Evolutionarily distinct “living fossils” require both lower speciation and lower extinction rates

D.J. Bennett, M.D. Sutton, S.T. Turvey

RRH: EVOLUTIONARY DISTINCTNESS AND THE LIVING FOSSIL

LRH: DOMINIC J. BENNETT, MARK D. SUTTON ET AL.

Abstract.– As a label for a distinct category of life, 'living fossil' is controversial. The term has multiple definitions and it is unclear whether the label can be genuinely used to delimit biodiversity. Even taking a purely phylogenetic perspective where a proxy for the living fossil is evolutionary distinctness (ED), an inconsistency arises: does it refer to 'dead-end' lineages doomed to extinction or 'panchronic' lineages that survive through multiple epochs? Recent tree-growth model studies indicate that speciation rates must have been unequally distributed between species in the past to produce the shape of the tree of life. Although an uneven distribution of speciation rates may create the possibility for a distinct group of living fossil lineages, such a grouping could only be considered genuine if extinction rates also show a consistent pattern, be it indicative of dead-end or panchronic lineages. To determine whether extinction rates also show an unequal distribution, we developed a tree-growth model where the probability of speciation and extinction is a function of a tip's ED. We simulated 1,000s of trees where the ED function for a tip is randomly and independently determined for speciation and extinction rates. We find that simulations where the most evolutionarily distinct tips have lower rates of speciation and extinction produce phylogenetic trees closest in shape to empirical trees. This implies that a distinct set of lineages with reduced rates of diversification, indicative of a panchronic definition, is required to create the shape of the tree of life.

Dominic J. Bennett and Mark D. Sutton. Department of Earth Sciences and Engineering, Imperial College London, London, UK.

Dominic J. Bennett and Samuel T. Turvey. Institute of Zoology, Zoological Society of London, London, UK.

E-mail: dominic.john.bennett@gmail.com.

Introduction

The concept of the living fossil has ignited both public and scientific interest since it was coined by Darwin (1859). In its most exciting interpretation a living fossil can provide a window into the past because groups that are commonly labelled as living fossils have remained unchanged (or nearly unchanged) for millions of years. For example, horseshoe crabs (Limulidae), ginkgos (*Ginkgo biloba*) and tuataras (*Sphenodon*) each ostensibly give an insight into life during the Ordovician, the Permian and the Triassic, respectively (Royer et al. 2003; Hay et al. 2008; Rudkin et al. 2008). For this reason it is a controversial concept, and some who interpret it to mean 'a group that has undergone no change' have argued for it to be abandoned (Casane and Laurenti 2013). This controversy originates in part from the inconsistency of definition. Definitions proposed have included: a species or lineage that persists with relatively little morphological change for 'extended periods' of time (Yoshida 2002); an evolutionary relict of a once large radiation (Nagalingum et al. 2011); or a species that was first discovered in the fossil record and only subsequently found to be still extant (Forey 1984) (= Lazarus taxon *sensu lato*). A diversity of definitions and thresholds does not necessarily imply that a concept is flawed; a similar diversity of definitions also exists for the opposing macroevolutionary phenomenon, the adaptive radiation (Schluter 2000), yet few doubt its reality. Definitions are impacted by methodological and material limitations, as well as subsequent scientific insights. More importantly for the living fossil is the uncertainty of whether any definition of 'living fossil' can provide meaningful delimitation of organisms. Even under a simple definition – 'a group that experiences low rates of speciation and extinction' – it is unclear whether or not living fossils simply represent the tail of a distribution of evolutionary performances (Schopf 1984). To date, most work on determining living-fossil status has focussed on single taxonomic groups and has been primarily descriptive (e.g. Eldredge and Stanley 1984). This approach, however, does not address the generality of the living fossil concept and its reality as an evolutionary phenomenon. Instead we test the concept from a phylogenetic tree perspective that allows us to use an existing metric (evolutionary distinctness; ED) that has been suggested as a quantitative measure of ‘living fossilness’ (Isaac et al. 2007; Cavin and Kemp 2011).

The uneven distribution of species among taxonomic groups – the ‘hollow curve distribution’ (HCD) – may intuitively point to evolutionary performances differing substantially between groups (Alfaro et al. 2009). Yule (1925) was the first to notice the HCD and related it to a property of phylogenetic tree shape, imbalance – the unequal number of descendants between sister groups. However, Yule demonstrated that this imbalance does not necessarily indicate unequal rates of diversification among species due to the phenomenon of preferential attachment: if speciation is random then clades with more species will speciate more and more because they increasingly represent a larger proportion of all species. Since Yule, however, the increase in the number and diversity of published modern-day phylogenetic trees has made it apparent that phylogenetic imbalance is found across all parts and scales of the tree of life at a level greater than expected for a random process (Blum and François 2006), indicating that evolutionary performance is not normally distributed (Mooers and Heard 1997). Indeed, various independent studies, which modelled the growth of phylogenetic trees through time using different methods, have all found that strong positive ‘age-dependent’ speciation is able to generate the level of imbalance observed in empirical phylogenetic trees (Purvis et al. 2011; Hagen et al. 2015; Rabosky and Goldberg 2015). In other words, for phylogenetic trees to have the level of observed imbalance, species-rich groups containing young species must tend to have substantially higher rates of speciation than older species-poor groups.

Palaeobiologists will be more familiar with logistic models where speciation rates drop as clades reach predefined carrying capacities (Sepkoski 1978, 1979, 1984; Raup 1985). Although many spin-off models have been developed, such as hierarchical models (Brayard et al. 2009), models that incorporate mass extinctions (Courtillot and Gaudemer 1996), and coupled logistic models (Miller and Sepkoski 1988; Roy 1996; Sepkoski et al. 2000), these all share the central feature that speciation rates drop as clades grow in size. Although these logistic models and tree-growth models have been developed and applied towards different ends – the former mathematically for the description of clade taxonomic richness through time while the latter for the simulation of tree-growth with birth and death parameters – they both find that speciation rates change dynamically. Age-dependent tree-growth models that place higher speciation rates on younger tips and logistic models that place higher speciation rates on clades that have fewer species – both of which are not mutually exclusive – produce better fits, respectively, to observed phylogenetic tree shape and fossil record occurrences than alternative models.

This prevalence of age-dependent speciation rate bias would suggest that the living fossil – as an evolutionarily distinct lineage that experiences reduced rates of evolutionary change – could be a real phenomenon. Biases in extinction rates across the tree of life, however, have little effect on phylogenetic imbalance (Hagen et al. 2015), making it difficult, using imbalance alone, to determine what past levels of extinctions rates were. Yet determining whether extinction as well as speciation is differentially distributed is important in understanding the nature and reality of living fossils. If extinction rates also show a positive age-dependent bias, low-diversity groups – which may be identified as living fossils – will experience reduced speciation and extinction rates, allowing them to persist for extended periods of time. We term this scenario the ‘Panchronic living fossil hypothesis’ (Pan). If extinction rates instead show a negative age-dependent bias, low-diversity groups will experience reduced speciation and increased extinction rates, in which case living fossil lineages are evolutionary relicts soon to go extinct. We term this scenario the ‘Dead-End living fossil hypothesis’ (DE). Alternatively, extinction rates may not show any age-dependent bias, in which case ‘living fossil’ as a label, when used generally to describe low-diversity groups, may be covering at least two distinct macroevolutionary phenomena – groups generated by both the Pan and DE phenomena, and potentially also by alternative phenomena – and would require splitting. We term this the ‘Hydra living fossil hypothesis’ (Hyd). Large-scale studies of the fossil record have consistently found extinction rates to be random with respect to lineage age for multiple groups (Van Valen 1973; Alroy 1996; Ezard et al. 2011), a signal that is even picked up during mass extinction events (Boyajian 1991). This would suggest that the Hyd scenario is the most likely, a finding that could discredit the use of ED as a proxy of living fossilness as it implies there is no consistent process for generating evolutionarily distinct species.

One means of determining whether extinction rates show the same age-dependent bias as speciation rates using tree shape is to look at the temporal distribution of internal nodes – a property of tree shape that we term ‘phylogenetic gravity’ after centre of gravity or centre of mass as calculated for clade richness through time (Gould et al. 1987) (Fig. 1.) Phylogenetic gravity is driven by temporal differences in births and deaths (Pybus and Harvey 2000). Low gravity or ‘regressive’ clades experience early bursts of diversification, causing most of their internal nodes in a phylogenetic tree to be distributed towards the root and the peak of their fossil occurrences to occur near their inception. High gravity or ‘progressive’ clades experience late bursts of diversification, causing most of their internal nodes to distribute towards the tips and their historical diversity to peak near the present. Additionally, however, the phylogenetic location of speciation and extinction events will also have an impact on gravity (Hagen et al. 2015). For example, when birth and death rates are the same between evolutionary scenarios, if speciation and extinction rates are opposing on equivalent parts of the tree (as would be the case in the DE scenario), we would expect gravity to be higher than if they were equal (such as in the Pan scenario) – because in the former scenario tips will be removed where there are few recently added tips, which increases the likelihood of loss of deep internal nodes. To date, however, most modellers of tree growth have focused on imbalance rather than gravity due to the relative scarcity of time-calibrated phylogenies (Purvis et al. 2011).

Commented [DJB1]: Insert Figure 1 here.

Here, we test how different models of evolution, each with different implications for the origin of living fossils, affect the shape of the tree of life. As there is no agreed-upon definition of the living fossil, our aim is to take a particular phylogenetic concept for what living fossils might represent, and test its self-consistency and utility for partitioning biodiversity. Because the majority of groups described as living fossils are often distinct lineages with few close relatives (Eldredge and Stanley 1984), we used the Evolutionary Distinctness (ED) metric – the number of closely related species measured by phylogenetic isolation – as a simple proxy for 'living fossilness'. For the living fossil concept to have general applicability, some or all of these low-diversity or evolutionarily distinct lineages must behave in similar ways over macroevolutionary time, as we would demonstrate if we discovered a strong signal for either a DE or Pan scenario. To this end, we developed a tree-growth model with age-dependent bias for speciation and extinction rates using a tip's ED value. An ED bias is equivalent to an age-dependent bias because time since last speciation correlates with a tip's ED value. With this model we simulated 10,000 trees under different scenarios by varying the bias in rates of speciation and extinction by the evolutionary distinctness of tips. To identify which hypothesis (DE, Pan, Hyd, or others) best recreates empirical evolutionary patterns, we compared simulated trees to over 2,000 empirical modern-day phylogenetic trees, representing sections of the tree of life, using both phylogenetic imbalance and a birth-death independent measure of phylogenetic gravity.

Materials and Methods

Evolutionary Distinctness Biased Markov Model (EDBMM).— We developed a tree-growth model based on the Equal Rates Markov Model (ERMM) that biases speciation and extinction rates by the ‘living-fossilness’ of a species – the Evolutionary Distinctness Biased Markov Model (EDBMM). The ERMM starts with a seed tree, and adds and drops tips at random per unit branch length given birth (*b*) and death (*d*) parameters. The EDBMM biases selection for speciation or extinction by a tip’s evolutionary distinctness value (ED). At every time step (*t*), the ED values of all the tips are calculated, and these are used to bias the random selection of tips for speciation and extinction. Tips are then added or dropped and the branch lengths of all extant tips are extended by $\boldsymbol{\frac{t}{n}}$, where n is the number of tips in the tree, so that birth and death rates are constant in time. A simulation can be stopped after a set time is elapsed or the tree has reached a set number of tips. (Fig. S1).

Because there is no fully agreed-upon definition of a living fossil, we use, as a proxy, ‘a low diversity lineage over time’; this is applicable to most lineages described as living fossils (Eldredge and Stanley 1984). This ‘living-fossilness’ proxy was calculated using ED as this is a measure of a species’ level of phylogenetic isolation, making it an estimate for species’ recent lineage diversification rate. We calculated ED using the metric of Isaac et al. (2007). This metric shares the total branch length of the tree between tips by equally apportioning branch lengths by the number of descendants. In this way it takes into account all branches and is conceptually intuitive (Fig. 2).

Commented [DJB2]: Insert Figure 2 (A and B) here.

Although speciation and extinction rates are held constant and determined by birth and death parameters, the relative probability of a tip being selected for speciation or extinction is determined by its ED value and two power parameters: σ (‘sigma’) and ε (‘epsilon’). The relative probability of the ith tip (*i*) speciating (*S*) at each time step (where *n* is the number of tips in the tree and *d* is their ED values) is given by equation (1), while its probability of going extinct (*E*) is given by equation (2). Given these probabilities and global birth (λ) and death (μ) rates, the probability of diversifying (*D*) for the ith tip in the next time step can be determined, equation (3).

$$p_i(S) = \frac{d_i^{\sigma}}{\sum_{j=1}^{n} d_j^{\sigma}} \qquad \text{eq. (1)}$$

$$p_i(E) = \frac{d_i^{\varepsilon}}{\sum_{j=1}^{n} d_j^{\varepsilon}} \qquad \text{eq. (2)}$$

$$p_i'(D) = \lambda \cdot p_i(S) - \mu \cdot p_i(E) \qquad \text{eq. (3)}$$

In other words, species with higher ED (hereafter ‘ED species’) experience increased rates of speciation and extinction for positive values of σ and ε respectively, and reduced rates of speciation and extinction for negative values. Values of 0 for σ and ε provide no overall bias towards ED species, making speciation and extinction rates random for all tips in the tree. (See Fig. 2 for visual representation of how different σ values influence probability of speciation for a 7-tipped tree.)

The σ-ε parameter space can be split into four quadrants, each with their own implications for the living fossil concept. We term each quadrant: ‘Phylogenetic Fuse’ (PF, +σ, -ε), ‘Dead-End’ (DE, -σ, +ε), ‘Ephemeral’ (Eph, +σ, +ε) and ‘Panchronic’ (Pan, -σ, -ε). Additionally, we identify five scenario points representing the extremes of each quadrant as well as the ‘Hydra’ (Hyd, σ=-1, ε=0) (see Figure 3A). In the PF scenario, ED species have low extinction rates and high speciation rates, such that species that could be described as living fossils are ‘fuses’ that will experience future evolutionary radiations. In the DE scenario, ED species have low speciation and high extinction rates; living fossils are evolutionary relicts doomed to disappear. In the Eph scenario, ED species have high speciation and extinction rates, making living fossils transient phenomena. In the Pan scenario, ED species experience low speciation and extinction rates, meaning that living fossils persist through time in low species numbers. In the Hyd scenario, ED species experience reduced speciation rates but extinction rates are random, making living fossils potentially the products of both Pan and DE scenarios.

Tree Simulations.— To determine which part of the σ-ε parameter space best recreates empirical tree shape, we simulated 10,000 trees using randomly generated σ and ε values between -1 to +1. We used seed trees of two tips, birth and death parameters of 2 and 1 respectively, and halted tree growth once they had reached a set number of tips determined by randomly selecting a value between 50-500 for each of the simulations. We also used the same parameters to generate 1,000 trees for each of the ‘extreme’ scenarios PF, DE, Eph, Pan, and the additional Hyd in order to compare the differences between these scenarios with larger numbers of trees.

In addition to exploring phylogenetic tree shape we also explored how idealised fossil records differed between Hyd, DE and Pan. We took the simulated phylogenies generated from the ‘extreme’ scenarios of Hyd, DE and Pan and their simulations were continued for twice the starting age of the initial tree using new birth and death parameters of 1 and 1 to remove the influence of a higher birth rate on clade shape. In this stage extinct tips were not removed. One hundred equally spaced time samples were then extracted from the tree. Each sample recorded the number of descendants for each clade excluding all tips that were either extinct or not yet extant before and after the time point. A clade was defined in the strict monophyletic sense as a node and all its descendants. This produced a species richness time series for every clade in the tree that accounts for both extant and extinct lineages, unlike other time series based on present-day phylogenetic trees (e.g. Pybus and Harvey 2000). We calculated various clade statistics from these idealised fossil records, including the centre of gravity/mass, which measures the relative time point when a clade reaches its peak (Gould et al. 1987), and centre of gyration, which measures variance around that peak. Statistics were calculated for all clades with total sizes of more than 100 descendants after the 25th time point to reduce the influence of edge effects.

Empirical Tree Search, Manipulation and Comparison.— Empirical trees were gathered for comparison with the simulated trees. We first identified large species-level phylogenies with over 1,000 taxa from the literature (see Table S1). These trees were supplemented by downloading published phylogenies from TreeBase (Donoghue and Sanderson 2002) using the R package treebase (Boettiger and Temple Lang 2012). TreeBase searches were constrained to species-level phylogenies and to trees with more than 50 taxa. Trees that were polytomous were rendered fully bifurcating using the function multi2di (Paradis et al. 2004) because, upon inspection, the majority of these polytomies were due to inference error as they did not separate known taxonomic groups. Because this function randomly splits polytomies, we generated a distribution of 100 trees to account for any bias. Trees with branch lengths that were not ultrametric (i.e. branch lengths not relative to time) were converted by rate-smoothing. To account for differences between rate-smoothing algorithms we ran three different methods: D8 (Britton et al. 2007), MPL (Britton et al. 2002) and PL (Sanderson 2002). To make comparisons with the simulated trees, we ensured empirical trees had the same number of tips by reducing any tree with over 500 tips by extracting non-overlapping clades of between 50-500 tips. In total, we sourced 2,061 empirical trees with a mean of 101 (25th, 50th and 75th quantiles: 60.0, 78.0, 110.5) tips (Fig. S2). The mean year of publication of the empirical trees was 2009 (SD 4). Of these trees, 1,432 (69%) were polytomous, 802 (39%) had given branch lengths, 64 (3%) were already ultrametric, and a mean 647 (31%), 492 (24%) and 370 (18%) were successfully made ultrametric using D8, PL and MPL respectively.

For all sourced empirical trees, the phylum, class and order represented by each tree were identified using the online Global Names Resolver taxonomic tool (Global Names Architecture 2015). The 215,752 unique names in all of the sourced trees represented 32 phyla, 87 classes and 260 orders. The number of trees represented by each phylum was highly unequal; the top five phyla -- Tracheophyta, Ascomycota, Chordata, Arthropoda and Basidiomycota -- represented 85% of all trees (Fig. S3). Sixty-one trees had names that could not be identified at the phylum, class or order rank.

For empirical and simulated trees we calculated ‘shape metrics’ for both imbalance and gravity. To avoid pseudoreplication, in cases where empirical tree data were available as distributions of trees rather than single trees – as produced from Bayesian analysis or by randomly removing polytomies – shape statistics were calculated for each tree in the distribution and the mean value was used in subsequent analysis.

To estimate imbalance we used both Colless (1982) and Sackin (1972) metrics. As imbalance metrics are affected by the number of tips in a tree, both measures were normalised with the expected value for a pure-birth tree of the same number of tips (‘Yule normalisation’) (Mooers and Heard 1997). Values greater than 0 indicate a tree that is more unbalanced than the Yule reference, and less than 0 indicate a tree that is more balanced. These metrics were calculated and normalised using the functions of the R package apTreeShape (Bortolussi et al. 2006).

The metric γ ("gamma") (Pybus and Harvey 2000) is used as a measure of gravity (termed "tippiness" by Manceau et al. 2015). However, γ was originally devised to measure stable (γ=0), increasing (γ>0) or decreasing (γ<0) lineage generation through time, and is therefore sensitive to differences in birth-death parameters between trees -- a factor we cannot control when sourcing empirical trees. Therefore, in addition to γ, we also calculated phylogenetic species variability (PSV) (Helmus et al. 2007) as a measure of phylogenetic gravity. As a measure of the relative distance between tips, PSV negatively covaries with gravity. Star trees where all tips are equally distant have low gravity and PSV values of 1. 'Dot' trees or 'line' trees where tips are separated from a single distant tip and internal nodes clustering have high gravity and PSV values tending towards 0 (Fig. S4). Although originally developed as a measure of community phylogenetic diversity, we show in supplementary analysis (see Figures S5-6) that PSV is less sensitive to the birth-death parameter (Pearson's Rs: -0.27 and 0.78, for PSV and γ) and more sensitive to ε (Pearson's Rs: 0.20 and -0.13, for PSV and γ) when birth-death varies randomly.

Normalisation was not required for gravity, as the metrics we use are independent of age and number of tips of the tree. Gravity metrics, however, require trees to be ultrametric, and as such only the ultrametric empirical trees (18-31% of sourced trees for the different rate-smoothing methods) were compared to simulated trees using gravity metrics. Because gravity and imbalance metrics partly correlate, we performed PCA using Sackin, Colless and PSV metrics to map the distances of empirical and simulated trees in in 'tree shape space'.

All of the methods described above were run in the R environment (v 3.2.3) and can be reproduced using online R scripts and functions (available: https://github.com/DomBennett/Project-EDBMM).

Results

Simulated Trees.— We found strong differences between the 10,000 simulated trees in each of the four quadrants of σ-ε parameter space. There was a strong negative correlation between σ and imbalance for both Colless and Sackin metrics (Pearson's Rs: -0.80 and -0.78 respectively; Figure S7). As such, the scenarios that represent positive σ in σ-ε space, PF and Eph, both produced trees more balanced than the Yule equivalent (Table 1). Because all empirical trees are less balanced than the equivalent Yule tree, we consider these scenarios to be implausible. Pan, Hyd and DE scenarios, however, represent negative σ in σ-ε space and produced trees more unbalanced than the equivalent Yule tree, consistent with empirical trees (Table 1).

Commented [DJB3]: Insert Table 1 here.

We found no significant differences between each of the four quadrants in terms of gravity, except for the DE scenario which produced trees with significantly lower γ values than all other simulated trees (37% lower, T-test: $p < 0.001$) and empirical trees (34% lower using D8, T-test: $p < 0.001$). Because in the DE scenario ED species have negative speciation and extinction rates, the most isolated branches of the tree are more likely to be pruned. This leads to the constant shrinking of the tree from the base, and causes the γ value to be low despite a birth-death parameter of 2-1. We term this the 'shrinking tree problem' (see Fig. 4). This explanation was corroborated by the average age of trees simulated in the DE quadrant being significantly lower than all other simulated trees (26% lower, T-test: $p < 0.001$).

Across all estimates of PSV, differences between each of the quadrants were not great, although Pan had the lowest PSV values, indicating higher gravity. In particular, PSV values were lowest for simulations with negative σ and mid-ranging or negative ε (Fig. S8). When excluding trees simulated with positive σ values, we found weak negative correlations between gravity and ε (Pearson's Rs: 0.18 (PSV) and -0.53 (γ); Fig. S9).

Commented [DJB4]: Insert Figure 3 (A, B and C) and Figure 4 here.

The clade dynamics show different shapes between the three extreme scenarios DE, Hyd and Pan. Although Pan produced trees with higher phylogenetic gravity than both DE and Hyd, the mean clade centre of gravity was lowest (Table 2). Pan also produced the most clades, which on average lasted less time than clades of other scenarios, indicating a high turnover. Furthermore, Pan had the highest gyration, indicating that although its clades tended to peak closer to the mid-point of their time span, this peak was more plateaued than in DE and Hyd. DE produced the fewest number of clades and had the highest gravity, indicating relatively fast clade extinction. With random extinction, clades produced with Hyd contained on average more species at peak (Table 2, Fig. 5).

Commented [DJB5]: Insert Table 2 and Figure 5 here.

Empirical trees.— Empirical trees were more unbalanced than the Yule equivalent for both Colless and Sackin metrics, 4.0 (SD 2.6) and 3.5 (SD 2.4) respectively. We did not find significant differences between γ generated with each of the three rate-smoothing methods (T-test$_{\mathrm{D8\sim PLl}}$: p = 0.062, T-test$_{\mathrm{D8\sim MPL}}$: p = 0.750; referred to later as γ_{D8}, γ_{PL} and γ_{MPL}). The mean γ for empirical trees was greater than expected for a simulated tree where birth and death parameters are equal, implying increasing lineage generation for most trees (Table 1). We also found no significant differences between PSV values generated from each of the three methods (T-test$_{\mathrm{D8\sim PL}}$: p = 0.59, T-test$_{\mathrm{D8\sim MPL}}$: p = 0.34; referred to later as PSV$_{\mathrm{D8}}$, PSV$_{\mathrm{PL}}$ and PSV$_{\mathrm{MPL}}$). The PSV values were mid-ranging, indicating a mid-level gravity for most trees (Table 1). Although we found significant correlation between gravity and imbalance (Pearson's product moment correlation, Colless~PSV$_{\mathrm{D8}}$: R = -0.21, p < 0.001), we deem this to be due to the physical impossibility of a tree with low gravity and high imbalance. We found significant differences between taxonomic groups at all three taxonomic ranks for both imbalance metrics but not for either of the gravity metrics (Table S2). The variance of shape metrics was larger between orders than between either classes or phyla, which may indicate that lower-ranked groups have diverging macroevolutionary dynamics compared to higher ranks (Fig. S10-12). For example, the variance of the means between taxonomic groups of Sackin and PSV$_{\mathrm{D8}}$ metrics increased 111% and 494%, respectively, from phyla to orders. However, we found that this increase in variance, or for any of the other metrics, was not significant (permutation tests of variances of the means from phyla to orders: p = 0.89 and p = 0.76 for Sackin and PSV$_{\mathrm{D8}}$ respectively) (Table S2).

Empirical and simulated compared.— In terms of imbalance, Pan and DE quadrants were both equally near to the empirical trees. However, for gravity as measured by PSV$_{D8}$, the Pan quadrant was closer to the empirical than DE (Table 1). To explore intra-quadrant differences, we plotted the distances of each 0.1x0.1 grid cell in σ-ε space to empirical trees for Colless, Sackin and PSV$_{D8}$ metrics (Fig. S13). Visually, this shows that simulated trees with extremely low values of σ and mid or negative levels of ε produce trees most similar to empirical trees in terms of imbalance and gravity. With a PCA of Sackin, Colless and PSV$_{D8}$ for the simulated and empirical trees, we found the grid cells closest to empirical trees mostly fell within the Pan quadrant (Fig. 3C).

This was further corroborated by a PCA of empirical trees and the simulated trees of each extreme scenario, which shows Pan to be closer to empirical trees in PC space than both of the other -σ scenarios Hyd and DE (Fig. 6). However, the level of variance in empirical tree shape was not matched by any of the simulated scenarios, particularly in gravity metrics (see Table 1). This higher variance in gravity could be the result of a greater diversity of birth-death rates; repeating the PCA above with empirical trees partitioned into $\gamma > 0$ and $\gamma < 0$ showed that empirical trees which are likely to have experienced reduced or negative diversification rates are closer in PC-space to DE and Hyd scenarios (Fig. S14). Additionally, we found a strong difference between the trees sourced from the literature, which had lower gravity than all simulated trees, than those originating from TreeBase (Fig. S15).

Commented [DJB6]: Insert Figure 6 here.

Discussion

Our phenomenological model of tree growth uses a tip's evolutionary distinctness value to bias speciation and extinction rates in order to test the reality of the living fossil. With our model, we confirm the results of previous studies where younger or more diverse clades have higher rates of net speciation (Purvis et al. 2011; Hagen et al. 2015; Rabosky and Goldberg 2015) by showing that scenarios where ED species have increased rates of speciation lead to trees with greater phylogenetic balance than in empirical trees. Scenarios where ED species have reduced rates of speciation (DE, Hyd and Pan) all produced levels of imbalance similar to empirical trees. Of these three scenarios, however, we found that the panchronic scenario (Pan), where speciation and extinction rates were reduced for ED species, produced trees closest in both gravity and imbalance to empirical trees. By eliminating the Hyd hypothesis we therefore show that the 'living fossil' label represents a true grouping and, by eliminating the DE hypothesis, that it does not merely represent 'evolutionary relicts' of once-large radiations that are soon to go extinct (Nagalingum et al. 2011). Instead, ED species are generally the product of the same macroevolutionary phenomenon and can therefore be grouped as a "real category" of biodiversity, and are able to persist without speciating or going extinct for long periods of time. Although our approach was only able to work with a purely phylogenetic interpretation of the living fossil based on ED, our results reveal this proxy more closely recalls the original living fossil concept proposed by Darwin (1859) as a "window into the past" than more recent concepts of living fossils as evolutionary dead-ends.

Our results differentiate between the fit of each of our three main hypotheses (Pan, Hyd and DE) to the shape of the tree of life in terms of phylogenetic gravity. While these results might theoretically be contingent on methodological particulars (metrics of gravity and rate-smoothing methods employed, and empirical tree sampling), we have good reasons to infer that this is not the case. Firstly, although PSV has never been used in this capacity before, it is still a useful measure of the shape of a tree, as it indicates whether the tree is more like a star or a line and is therefore equivalent to a metric of gravity. In supplementary analysis where we simulated trees with varying birth and death parameters, we demonstrated that PSV, although still strongly correlating with the birth parameter, also correlates more with ε than γ does. For this reason, we favoured PSV over γ for comparisons of gravity because the differences in γ between and among empirical trees were much greater than that produced by our simulations, obscuring any signal. Secondly, several rate-smoothing methods were employed, and all produced similar results. Additionally, we recovered the same tendency for empirical trees to have high levels of gravity ($\gamma > 0$ and PSV ~ 0.5) in trees that did not require rate-smoothing (i.e. were already ultrametric). We did, however, find significant differences between trees sourced from the literature and those from TreeBase. We were unable to break up the empirical trees based on the methods used to create them due to the lack of readily available data. Doing this, we may have found further sources of bias that can impact tree shape, and therefore concede that this may be a weakness in our estimation of mean gravity of empirical trees. Thirdly, the higher levels in gravity for empirical trees might reflect a bias in the sourced data towards groups in the process of radiating. We think it unlikely that this is the case, however, as we sampled from a broad range of taxonomic groups. Furthermore, our results indicate that the range of PSV values for radiating and non-radiating clades (i.e. the PSV values for those trees with positive and negative γ) are still greater than those produced by the Hyd or DE scenarios; higher births to deaths alone is unable to generate the higher gravity seen in empirical trees.

Although our results indicate that extinction rates show negative age dependence, there is little evidence for this from the fossil record. Most notably, Van Valen’s Red Queen hypothesis (Van Valen 1973) was proposed on the discovery that extinction rates appeared to be independent of lineage age for a range of taxonomic groups. Boyajian (1991) corroborated this by showing that extinction rates were independent of lineage age even during periods of mass extinction for fossil marine families. More recently, Ezard et al. (2011) showed that extinction rates were more influenced by environmental factors in foraminiferans, rather than biotic factors such as relative clade size. Such evidence would point to a Hyd scenario of evolution instead of a Pan scenario. However, in some instances extinction rates have been found to demonstrate positive age dependence. Quental and Marshall (2013) demonstrated that an island biogeography model of clade rise and fall could explain the trajectories of recently extinct mammalian orders, a process akin to the DE scenario. Additionally, Wilshire et al. (2014) developed a new method for measuring age dependence in extinction rates and found that foraminiferans all showed positive age dependence.

Given this fossil evidence, how can we assert our claim that extinction rates show negative age dependence as per our results? We would suggest that there are multiple factors determining extinction rates, which are generally independent of lineage age and its correlate, evolutionary distinctness. A Pan scenario, however, can be recreated with only the most evolutionarily distinct lineages, which represent a small subset of total biodiversity, having reduced extinction rates. Indeed, as a Pan model continues to run, the relative skew of evolutionary distinctness continues to grow, making the less evolutionarily distinct tips of the tree have random extinction with respect to one another. Statistically detecting such a non-linear relationship for a small subset of the most evolutionarily distinct species from the fossil record would be difficult. However, by taking a tree-growth modelling approach, we demonstrate how Hyd and DE scenarios are implausible general models of evolution. Both Hyd and DE scenarios lead to the continual reduction in tree age due to the loss of the most evolutionarily distinct branches – the ‘shrinking tree problem.’ Only a Pan model of evolution can recreate highly evolutionarily distinct lineages such as horseshoe crabs, ginkgos and tuataras while maintaining tree imbalance. Trees produced with the Hyd and DE scenarios had respectively a third and a half less the age of Pan trees, which would make the persistence of such isolated lineages highly unlikely through geological time. Slowly evolving, species-poor lineages are required to maintain the age of a tree.

In Raup's (1985) terminology, our modelling framework fits into a time-inhomogeneous category. This is because although the global parameters of speciation and extinction remained constant, the relative rates differed between clades as a function of the proportion of evolutionarily distinct species each clade represented. By placing differential rates of diversification on potentially competing clades, our modelling approach is related to the density-dependence logistic models of Raup (1985), Sepkoski (1978, 1979, 1984) and others. Furthermore, different scenarios resemble the simulations of previous modellers. The DE scenario that causes the sharp rise and fall of clades as a result of increasing and decreasing extinction and speciation rates respectively for older clades is most related to the island biogeography model of clade rise and fall of Quental and Marshall (2013). The Hyd scenario that modifies speciation rates and keeps extinction rates constant is akin to the coupled logistic models of directly competing clades (Miller and Sepkoski 1988; Roy 1996; Sepkoski et al. 2000). Despite these similarities, the differences in process from our approach should be highlighted: these models generally attempt to explain taxonomic diversity against time whereas our simulations modelled phylogenetic trees, and we make no explicit assumptions about the roles of biotic or abiotic forces shaping diversity nor attempt to capture potential competition (see review by Voje et al. 2015).

Gould et al. (1987) found clades tended to be bottom-heavy or 'regressive' for a wide range of taxonomic groups. All of the clades in our simulations were top-heavy, although Pan produced the least top-heavy clades. Our results, however, are not directly comparable to those generated from the fossil record. This is due to differences in how a clade is defined between our analyses and in the fossil record. We used a strict monophyletic definition because we were working with phylogenetic trees rather than simulating morphological shifts; those working on the fossil record do not have access to full phylogenetic information and have hence preferred a morphological paraclade/clade definition. In Raup's (1980) terminology we used a phylogeneticist approach, while Gould et al. (1987) used an evolutionary taxonomist approach. We would therefore expect to calculate higher estimates of centre of gravity in our simulated clades with respect to the results of Gould et al. (1987) because the phylogeneticist approach counts all descendants from a node, while the evolutionary taxonomist approach would only count those descendants that are deemed to still be part of the higher clade. Furthermore, the phylogeneticist approach should always lead to top-heavy estimates of gravity due to double counting as a result of the hierarchical nature of a tree: for every diversifying node there is a parent node, yet for every node soon to go extinct there are no child nodes. For these reasons we argue that the clade trajectories of the Pan scenario most closely match those found in the fossil record.

It should be stressed, however, that we have not set out to develop a 'general model of evolution'. We set out to determine what biases in diversification with respect to evolutionary distinctness best reproduce observed phylogenetic trees. Factors other than evolutionary distinctness and its correlates may prove to be more important in shaping other macroevolutionary patterns. For example, the high variance in γ in empirical trees – a factor none of our simulations were able to recreate – may reflect the importance of adaptive radiations (Simpson 1944). Indeed, unexpectedly high levels of imbalance have previously been used to identify adaptive radiations (Purvis et al. 2011). Excluding the signal generated by radiations, however, indicates that turnover – speciation and extinction without net gain or loss in species number – is the secondary key process. Alfaro et al. (2009) found that up to 85% of present-day vertebrate species could be the result of adaptive radiations, with the remaining 15% the result of turnover. Turnover has been noted in the fossil records of many major clades, e.g. Carnivora (Liow and Finarelli 2014). The Pan scenario is essentially a turnover process as tips have the same rate of speciating or going extinct, with these rates being determined by ED values. Turnover alone is able to explain the phenomenon of high values of γ despite trees as a whole not experiencing higher births to deaths (Boettiger and Temple 2012), as typically seen in the fossil record (Alroy et al. 2008). Lineages with high turnover will continually add new nodes towards the tips, while lineages with low turnover will maintain tree age and provide few nodes towards the root.

Evolutionary distinctness is an increasingly accepted proxy for the living fossil. Cavin and Kemp (2011) used it to identify lungfish as the most evolutionarily distinct vertebrate species – much more so than coelacanths – and recommended them for the ‘living fossil’ tag. Evolutionary distinctness, however, is only a proxy for the living fossil as it cannot capture all elements of the living fossil concept, most notably ‘little change through time’ (Eldredge and Stanley 1984). Regardless of scenario (Pan, Hyd or DE), under a Brownian motion model of character evolution we would expect ED species to be more ecologically and morphologically distinct as they have the greatest evolutionary distance from other species (Isaac et al. 2007). Few studies have been performed to investigate this question. Although Redding et al. (2010) showed that high ED primates were ‘odder’ compared to low ED primates, this appears not to be the case from the fossil record when considering lineage ages and character states. Older lineages of fossil crinoids have been found to be less morphologically distinct than younger lineages (Liow 2004) and trachyleberidid ostracodes were found to be morphological equivalent between older and younger lineages (Liow 2006). Wagner and Estabrook (2014) showed that paired character traits across a wide range of fossil groups and periods showed hierarchical stratigraphic compatibility: character states show morphological progression through the loss of earlier states. It could be argued that species that may still retain these ‘primitive’ character states as they disappear globally are more likely to be evolutionary distinct as they are less likely to produce future progeny. Under a punctuated equilibrium model of evolution where most character evolution occurs during speciation events (Gould and Eldredge 1993), only a Pan scenario would recreate the morphological primitiveness of living fossils as it would generate high ED species through long periods of low diversification rates rather than recent rapid extinction. There is recent evidence that speciation events do indeed accelerate such change (Venditti and Pagel 2010; Ezard et al. 2013). Furthermore, recent pan-vertebrate genomic analyses of protein-coding sequences and genomic rearrangements indicate that species commonly identified as living fossils do experience reduced rates of change (Amemiya et al. 2013; Venkatesh et al. 2014). Regardless, we do not suggest that ED species equate to living fossils under all possible definitions. By demonstrating that positive feedback in diversification rates is required for producing the observed shape of the tree of life, we show that evolutionary performance is unequally distributed and that ‘living fossil’ as a label can have more meaning than just the tail end of a normal distribution. It can be used to describe a true grouping of lineages with substantially lower rates of diversification than seen in other parts of the tree of life, rates that are reflective of the hollow curve distribution itself, and representing an emergent feature of a turnover-based model of evolution. Our aim was not to determine what the processes are that allow living fossils to arise, but to validate or invalidate potential definitions of the living fossil.

By demonstrating the validity of a panchronic definition, we hope to have opened the door for future research that seeks to determine what these processes may be.

Acknowledgements

Thanks to A. Purvis, L. Signorile, S. Dures, M. Bennett and K. A. Jønsson for useful initial discussions, to L. H. Liow for introducing the term ‘panchronic’, and to P. Wagner and an anonymous reviewer whose insightful comments and suggestions have greatly improved this manuscript.

Literature cited

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, … L.J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences of the United States of America 106(32):13410–13414.

Alroy, J. 1996. Constant extinction, constrained diversification, and uncoordinated stasis in North American mammals. Palaeogeography, Palaeoclimatology and Palaeoecology 127:285–311.

Alroy, J., M. Aberhan, D. J. Bottjer, M. Foote, F. T. Fürsich, P. J. Harries, … C. C. Visaggi. 2008. Phanerozoic trends in the global diversity of marine invertebrates. Science 321:97–100.

Amemiya, C. T., J. Alföldi, A. P. Lee, S. Fan, H. Philippe, I. Maccallum, … K. Lindblad-Toh. 2013. The African coelacanth genome provides insights into tetrapod evolution. Nature 496:311–6.

Blum, M., and O. François. 2006. Which random processes describe the Tree of Life? A large-scale study of phylogenetic tree imbalance. Systematic Biology 55(4):685–691.

Boettiger, C., and D. Temple Lang. 2012. Treebase: An R package for discovery, access and manipulation of online phylogenies. Methods in Ecology and Evolution 3(6):1060–1066.

Bortolussi, N., E. Durand, M. Blum, and O. François. 2006. apTreeshape: Statistical analysis of phylogenetic tree shape. Bioinformatics 22(3):363–364.

Boyajian, G. E. 1991. Taxon age and selectivity of extinction. Paleobiology 17(1):49–57.

Brayard, A., G. Escarguel, H. Bucher, C. Monnet, T. Brühwiler, N. Goudemand, … J. Guex. 2009. Good genes and good luck: ammonoid diversity and the end-Permian mass extinction. Science 325:1118–1121.

Britton, T., B. Oxelman, A. Vinnersten, and K. Bremer. 2002. Phylogenetic dating with confidence intervals using mean path lengths. Molecular Phylogenetics and Evolution 24(1):58–65.

Britton, T., C. L. Anderson, D. Jacquet, S. Lundqvist, and K. Bremer. 2007. Estimating divergence times in large phylogenetic trees. Systematic Biology 56(5):741–752.

Cavin, L., and A. Kemp. 2011. The impact of fossils on the evolutionary distinctiveness and conservation status of the Australian lungfish. Biological Conservation 144(12):3140–3142.

Casane, D., and P. Laurenti. 2013. Why coelacanths are not "living fossils": a review of molecular and morphological data. BioEssays 35(4):332–8.

Colless, D. H. 1982. Review of phylogenetics: the theory and practice of phylogenetic systematics. Systematic Zoology 31:100–104.

Darwin, C. 1859. On the origin of species by means of natural selection. London: J. Murray.

Eldredge, N., and S. M. Stanley. 1984. Living Fossils (Casebooks in Earth Sciences). Springer.

Ezard, T. H. G., T. Aze, P. N, Pearson, and A. Purvis. 2011. Interplay between changing climate and species' ecology drives macroevolutionary dynamics. Science 332:349–51.

Ezard, T. H. G., G. H. Thomas, and A. Purvis. (2013). Inclusion of a near-complete fossil record reveals speciation-related molecular evolution. Methods in Ecology and Evolution 4(8):745–753.

Forey, P. 1984. The coelacanth as a living fossil. *in* N. Eldredge and S. M. Stanley, ed. Living fossils (Casebooks in Earth Sciences). Springer 166–169.

Gould, S. J., and N. Eldredge. 1993. Punctuated equilibrium comes of age. Nature 366(6452):223–227.

Hagen, O., K. Hartmann, M. Steel, and T. Stadler. 2015. Age-dependent speciation can explain the shape of empirical phylogenies. Systematic Biology 64(3):432–440.

Hay, J.M., S. Subramanian, C. D. Millar, E. Mohandesan, D. M. Lambert. 2008. Rapid molecular evolution in a living fossil. Trends in Genetics 24:106–109.

Helmus, M. R., T. J. Bland, C. K. Williams, and A. R. Ives. 2007. Phylogenetic measures of biodiversity. The American Naturalist 169(3):E68–E83.

Isaac, N. J. B., S. T. Turvey, B. Collen, C. Waterman, and J. E. M. Baillie. 2007. Mammals on the EDGE: conservation priorities based on threat and phylogeny. PloS One 2(3):e296.

Liow, L. H. 2004 A test of Simpson's "Rule of the Survival of the Relatively Unspecialized" using fossil crinoids. The American Naturalist, 164:431–443.

Liow, L. H. 2006 Do deviants live longer? Morphology and longevity in trachyleberidid ostracodes. Paleobiology 32:55–69.

Liow, L. H., and J. A. Finarelli. 2014. A dynamic global equilibrium in carnivoran diversification over 20 million years. Proceedings of Royal Society London B: Biological Sciences 281:20132312.

Gould, S. J., N. L. Gilinsky, and R. Z. German. 1987. Asymmetry of lineages and the direction of evolutionary time. Science 236:1437–1441.

Manceau, M., A. Lambert, and H. Morlon. 2015. Phylogenies support out-of-equilibrium models of biodiversity. Ecology Letters 18:347–356.

Miller, A. I., and J. J. Sepkoski. 1988. Modeling bivalve diversification: the effect of interaction on a macroevolutionary system. Paleobiology 14(4):364–369.

Mooers, A., and S. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. Quarterly Review of Biology 72:31-54.

Nagalingum, N. S., C. R. Marshall, T. B. Quental, H. S. Rai, D. P. Little, and S. Mathews. 2011. Recent synchronous radiation of a living fossil. Science 334:796–799.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20(2):289–290.

Piel, W. H., M. Donoghue, and M. Sanderson. (2002). TreeBASE : A database of phylogenetic information. To the Interoperable "Catalog of Life" with Partners - Species 2000 Asia Oceanea (171):41–47.

Purvis, A., S. A. Fritz, J. Rodríguez, P. H. Harvey, and R. Grenyer. (2011). The shape of mammalian phylogeny: patterns, processes and scales. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 366(1577):2462–77.

Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proceedings of the Royal Society B: Biological Sciences 267(1459):2267–2272.

Quental, T. B., and C. R. Marshall 2013. How the Red Queen drives terrestrial mammals to extinction. Science 341:290–2.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Systematic Biology 64(2):340–355.

Global Names Architecture 2015. Global Names Resolver. [online] Available at:http://resolver.globalnames.biodinfo.org/ [Accessed 14 Jul. 2015].

Raup, D. M. 1985. Mathematical models of cladogenesis. Paleobiology 11(1):42–52.

Redding, D. W., C. V. DeWolff and A. Ø. Mooers. (2010). Evolutionary Distinctiveness, Threat Status, and Ecological Oddity in Primates. Conservation Biology 24(4):1052–1058.

Royer, D.L., L.J. Hickey, and S.L. Wing. 2003. Ecological conservatism in the "living fossil" Ginkgo. Paleobiology 29(1):84–104.

Roy, K. (1996). The roles of mass extinction and biotic interaction in large-scale replacements: A reexamination using the fossil record of stromboidean gastropods. Paleobiology 22(3):436–452.

Rudkin, D.M., G. A. Young, and G. S. Nowlan, 2008. The oldest horseshoe crab: a new xiphosurid From Late Ordovician Konservat-Lagerstätten deposits, Manitoba, Canada. Palaeontology 51(1):1–9.

Sackin, M. J. 1972. “Good” and ”bad” phenograms. Systematic Zoology 21:225–226.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19(1):101–109.

Schluter, D. 2000. The Ecology of Adaptive Radiation. Oxford University Press, Oxford.

Schopf, T.J.M., 1984. Rates of evolution and the notion of “living fossils.” Annual Review of Earth Planetary Science 12:245–292.

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

Sepkoski Jr., J. J. 1978. A kinetic model of Phanerozoic taxonomic diversity I. Analysis of marine orders. Paleobiology 4(3):223–251.

Sepkoski Jr., J. J. 1979. A kinetic model of Phanerozoic taxonomic diversity II. Early Phanerozoic families and multiple equilibria. Paleobiology 5(3):222–251.

Sepkoski Jr., J. J. 1984. A kinetic model of Phanerozoic taxonomic diversity III. Post-Paleozoic families and mass extinctions. Paleobiology 10(2):246–267.

Sepkoski Jr., J. J., F. K. McKinney, and S. Lidgard. 2000. Competitive displacement among post-Paleozoic cyclostome and cheilostome bryozoans. Paleobiology 26(1):7–18.

Venditti, C., and M. Pagel. 2010. Speciation as an active force in promoting genetic evolution. Trends in Ecology and Evolution 25(1):14–20.

Venkatesh, B., A. P. Lee, V. Ravi, A. K. Maurya, M. M. Lian, J. B. Swann, … W. C. Warren. 2014. Elephant shark genome provides unique insights into gnathostome evolution. Nature 505(7482):174–9.

Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1:1–30.

Voje, K. L., Ø. H. Holen, L. H. Liow, and N. C. Stenseth. 2015. The role of biotic forces in driving macroevolution: beyond the Red Queen. Proceedings of the Royal Society B: Biological Sciences 282:1–9.

Wagner, P. J. and G. F. Estabrook. 2014. Trait-based diversification shifts reflect differential extinction among fossil taxa. Proceedings of the National Academy of Sciences 111:16419–16424.

Wiltshire, J., F. W. Huffer, and W. C. Parker. 2014. A general class of test statistics for Van Valen's Red Queen hypothesis. Journal of Applied Statistics 41(9):2028–2043.

Yoshida, K. 2002. Long survival of "living fossils" with low taxonomic diversities in an evolving food web. Paleobiology 28(4):464–473.

Yule, G. U. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Philosophical Transactions of the Royal Society B: Biological Sciences 213(1925):21–87.

Figure Captions

Figure 1. The two dimensions of phylogenetic tree shape for a 16 tipped tree: balance and gravity, a new term we coin after ‘centre of gravity’. Trees with most nodes towards the tip of the tree would be more likely to fall over (high centre of gravity), whereas tips towards the root would be less likely to topple (low centre of gravity).

Figure 2. (a) How to calculate evolutionary distinctness from a phylogenetic tree: divide every branch length by the number of descendants, sum these proportional branch lengths for every tip in the tree to calculate evolutionary distinctness scores. The total of these scores equals the total branch length of the tree. (b) How different ED values impact speciation probability, S, in the next time step given different σ values for tree in Figure (2a). Under an equal rates Markov model where rates of diversification are equal between tips (σ=0), all the tips have an equal chance of diversifying. This will maintain tree imbalance, as nodes with more descendent tips have a greater chance of diversifying, gravity will remain at a mid-level provided birth and death parameters are equal. Modelling this same tree under an Evolutionary Distinctness Biased Markov model will lead to, depending upon the model’s parameters, different outcomes. For example, under a Pan simulation (σ=-1) the most evolutionarily distinct tip (4.00000) will become ever more distinct as it will have lower rates of speciation and extinction. While the tips with lower scores will be more likely to speciate and go extinct, extenuating tree imbalance and increasing tree gravity.

Figure 3. (a) σ-ε (“sigma”-“epsilon”) parameter space indicating the four quadrants for living fossils: Phylogenetic Fuse (PF), Dead-End (DE), Ephemeral (Eph) and Panchronic (Pan). The

point at 0,0 indicates an Equal Rates Markov Model. The extreme values for each scenario are marked with “X”s. (b) Example12-tipped trees, their expected shapes and corresponding ED values for the five main simulation scenarios. PF and Eph produce balanced trees with moderate levels of gravity. Pan, Hyd and DE all produce unbalanced trees but with different values for gravity, as demonstrated in the more distributed ED values as gravity becomes higher. (c) Distances of 0.1x0.1 grid cells of mean PC1 and PC2 values in simulated space to empirical trees. The smaller the value, the closer to empirical tree shape. Values generated from a PCA of the Colless, Sackin and PSV metrics of simulated and empirical trees. Both PCs are weighted equally.

Figure 4: Chromatophyla of trees with fossil taxa produced under the three most likely scenarios of ED-biased evolution: DE, Hyd, and Pan. Darker branches indicate greater numbers of descendants; darker areas of the tree indicate points of diversification. Under the DE scenario, the resulting tree without fossils is younger than the equivalents under both Pan and Hyd due to the pruning of ED tips – ‘the shrinking tree problem’. The DE tree’s most distinct branches are extinct, the Pan tree’s ED branches are all extant and the tree is as old as the simulation time span. (See online for colour version.)

Figure 5. Clade dynamics between the 25th and 75th time intervals for the 100 largest clades produced by exemplary simulations of the DE, Hydra and Pan scenarios. Each clade trajectory through time has a different colour (see online for colour version).

Figure 6. PCA of trees simulated using the extreme scenarios and empirical trees using Yule-normalised Colless and Sackin and PSV. Error bars are standard errors. The three points for empirical trees represent the trees produced using different rate-smoothing methods. PC1 correlates with Sackin and Colless and PC2 correlates most with PSV, making the two PCs equivalent to imbalance and gravity dimensions of tree shape.