of the bilaterian animals and placental mammals. Systematic Biology 73:223–234.

Coiro, M., J. A. Doyle, and J. Hilton. 2019. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? New Phytologist 223:83–99.

Cooney, C. R. and G. H. Thomas. 2021. Heterogeneous relationships between rates of speciation and body size evolution across vertebrate clades. Nature Ecology & Evolution 5:101–110.

Crisp, M. D. and L. G. Cook. 2005. Do early branching lineages signify ancestral traits? Trends in Ecology & Evolution 20:122–128.

Dos Reis, M., P. C. Donoghue, and Z. Yang. 2016. Bayesian molecular clock dating of species divergences in the genomics era. Nature Reviews Genetics 17:71–80.

Duchêne, D. A., X. Hua, and L. Bromham. 2017. Phylogenetic estimates of diversification rate are affected by molecular rate variation. Journal of Evolutionary Biology 30:1884–1897.

Eo, S. H. and J. A. DeWoody. 2010. Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proceedings of the Royal Society B: Biological Sciences 277:3587–3592.

Erwin, D. H. 2000. Macroevolution is more than repeated rounds of microevolution. Evolution & Development 2:78–84.

Forbes, A. A., R. K. Bagley, M. A. Beer, A. C. Hippee, and H. A. Widmayer. 2018. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. BMC Ecology 18:1–11.

Gleeson, J. P., J. A. Ward, K. P. O'Sullivan, and W. T. Lee. 2014. Competition-induced criticality in a model of meme popularity. Physical Review Letters 112:048701.

Goldie, X., R. Lanfear, and L. Bromham. 2011. Diversification and the rate of molecular evolution: no evidence of a link in mammals. BMC Evolutionary Biology 11:286.

Greenberg, D. A. and A. Ø. Mooers. 2017. Linking speciation to extinction: diversification raises contemporary extinction risk in amphibians. Evolution Letters 1:40–48.

Henao-Diaz, L. F. and M. Pennell. 2023. The major features of macroevolution. Systematic Biology 72:1188–1198.

Holmes, J. D. and G. E. Budd. 2022. Reassessing a cryptic history of early trilobite evolution. Communications Biology 5:1177.

Hua, X. and L. Bromham. 2017. Darwinism for the genomic age: connecting mutation to diversification. Frontiers in Genetics 8:12.

Hugall, A. F. and M. S. Lee. 2007. The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61:2293–2307.

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

943 Jenner, R. A. 2022. Ancestors in evolutionary biology: linear thinking about branching
944 trees. Cambridge University Press.

945 Jobson, R. W. and V. A. Albert. 2002. Molecular rates parallel diversification contrasts
946 between carnivorous plant sister lineages 1. Cladistics 18:127–136.

947 Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. Mammalian Protein
948 Metabolism 3:21–132.

949 Kapli, P., T. Flouri, and M. J. Telford. 2021. Systematic errors in phylogenetic trees.
950 Current Biology 31:R59–R64.

951 Kendall, D. G. 1948. On the generalized" birth-and-death" process. The Annals of
952 Mathematical Statistics 19:1–15.

953 Khurana, M. P., N. Scheidwasser-Clow, M. J. Penn, S. Bhatt, and D. A. Duchêne. 2024.
954 The limits of the constant-rate birth–death prior for phylogenetic tree topology
955 inference. Systematic Biology 73:235–246.

956 Lancaster, L. T. 2010. Molecular evolutionary rates predict both extinction and speciation
957 in temperate angiosperm lineages. BMC Evolutionary Biology 10:1–10.

958 Lanfear, R., S. Y. Ho, D. Love, and L. Bromham. 2010. Mutation rate is linked to
959 diversification in birds. Proceedings of the National Academy of Sciences
960 107:20423–20428.

961 Lee, M. S., J. Soubrier, and G. D. Edgecombe. 2013. Rates of phenotypic and genomic
962 evolution during the Cambrian explosion. Current Biology 23:1889–1895.

963 Lepage, T., S. Lawi, P. Tupper, and D. Bryant. 2006. Continuous and tractable models for
964 the variation of evolutionary rates. Mathematical Biosciences 199:216–233.

965 Magallon, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm
966 clades. Evolution 55:1762–1780.

967 Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating
968 species-specific diversification rates. Nature Ecology & Evolution 3:1086–1092.

969 Marshall, C. R. 2017. Five palaeobiological laws needed to understand the evolution of the
970 living biota. Nature Ecology & Evolution 1:0165.

971 Martin, R. A. 2017. Body size in (mostly) mammals: mass, speciation rates and the
972 translation of gamma to alpha diversity on evolutionary timescales. Historical Biology
973 29:576–593.

974 McPeek, M. A. and J. M. Brown. 2007. Clade age and not diversification rate explains
975 species richness among animal taxa. The American Naturalist 169:E97–E106.

976 Moen, D. and H. Morlon. 2014. Why does diversification slow down? Trends in Ecology &
977 Evolution 29:190–197.

978 Nee, S. 2006. Birth-death models in macroevolution. Annual Review of Ecology, Evolution
979 and Systematics 37:1–17.

980 Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be
981 estimated from molecular phylogenies. Philosophical Transactions: Biological Sciences
982 344:77–82.

983 Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process.
984 Philosophical Transactions of the Royal Society of London B: Biological Sciences
985 344:305–311.

986 Omland, K. E. 1997. Correlated rates of molecular and morphological evolution. Evolution
987 51:1381–1393.

988 Quintero, I., N. Lartillot, and H. Morlon. 2024. Imbalanced speciation pulses sustain the
989 radiation of mammals. Science 384:1007–1012.

990 Rabosky, D. L. 2010. Primary controls on species richness in higher taxa. Systematic
991 Biology 59:634–645.

992 Rabosky, D. L., M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang,
993 and J. G. Larson. 2014. Bamm tools: an r package for the analysis of evolutionary
994 dynamics on phylogenetic trees. Methods in Ecology and Evolution 5:701–707.

995 Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E.
996 Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the
997 largest vertebrate radiation. Nature Communications 4:1958.

998 Ritchie, A. M., X. Hua, and L. Bromham. 2022a. Diversification rate is associated with
999 rate of molecular evolution in ray-finned fish (Actinopterygii). Journal of Molecular
1000 Evolution 90:200–214.

1001 Ritchie, A. M., X. Hua, and L. Bromham. 2022b. Investigating the reliability of molecular
1002 estimates of evolutionary time when substitution rates and speciation rates vary. BMC
1003 Ecology and Evolution 22:1–19.

1004 Rolland, J., L. F. Henao-Diaz, M. Doebeli, R. Germain, L. J. Harmon, L. L. Knowles,
1005 L. H. Liow, J. E. Mank, A. Machac, S. P. Otto, et al. 2023. Conceptual and empirical
1006 bridges between micro- and macroevolution. Nature Ecology & Evolution 7:1181–1193.

1007 Sarver, B. A., M. W. Pennell, J. W. Brown, S. Keeble, K. M. Hardwick, J. Sullivan, and
1008 L. J. Harmon. 2019. The choice of tree prior and molecular clock does not substantially
1009 affect phylogenetic inferences of diversification rates. PeerJ 7:e6334.

1010 Shafir, A., D. Azouri, E. E. Goldberg, and I. Mayrose. 2020. Heterogeneity in the rate of
1011 molecular sequence evolution substantially impacts the accuracy of detecting shifts in
1012 diversification rates. Evolution 74:1620–1639.

1013 Shutovskyi, A. M. 2023. Some applied aspects of the Dirac delta function. Journal of
1014 Mathematical Sciences 276:685–694.

1015 Smith, S. A. and J. M. Beaulieu. 2024. Ad fontes: divergence-time estimation and the age
1016 of angiosperms. New Phytologist 244:760–766.

Soltis, P. S. and D. E. Soltis. 2016. Ancient WGD events as drivers of key innovations in angiosperms. Current Opinion in Plant Biology 30:159–165.

Stadler, T., J. H. Degnan, and N. A. Rosenberg. 2016. Does gene tree discordance explain the mismatch between macroevolutionary models and empirical patterns of tree shape and branching times? Systematic Biology 65:628–639.

Stanley, S. M. 1990. The general correlation between rate of speciation and rate of extinction: Fortuitous causal linkages. Pages 103–127 *in* Causes of Evolution: A Paleontological Perspective (R. M. Ross and W. D. Allmon, eds.). University of Chicago Press, Chicago.

Stanley, S. M. 1998. Macroevolution: pattern and process. Johns Hopkins University Press, Baltimore MD.

Strathmann, R. R. and M. Slatkin. 1983. The improbability of animal phyla with few species. Paleobiology 9:97–106.

Venditti, C. and M. Pagel. 2010. Speciation as an active force in promoting genetic evolution. Trends in Ecology & Evolution 25:14–20.

Vezzani, A., E. Barkai, and R. Burioni. 2019. Single-big-jump principle in physical modeling. Physical Review E 100:012108.

Warnock, R. and A. Wright. 2021. Understanding the Tripartite Approach to Bayesian Divergence Time Estimation. Elements of Paleontology Cambridge University Press.

Webster, A. J., R. J. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478–478.

Webster, B. L., R. R. Copley, R. A. Jenner, J. A. Mackenzie-Dodds, S. J. Bourlat, O. Rota-Stabelli, D. Littlewood, and M. J. Telford. 2006. Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral ecdysozoan. Evolution & Development 8:502–510.

Xiang, Q.-Y., W. H. Zhang, R. E. Ricklefs, H. Qian, Z. D. Chen, J. Wen, and J. H. Li. 2004. Regional differences in rates of plant speciation and molecular evolution: a comparison between eastern Asia and eastern North America. Evolution 58:2175–2184.

Yu, D., Y. Ren, M. Uesaka, A. J. Beavan, M. Muffato, J. Shen, Y. Li, I. Sato, W. Wan, J. W. Clark, et al. 2024. Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. Nature Ecology & Evolution 8:519–535.