How big is a genus? Towards a nomothetic systematics

Running Head: How big is a genus?

Keywords

genus, species-within-genus statistics, Linnean taxonomy, taxonomic rank, birth-death process, macroevolution

Manuscript components

Manuscript (below), with 4 colour figures and 1 table

Supplementary Data SD1. 'MBL 2017': Software used to generate simulated phylogenetic trees and synthetic taxonomy. The package contains 15 files. MBL2017 can be executed on PC or Mac but requires the Qt library (www.qt.io).

Supplementary Data SD2. Supplementary explanation of results, including description of taxonomic sorting algorithms, example taxonomically-sorted output from tree simulations, data quality approach to real-world taxonomic data, frequency distributions from simulated data and 'real-world' data, and quantitative comparisons among real-world datasets.

Abstract

A genus is a taxonomic unit that may contain one species (monotypic) or thousands. Yet counts of genera or families are used to quantify diversity where species-level data are not available. High frequencies of monotypic genera (~30% of animals) have previously been scrutinised as an artefact of human classification. To test whether Linnean taxonomy conflicts with phylogeny, we compared idealised phylogenetic systematics *in silico* with real-world data. We generated highly-replicated, simulated phylogenies under a variety of fixed speciation/extinction rates, imposed three independent taxonomic sorting algorithms on these clades (2.65×10^8 simulated species), and compared the resulting genus size data with quality-controlled taxonomy of animal groups (2.8×10^5 species). 'Perfect' phylogenetic systematics arrives at similar distributions to real-world taxonomy, regardless of the taxonomic algorithm. Rapid radiations occasionally produce a large genus when speciation rates are favourable; however, small genera can arise in many different ways, from individual lineage persistence and/or extinctions creating subdivisions within a clade. The consistency of this skew distribution in simulation and real-world data, at sufficiently large samples, indicates that specific aspects of its mathematical behaviour could be developed into generalised or nomothetic principles of the global frequency distributions of higher taxa. Importantly, Linnean taxonomy is a better-than-expected reflection of underlying evolutionary patterns.

Introduction

The classification of organisms (systematics) does not always conform to their evolutionary history (phylogenetics). The identification of species pre-dates any kind of evolutionary paradigm, and indeed pre-dates any kind of science (Hopwood, 1959; Mayr, 1982), so it is reasonable for specialists to consider how to reconcile older and widely-used systems of classification with tree-based thinking. Treatment of taxonomic ranks above the species level is the subject of extensive ongoing debate in the field of biological systematics and macroevolution (Hendricks et al., 2014; Giribet, Horminga & Edgecombe, 2016). Many authors suggest that species are real products of evolution, while higher-ranked groupings are arbitrary constructs (e.g., Stork et al., 2015). Meanwhile, Linnean ranked taxa, that represent nested groups of species, are accepted as biologically 'real' in other fields of science and beyond.

Most fields of biology simply use taxonomic names to address their own questions. Taxonomic 'surrogacy' (using counts of families or genera to measure biodiversity), is applied where species-level identifications are not readily available (Gaston & Williams, 1993; Ricotta, Ferrari, & Avena 2002; Bertrand, Pleijel & Rouse 2006; Heino, 2014). At small scales, environmental impact assessments of a single local ecosystem will generally yield equivalent results whether all present taxa are identified to species level or not (taxonomic sufficiency: Ellis, 1985; Timms et al., 2013). Taxonomic surrogacy is also used in synoptic study of the global fossil record, where species-level identifications may not be available because of preservational limitations. Counting the succession of fossil genera and families – not species – is the basis for the current understanding of macroevolution and global extinction patterns (Raup & Sepkoski, 1986; Lu, Yogo & Marshall, 2006; Alroy et al., 2008; Hendricks et al., 2014).

A genus can contain many species, or it can contain a single species. The issue of inconsistent genus size has been mooted as a major impediment to studying extinction, though it has rarely been addressed directly (Quental & Marshall, 2010). Taxonomic conventions for what constitutes sufficient distinction for a particular rank are not formally articulated, but appear to differ among organismal groups (Avise & Liu, 2011). A better understanding of the diversity represented by the genus rank is important for attempts to estimate species diversity in any field that uses taxonomic surrogacy. The genus is the lowest commonly-used rank among supraspecific classifications and the most widely used for taxonomic surrogacy; in this study we focus on the genus to enable the gathering of a large empirical dataset.

Many groups of living animals and plants have a high frequency of monotypic genera, and decreasing numbers of larger genera; this skew distribution is termed the 'hollow curve' and has been recognised and discussed since the early 20th Century (e.g. Yule, 1925; Kendall, 1948; Holman, 1985). Such diversity patterns have many applications beyond the field of systematics itself. Early work compared the skew distributions seen in taxonomic rank and other natural patterns, such as body size and species-area curves (Yule, 1925; Anderson, 1974), though the interactions of these processes are not straightforward. Building directly on the observation that ranked taxonomic frequency distributions appear consistent, the 'hollow curve' pattern has been used to predict global species richness from higher ranked taxa (Mora et al., 2011). Global taxonomic initiatives for living diversity face the same data limitations as studies of macroevolutionary trends in the fossil record: most higher-rank taxa have been discovered while a large proportion of species remain undescribed (Costello, May & Stork, 2013), and they are dependent on primary taxonomic datasets that may themselves be controversial (e.g. Bass & Richards, 2011). A demonstration that the hollow curve is an emergent property of evolutionary processes and consistent across various groups of

100 organisms, rather than a potentially inconsistent taxonomic artefact, would thus have
101 considerable power.

102 This hollow curve has been repeatedly observed for almost a century, yet often
103 considered puzzling (Yule, 1925; Holman, 1985; Aldous, 2001; Aldous, Krikun & Popovic,
104 2011). Some of the variability in genus size has even been attributed to taxonomic cultural
105 factors, such as personality-driven tendencies in individual taxonomists toward 'splitting' or
106 'lumping' or human preferences for classification in smaller or larger groups (Fenner, Lee &
107 Wilson, 1997; Scotland & Sanderson, 2004). Previous studies of genus size have focussed on
108 'top down' approaches, developing simulations that accurately replicate the observed size-
109 frequency distribution of taxonomic datasets (e.g. Yule, 1925; Maruvka, 2013), or compare
110 observed patterns with specific probability distributions (e.g. Scotland & Sanderson, 2004).
111 Our aim in the present study is to use a 'bottom up' approach, starting with species evolution
112 and applying a perfectly objective classification, to examine whether or not the skew
113 distribution in higher taxa is in conflict with underlying phylogenetic processes.

114 Within a phylogeny, sister-taxa are not necessarily of equivalent rank. The sister
115 taxon of a genus may also be a genus, or it may be a species, a family or other higher taxon,
116 or an unranked group of genera. This has raised questions about the viability of ranked taxa
117 in a phylogenetic framework, though it is not necessarily problematic (Giribet et al., 2016).
118 Importantly, it also means that observed patterns in established taxonomic classification are
119 not equivalent to phylogenetic 'imbalance' or the relative size of nested and adjacent clades
120 (Aldous, Krikun & Popvic, 2008). This is because the size-frequency distributions of
121 subclades predicted *a priori* by birth-death processes may not be equivalent to those of
122 taxonomic units recognised *a posteriori*.

123 Species richness in living clades is controlled not by speciation alone, but also by
124 times of lineage persistence and extinction events, as these create 'space' within a clade, gaps

that separate living species into discrete groups that may be treated as higher taxonomic entities (e.g. genera). Extinction processes are a critically important process to producing the species richness in a clade (Marshall, 2017). Extinction is inevitable over evolutionary time, and lineage loss within a clade creates discontinuities in phenotypic or genetic gradients, while accumulated branch lengths over clade evolution results in more diversity and hence more potential for generic splitting. Thus, there is only one evolutionary pathway to a large genus (a single rapid radiation), but many ways to create a small genus, such as a persistent, unbroken and relatively unchanging evolutionary lineage, or the extinction of other closely-related species in a clade, or lineage persistence or extinction events nested within a larger clade that separate species into multiple genera. This may explain why clade size, like many natural phenomena, has a hollow curve (Yule, 1925; Strand & Panova, 2014).

Literature in phylogenetics is often focussed on analysing rapid radiations and the causative explanations of their evolutionary history (e.g. Bond & Opell, 1998; Alfaro et al., 2009; Harmon & Harrison, 2015). Our goal here was to return to basic principles and examine large-scale emergent patterns in diversification, regardless of individual clade history, that could provide a more fundamental basis to identify where taxonomically defined genera may constitute genuine outliers.

It is unclear to what extent these repeatedly observed skew distributions in conventional taxonomic genus size are influenced by the real evolutionary history of clades, and consequently it is unclear whether supraspecific diversity can be confidently translated to a probabilistic approximation of species diversity. That is, if a taxon is only identified to genus level, is it possible to establish a probability envelope of how many species it represents globally? To address this question, we compared empirical and simulation data to determine the range of behaviour in genus size frequency distributions, and the variability of these distributions under different taxonomic algorithms and evolutionary rates. Consistent

behaviours in ‘real world’ taxonomy and in evolutionary simulations would indicate that generalised principles of systematics could lead to robust quantification of diversity from taxonomic surrogacy.

Early work on mathematical approaches to macroevolution used birth-death models (Kendall, 1948) to explore the impact of speciation and extinction rates on patterns of cladogenesis (Rannala et al., 1998; Huelsenbeck & Lander, 2003). David Raup (1933-2015) and colleagues produced a computer program they referred to as ‘MBL’ after a meeting in the Marine Biological Laboratory at Wood’s Hole, Massachusetts (Raup et al., 1973; Raup & Gould, 1974). Their explorations of the performance of birth-death models with this tool demonstrated the importance of the interplay of speciation and extinction rates (Sepkoski, 2012). These systems continue to provide a robust and elegant framework to explore macroevolutionary dynamics (Nee, 2006; Budd & Jackson, 2016).

Tree simulation based on birth-death systems, with high replication resulting from modern computing power, is here used to assess whether or not genus size distribution in real-world taxonomic data can be reproduced using simple models. We imposed three algorithmic taxonomic classifications on large samples of simulated trees, to compare a range of speciation and extinction parameters and their potential impacts on genus size trends. We also analysed a broad sampling of taxonomic data from living metazoans, to assess the consistency of size-frequency patterns. The present work thus uses a ‘null model’ approach to assess the degree of disparity between deliberately idealised simulations with empirical data drawn from real historical taxonomy. This framework is designed to address the question of whether ranked groups are arbitrary, or whether they can be reconciled with underlying phylogenetic patterns, and presents a significant first step in developing a predictive approach to infer species-diversity information from data with genus-level resolution.

Methods

Real-world taxonomy

We gathered comprehensive taxonomic datasets for a broad selection of animal groups. These datasets were selected primarily based on taxonomic completeness and global species coverage, and their acceptance and/or use by the community of relevant taxonomic experts. In each dataset, taxa were treated to the same stringent quality checking. Each database was filtered to exclude fossil species where present, and line checked to remove incomplete binomial epithets or false duplication due to genuine typographical errors. To facilitate comparisons across groups with potentially very different taxonomic conventions, it is necessary to impose certain *a priori* filters that could be applied to all the datasets. We did not include subspecies or subgenera in this analysis (following e.g. Alroy et al., 2001; Heim & Peters, 2011), because taxonomic species and genus ranks are the universal binomial epithet that are consistently available for all taxa. While all species are assigned to a genus, not all species are associated with a subgenus, and not all species are split into subspecies. Some prior studies on well-curated datasets of marine taxa 'elevated' subgeneric taxa to genus level (e.g. Raup, 1978). We consider such adjustments to be taxonomic revision that is the prerogative of relevant experts, and an aim of our study is to demonstrate whether the generic concept *as normally expressed* is comparable between groups, at least in terms of size distributions. We hence did not make any adjustments to the classification presented in the global taxonomic datasets we used here, even in the few groups where we have an appropriate level of expertise. Fossils were excluded both to ensure consistency across different datasets, but also to facilitate comparison with our simulations where all extinct species are excluded. We did not impose any further taxonomic refinement or interpretation, but where datasets recorded synonyms and reported them as such, only the valid accepted form was included in our analysis. These datasets include both monophyletic

and non-monophyletic groupings. (Further, within the large non-monophyletic dataset of marine invertebrates, some contained subgroups are incomplete because of non-marine species not included in the database.) We used these data to quantify the number of species in each valid genus for birds (Gill & Donsker, 2014), fish (Froese & Pauly, 2015), marine invertebrates (Boxshall et al., 2015), odonate insects (Schorr & Paulson, 2015), reptiles (Uetz & Hošek, 2014), and mammals (Wilson & Reeder, 2005).

Model background

Branching phylogenies can be modelled using 'birth-death' type models, and some emergent patterns can be understood from relatively simple mathematical properties that have been productively applied to macroevolutionary studies, and have a long history in mathematical literature (e.g. Watson, 1875). The standard birth-death type model begins with a single parent lineage. At each iterative time-step there is a set probability that the lineage will split into two daughter lineages (a 'birth' with probability noted lambda, λ), go extinct (a 'death' with probability noted mu, μ) or persist unchanged (with probability 1-λ-μ). The interactions of these parameters control several important properties of the descendent clade (fig. 1). Firstly, the probability of total extinction of the descendant clade is determined by the ratio μ/λ: if the extinction rate is higher than the speciation rate, then the descendant clade will eventually go extinct; otherwise the probability of total extinction decreases as μ/λ drops. This ratio is illustrated in figure 1 as the shades of grey in the probability space, where the black half above the diagonal $\mu=\lambda$ indicates inevitable total extinction. Secondly, the expected number of living descendent lineages at time t increases exponentially dependent on the difference (λ-μ) between speciation and extinction rates. This second property has been more frequently discussed in previous literature, especially in terms of the potential for rapid exponential growth of clades when the speciation rate exceeds the extinction rate (Raup,

1985). In biologically realistic scenarios, the values are near balanced (Marshall, 2017). This constraint, and the interaction of λ and μ have several interesting emergent properties. Any pair of parameters that have the same difference (λ-μ = constant), have the same (average) number of descendents in a fixed span of time (fig. 1). Thus, if the speciation rate (λ) is lower than the extinction rate (μ), the expected number of descendent species goes to zero (λ-μ < 0), and the clade inevitably goes extinct ($\mu/\lambda > 1$). If the speciation rate is much higher than the extinction rate, the population rapidly explodes into biologically unrealistic species richness.

Synthetic taxonomy

In the case of the present models, fixed speciation (λ) and extinction (μ) rates were used within each individual simulation in order to constrain the behaviour of the simulation. However, each individual simulation was relatively short (400 generations) so results are combined from large-scale replication.

We generated synthetic trees using a fast C++ implementation of the MBL model (Raup et al., 1973; Supplementary Data, SD1). Random numbers were imported as 32-bit unsigned integers from a 100Mb set of quantum random numbers downloaded from https://qrng.anu.edu.au (see Symul, Assad & Lam, 2011). Tree growth was initiated with one lineage at time t=0, and iterated for 400 generations. The code was tested through comparison of 10,000-tree runs with predicted theoretical values of rates of total extinction and mean survivorship at t=400. Observed values for both lay within 0.1% of predicted values (Supplementary Data SD2). We set no limit on tree size (unlike Raup et al., 1973, who were constrained by available computer memory). The software interface allows readers to run these simulations and to manipulate generation time, and threshold values for the taxonomic algorithms (Supplementary Data SD1).

We selected five pairs of values for the parameters λ (speciation probability at each iteration) and μ (extinction probability at each iteration) for use in this study. These were selected to give the same value of λ-μ = 0.01, and hence to provide the same value for mean number of species at t=400 in all cases (calculated as $e^{t\,(\lambda-\mu)} = e^4 \approx 54.6$ living species at time t=400). The parameter pairs were: λ=0.015, μ=0.005; λ=0.025, μ=0.015; λ=0.055, μ=0.045; λ=0.125, μ=0.115 and λ=0.200, μ=0.190 (fig. 1). For each parameter-pair we generated 10,000 successful trees – i.e. all trees that experienced total extinction before t=400 were discarded and the simulation was continued until 10,000 lineages survived to t=400.

In the surviving trees, we excluded all extinct lineages and only considered the species (tips) extant at t=400. We then imposed synthetic taxonomies to delineate species alive at the final sampling into 'genera'. Three approaches to taxonomy were used, here termed *Relative-Difference Taxonomy* (RDT), *Internal-Depth Taxonomy* (IDT), and *Fixed-Depth Taxonomy* (FDT). All three algorithms produce only monophyletic genera, identified using different features of the internal topology of the tree (fig. 2; Supplementary Data fig. SD2.1).

Relative-Difference Taxonomy (RDT) makes no assumption that genera should be similar in age and implements a relatively complex set of rules, to formally articulate sorting from the general principles of phylogenetic systematics. This asserts that a genus should be a grouping containing those species that are relatively phylogenetically closer to each other than they are to anything outside the genus group. In our algorithm, all sister-species pairs were *de facto* united in a genus, along with any additional taxa that formed a clade without exceeding the relative-distance threshold. Where the threshold is 0.5, this means more than doubling the phylogenetic distance between nodes. We tested the algorithm's sensitivity to the relative distance threshold with four different values (0.3, 0.5, 0.6, 0.75). All extant

species not placed in a genus by this pairing/expansion algorithm are left as monospecific genera (fig. 2; Supplementary Data, fig. SD2.1).

Internal-Depth taxonomy (IDT) operates on a similar principle of relative differentness but uses an unrelated algorithm. Under IDT, a genus is a group of species lineages whose internodal distances are always less than a fixed threshold. Where a lineage persists without splitting for longer than the threshold distance, the downstream branches establish a new genus, and any paraphyletic genera are automatically split into monophyletic units. Four threshold values were tested, at 3.75%, 5%, 10%, and 15% of total simulation time (15, 20, 40, and 60 time-iterations).

Fixed-Depth taxonomy (FDT) defines a genus to comprise all species diverging for less than a constant amount of time. Avise and Johns (1999), for example, suggested divergence in the interval 2–5Ma for contemporary species. FDT groups into one genus all species whose most recent common ancestor occurred at or after a 'threshold' number of time-iterations from the end of the simulation. This threshold was tested at 3.75%, 5%, 10%, and 15% of total simulation time (15, 20, 40, and 60 time-iterations) for this study. The approach provides a naïve but easily understood taxonomy in which there is an absolute upper limit to the degree to which any two congeneric species can be separated from each other.

Simulations were repeated with four different thresholds for each algorithm, thus producing 12 taxonomic schemes for each speciation/extinction rate parameter set. Our software allows sorting to be completed in parallel for the three algorithms, thus 20 simulations were performed (4 threshold sets on each of 5 rate parameter pairs). Each simulation was run until 10,000 trees were produced.

Results

298 *Real-world taxonomy*

299 Size-frequency data of genus-level species richness are remarkably consistent among
300 all sampled datasets (fig. 3; table 1; supplementary data SD2). The largest fraction of genera
301 in any group is monotypic genera (size = 1 species), decreasing nonlinearly in frequency with
302 increasing genus size. The proportion of monotypic genera was around one-third of genera in
303 all sampled groups (28% to 43%; Table 1). The behaviour of the non-monophyletic groups
304 sampled (fish, marine invertebrates) did not differ from the other datasets. The same
305 universal behaviour emerges in sufficiently large samples. The general pattern of (a) a
306 skewed frequency-distribution of genus size, and (b) approximately one-third of genera being
307 monotypic, holds true in other subsampled partitions of monophyletic taxonomic orders (data
308 not shown).

309 The frequency distribution patterns among different organisms are visually similar
310 and may be statistically equivalent. While the distributions differ slightly in terms of the
311 proportion of monotypic taxa (the spread of values see on the left side in Fig 3), the question
312 of relevance is whether these frequency distributions deviate significantly from each other
313 over the whole span of genus sizes. Statistical tests to compare discrete distributions may
314 have limited information value, but pairwise two-tailed Kruskal-Wallis tests on proportional
315 frequencies (i.e. percentages of genera in each species-richness size for each taxonomic
316 group) found no significant difference at $\alpha=0.05$ between any two groups (all pairwise
317 comparisons $p < 0.039$), with the single exception of mammals and birds (pairwise
318 comparison, $D = 0.255$, $p = 0.0914$; Supplementary Data SD2). Mammalia is the smallest
319 dataset included in the analysis, and that deviation was driven by the size of the largest
320 mammal genera. The two largest mammal genera, are *Myotis* bats with 102 spp. and
321 *Crocidura* shrews at 173 spp. (the largest bird genus, *Zosterops*, is 87 spp.). Datasets were
322 compared based on percentages to accommodate the range of total size, and thus the one

large mammal genus represents a larger proportion of total mammal genus diversity. Mammal genera have a broader range of species-richness relative to birds, but neither of these two groups was significantly different from any other group, including the total group.

Size-frequency distributions followed a similar pattern in all groups; however, the sizes of the largest genera were distinctly different. The largest marine invertebrate genera are an order of magnitude larger than other groups that we examined (fig. 3; table 1). Nonetheless, the proportions of monotypic genera were consistent (table 1) and the overall frequency distributions are statistically equivalent (see above). Maximum genus size was also independent of taxonomic group, and did not correlate with the number of genera or total group species richness (genera: $p = 0.740$, species: $p = 0.780$).

Synthetic taxonomy

The real-world taxonomic data (fig. 3) and all three taxonomic rule-sets in simulation (RDT, IDT, FDT) consistently recovered broadly hollow-curve distributions of genus size, with proportionally higher numbers of small genera and smaller numbers of large genera (fig. 4). In summative simulation data (combining heterogenous speciation and extinction rates), the distributions are strongly similar to real-world data, and the proportion of monotypic genera is equivalent to that in real-world taxonomy (fig. 4d). Simulations, however, recovered maximum genus sizes that were substantially smaller than some reported from organismal taxonomy.

To exclude the possibility that maximum genus size was constrained primarily by clade size, we visualised the maximum genus size for every individual tree (10,000 trees per parameter set) under the three different taxonomic sorting algorithms (supplementary data

SD2). Under a combination of higher speciation/extinction parameters, and under higher (more lenient) threshold values, the maximum genus size does increase slightly with increasing clade size, but has a clear upper threshold that is orders of magnitude lower than the clade size. Genus size is hence not saturated or constrained by simulation tree size.

In simulation, the largest genus size recovered was a single instance of a genus with 675 species, under a broad threshold in IDT that was selected to examine extreme behaviour (supplementary data SD2, fig SD2.2; IDT threshold = 15%). In that simulation the frequency distribution of genus size becomes extremely flat with only 6% of species in monotypic genera, significantly diverging from patterns seen in 'real world' taxonomy. The largest genera recovered under more moderate threshold values were all under 350 species (fig 4).

The distributions of genus size from RDT simulations did not change substantially with different speciation/extinction-rate parameter pairs (fig. 4a). Changes in threshold value had no substantial effect on the resulting patterns (fig. 4a, Supplementary Data SD2, fig SD2.2). In these simulations, two-species genera are recovered most frequently, and the second-largest group is monotypic genera. This somewhat violates the expected 'hollow curve' where monotypic groups are otherwise the largest fraction of genera. This artefact arises from the RDT rules, in which any pair of sister-species form a genus regardless of the depth of their common ancestor. However, the artefact does not appear to extend to the rest of the curve, and we note that the combined proportion of one- and two-species genera is similar across all taxonomic algorithms. While this has some implications for the use of topological criteria (discussed below), we do not consider than the overall pattern undermines the expectation of dominant monotypes in taxonomy.

The proportion of monotypic genera, and the size of the largest genera recovered, were less sensitive to changing parameters than under either FDT or IDT. Among all the

parameter sets tested, the proportion of monotypic genera ranged from 36.6% to 47.2%, and the size of the largest genera recovered ranged from 8 to 36 species per genus (Supplementary Data SD2, fig SD2.2, SD2.3), closely in line with proportions in real-world taxonomy (Table 1).

The IDT algorithm consistently recovered larger maximum genus sizes than the other two algorithms. Increasing rates of speciation resulted in broader and flatter genus size-frequency distributions (fig. 4b). This 'flattening' decreased the left skew of the frequency distribution as evidenced in both a relatively lower proportion of monotypic species and larger maximum genus sizes. Speciation parameters at both extremes of our range of test values produce frequency distributions that deviate from the patterns seen in real-world taxonomic data. Variation in the threshold value did not alter the overall shape of the frequency distribution under any particular parameter set (fig. 4b), but increasing the threshold value caused the same flattening effect as increasing speciation rate (Supplementary Data SD2, fig SD2.2). The proportion of monotypic genera, and the size of the largest genus co-vary, ranging from 6.1% monotypic with a maximum genus size of 674 species, under the highest speciation rate and highest threshold tested (λ=0.20, threshold 15%) to up to 79.2% and a largest genus size of 10 species under the lowest parameters (λ=0.015, threshold 3.75%).

Fixed-Depth taxonomy (FDT) recovers distribution patterns that are similar to IDT However, fixed-depth taxonomy is much more sensitive to changes in speciation-/extinction-rate parameters, varying slightly more than IDT with changing speciation rates, and like IDT an increase in speciation rate resulted in increasingly broad genus size-frequency distributions (fig. 4c). Under all variations, the proportion of monotypic genera ranged from only 4% of genera monotypic to 79% of genera monotypic (Supplementary Data SD2, fig SD2.2). For the lowest speciation rate applied (λ=0.015), up to 73.7% of FDT simulated

genera were monotypic under a 10% threshold, compared to 13.2% of genera monotypic under the highest speciation rate applied (λ=0.200). FDT recovers lower maximum genus sizes than IDT. Increasing rates of speciation produced increasingly larger maximum genus sizes, ranging from 10 species per genus under the lowest speciation rate to a genus with 75 species under the highest simulated speciation rate, or up to 156 species in the largest single genus from a 15% threshold (fig. 4c). Increases in threshold values, like IDT, created the same effect on the resulting frequency distribution as increasing speciation rate parameters (Supplementary Data SD2, fig SD2.2).

Combining the data for all five speciation/extinction parameter sets provides a visualisation of the central tendency of the behaviour for each algorithm (fig 4d). All three taxonomic algorithms produced frequency distributions that were similar to each other and strongly similar to the hollow curve distributions found in real-world taxonomy.

Discussion

Size-frequency distributions

Discussion abounds over the potential inconsistency of taxonomic delimitations (Gift & Stevens, 1997). Different organismal groups are classified with different interpretations of rank, especially comparing invertebrate and vertebrate groups (Avise & Johns, 1999; Avise & Liu, 2011). This inconsistency or apparent instability may seem to be a fundamental handicap to modernising systematic classifications. In this context it is interesting that the size frequency of metazoan genera converges on a strongly consistent pattern, and that pattern also agrees mathematically with distributions that emerge from idealised phylogenetic simulations.

Our results demonstrate that the sizes of higher ranks behave in a predictable fashion, supporting their use as a proxy for specific diversity (taxonomic surrogacy) in synoptic

studies. These patterns emerge consistently, at sufficiently large samples. Taxonomic surrogacy has many practical advantages for measuring biodiversity, which underlie the widespread use of that approach. Work on morphological disparity in living species has supported the utility of higher ranked taxa (Triantis et al., 2016). And, even more frequently, synoptic work on the fossil record has reinforced the importance of evolutionary information from higher ranks (Raup & Boyajian, 1988). For a few well-studied groups, there is demonstrable congruence in species phylogeny and morphologically defined genera (e.g. Jablonski & Finarelli, 2009; Humphreys & Barraclough, 2014; Holt & Jønsson, 2014). These provide significant hope or reassurance that it is theoretically possible to apply traditional Linnean classifications where taxonomic ranks have a clearly articulated evolutionary or temporal delimitation. Nonetheless, the question of whether genera represent real biological or evolutionary entities has not been directly addressed outside those very few groups for which phylogenetic studies with dense taxon sampling are available. A lack of certainty about which patterns are universal or artefactual remains a persistent criticism of the transferable meaning of ranked taxonomy (Lee, 2003).

The dominance of monotypic genera, and the rarity of large genera, is an established consistent pattern that has been 're-discovered' repeatedly for more than a century (Aldous, 2001). Indeed, the pattern should be expected from birth-death models (Kendall, 1948). One of our taxonomic algorithms recovered a high number of two-species genera, but only under a highly unrealistic taxonomic scenario (forcing sister-species to share a genus even if they deeply divergent). There is a significant body of work on the long-tailed distribution of species richness among genera (Yule, 1925; Maruvka et al., 2013), but the idea still persists that supraspecific groups are more arbitrary than species definitions and the skewed frequency distribution might be an artefact of taxonomic practice (e.g. Scotland & Sanderson, 2004; Strand & Panova, 2014). Our new data show, however, that this frequency pattern is

445 strongly consistent across independent groups, with different taxonomic approaches and
446 evolutionary histories. Our modelling demonstrates that it can arise from the interaction of
447 phylogeny and taxonomy alone.

448 The difference between taxonomic units and nested clades is a persistent
449 misunderstanding in controversies about the utility of ranked taxonomy (Giribet et al., 2016).
450 Even though our simulated genera are all monophyletic, the sister taxon of a genus is rarely
451 another genus. This is not problematic; it is a reflection of the intentionally relativistic nature
452 of ranked taxonomy. The patterns of nested clades in phylogenetic trees are informative to
453 evolutionary processes, but they are not equivalent to taxonomy. Mathematical patterns that
454 arise from topology have been referred to as tree 'imbalance' in computational phylogenetics
455 (Mooers & Heard, 1997). Perfectly balanced bifurcating trees can only arise under very
456 narrowly constrained circumstances, so phylogenetic imbalance, or a skew distribution in the
457 size of daughter clades, is the expected condition and arises from random splitting in birth-
458 death models (Nee, 2006). Metrics of tree imbalance examine nested clades; real applied
459 taxonomy and our synthetic taxonomy are not so restricted. Even though our simulated
460 genera are all monophyletic, the sister taxon of a genus is rarely another genus. Most
461 phylogenetic simulations differ from patterns observed in taxonomy in that the models
462 recover far fewer monotypic clades (Scotland & Sanderson, 2004). This is in contrast to our
463 compiled real-world datasets, which show a consistent proportion of monotypic genera, and
464 our simulations, which recover frequencies of monotypes that closely match real-world data
465 (fig. 4d).

466 Substantial previous research has explored genus-size, or more generally clade-size,
467 frequency distribution with simulation and modelling. In this context we differentiate
468 between what we term 'top down' and 'bottom up' approaches. 'Top down' includes any
469 model that directly generates the size or origination of higher taxa as units themselves. The

most direct ‘top down’ models have examined the patterns in real-world, empirical data for taxonomic classification, and then derived comparable mathematical descriptions that could be used to understand underlying evolutionary patterns (e.g. Yule, 1925; Maruvka et al., 2013). Others used phylogenetic simulations from branching processes with the origination of higher taxa embedded as a term included in the model, and examined the species richness of directly-generated genera or ‘paraclades’ (e.g. Patzkowsky, 1995), comparing simulation results with empirical data (Przeworski &Wall, 1998; Foote, 2012). A very few prior studies used a ‘bottom up’ approach (as we did herein), by which we mean that they first generated a simulated species phylogeny, and then applied classification. However, this approach previously was primarily used as a tool to examine cladogenesis and lineage origination over time (Sepkoski &Kendrick, 1993; Robeck, Maley & Donoghue 2000). Our novel ‘bottom up’ approach, or synthetic taxonomy, is the most direct approximation of the process of classifying living taxa in context of their evolutionary relationships.

Previous ‘top down’ models fitted to observed genus-size distributions produced closer matches to real-world data than we obtain here through artificial taxonomy, because that was their explicit aim (Maruvka et al., 2013). Other studies have also obtained good fits to empirical data with birth-death models that include direct simulation of higher taxa as cladogenic events (Foote, 2012). By contrast, our results come from a new bottom-up approach that compares ways that species might be partitioned into genera, given total knowledge of their phylogeny in simulation. This is an important distinction, because we are modelling the patterns of species origination, not controlling the origination of genera nor deriving a model to emulate their observed patterns.

Our approach was designed to address the central question of whether human-determined, historical taxonomy can be rationalised with phylogenetic patterns. While we had no *a priori* expectation that synthetic phylogenetically driven taxonomy should replicate

real-world data, there are clear similarities. None of the algorithms we used to recover simulated 'genera' were intended to closely mimic any taxonomic process. Rather we aimed to test the consistency of emergent patterns under several different idealised, monophyletic taxonomic definitions. We also used large sample sizes compared to real taxonomy, on the order of 10^8 simulated living species, compared to maximum global species estimates on the order of 10^7 (Mora et al., 2011; Scheffers et al., 2012; Stork et al., 2015). The observations and data discussed here represent large-scale emergent patterns in global biodiversity. In smaller sample sizes, the contingencies of either taxonomic history, or evolutionary history, could lead to the deviations that have previously been interpreted as evidence that the overall skew distributions are artefactual.

Skew distributions are common in natural systems, despite great variety in underlying mechanisms for sorting objects into frequency groups. Certain standard skew distributions approximately mimic the frequencies of genera of different sizes (Reed & Hughes, 2002), as well as patterns of word frequencies in language, or the sizes of corporations or cities (Reed & Jorgensen, 2004). Emergent global patterns in taxonomic diversity do not belie the many particular mechanisms that lead to the origination of large or small genera in particular clades. Large corporations are the minority of companies, but that does not mean that all large corporations are successful for the same reason(s). The same applies to the species richness of genera. Similarly, any particular explanation for the evolutionary dynamics in a particular group (a key adaptation, or contraction through extinction) may not undermine its role in a larger stochastic process. Smaller samples can easily find a pattern that appears to deviate from central tendency, which has previously caused some doubt about whether this skew distribution is artefactual (e.g. Strand & Panova, 2014). We contend that the repeated finding of nearly identical patterns in taxonomic datasets at varying scales (e.g. Yule, 1925; Holman, 1985; Mora et al., 2011; Maruvka et al., 2012; Strand & Panova, 2014; and herein)

520 is evidence that skew distribution in taxonomic size frequency is mathematically valuable.
521 The new insight afforded by our simulations it is that this is a realistic product of species
522 evolution.

523 The question of monophyly in real-world taxonomic data could influence patterns at
524 multiple levels. The frequency distribution of genus size does not change when restricted to
525 phylogenetically-defined clades; we selected 'real world' taxonomic datasets based on
526 taxonomic completeness and acceptance by relevant experts, and they include both
527 monophyletic clades (e.g. Aves), and non-monophyletic assemblages (marine invertebrates,
528 fish). Yet the overall frequency distributions appear similar. Within each dataset, most
529 genera are defined by morphology; most genus names pre-date molecular phylogenetics, and
530 the vast majority of species lack sequence data (Appeltans et al., 2012). Most genera and
531 families (especially in under-studied groups) have also not been tested for monophyly,
532 although the absence of a test does not imply that all will fail. But this pattern cannot be
533 blamed on 'lumping', 'splitting', or cryptic species complexes. Some genera included in our
534 datasets are undoubtedly paraphyletic, though previous simulations have shown this does not
535 necessarily affect overall patterns, at least when including extinct lineages (Sepkoski &
536 Kendrick, 1993). The emergence of a hollow-curve distribution in real-world taxonomic data
537 is not dependent on genera being monophyletic, yet it also emerges consistently from
538 simulations using strict monophyly.

539 Future generalisations about species diversity should account for the underlying
540 frequency distribution of genus size. In a strongly skewed distribution, central-tendency
541 measures such as the arithmetic mean are relatively uninformative. Many authors (e.g. Qian
542 & Ricklefs, 2000; Krug, Jablonski & Valentine, 2008; Mora et al., 2008; Foote, 2012) have
543 relied on a species-per-genus ratio, or used such a ratio as a proxy for maximum genus size.
544 While many authors have discussed or made adjustments for genus size distributions,

nonetheless this approach is equivalent to using an average of species-per-genus, and implicitly assumes an underlying normal distribution for genus size. Though authors may have a thorough understanding of the taxonomic patterns within their group or even the global patterns discussed here, it should be emphasised that taxonomic metadata are applied to many other fields of science. Other work has highlighted the potential pitfalls of extrapolations based on unsubstantiated assumptions of a universal species-per-genus ratio (e.g. Scheffers et al., 2012). The modal genus size is very likely to always be 1 (Aldous, 2001), and the mean is hence not a useful measure of central tendency in genus diversity. Future studies can expand on the present work to estimate diversity using a modelling approach for reconstructing species diversity from a more accurate generalised probability distribution for genus diversity.

Large genera

Evolutionary biology is intellectually focussed on large and rapidly evolving groups (Seehausen, 2006; Rabosky & Lovette, 2008; Losos et al., 1998; Thorpe & Losos, 2004). The 'success' of a genus is considered nearly synonymous with its species richness (Minelli, 2015). Indeed, a substantial proportion of species are included in large genera – in reptiles the five largest genera (*Anolis*, *Liolaemus*, *Cyrtodactylus*, *Atractus*, *Hemidactylus*) comprise slightly more than 10% of nominal reptile species, and the species in monotypic genera account for less than 10% of species in each of the taxonomic datasets included herein.

Among relatively under-studied groups, large genera are often 'bucket' taxa awaiting taxonomic revision, rather than interesting evolutionary phenomena. In our datasets, there are only five genera with more than 500 species (all marine invertebrates). Some have additional structure; the gastropod *Conus*, for example, was recently divided into 57 sub-genera (Puillandre et al., 2015). Flowering plants and fungi, not sampled here, contain some of the

largest eukaryotic genera with thousands of species (Minelli, 2015); these too often have recognised additional phylogenetic structure and are split into many subgenera. Among all groups, most very large genera appear to represent units that are not 'real' either in that they are non-monophyletic or not appropriate to the rank of genus.

In order to compare like with like, across a broad range of organisms, we considered that it was better use the taxonomic ranks assigned by experts rather than impose our own re-interpretation. For instance, some groups have sub-generic divisions that could arguably be the equivalent to the genera of other groups; we did not impose this equivalence as it would involve overturning the decision of experts as to what relevant level of distinctiveness is required to differentiate a genus in that group. It is interesting then that using a sampling of the current taxonomic *status quo* recovered consistent patterns of genus-size distribution across all the animal groups we investigated.

The main goal of our study was to determine whether taxonomic rank in general, but genera in particular, can predict species biodiversity; one immediate outcome is that our findings can be used to assess where biological groups may deviate from that null model. We suggest for example that this is further evidence to support critical re-examination of unusually large genera especially among marine invertebrates, and unusually high frequencies of small genera, such as in mammals.

Rates of evolution

There are real, predictable patterns in systematics, and the skew distribution in generic size occurs across variety of rate parameters and taxonomic algorithms. Our simulations deliberately used fixed rates of speciation and extinction to facilitate comparisons between rate parameters; this led to well-constrained behaviour in the resulting trees. There is

a clear mathematical behaviour to trees, influenced by speciation and extinction rates, which translates to mathematical behaviours of clades (Aldous et al., 2008).

Our taxonomic algorithms were also deliberately defined in an idealised way that is not realistically similar to practical taxonomy. Taxonomy almost always operates with limited data, inferring relationships based on key characters with established utility (whether molecular or morphological), as available for the specimens under study. In simulation, we have omniscient knowledge of the underlying phylogeny, so this provides a way to assess how constrained or variable genus size frequency would be, in comparing perfectly complete and accurate phylogenies under a range of evolutionary rates.

Our first approach to simulated taxonomy, RDT (fig. 4b), extends the phylogenetic species concept so that ranks are assigned based on the relative similarity of proximate monophyletic groupings (*sensu* Cracraft, 1983). The second approach, IDT (fig. 4c) is conceptually similar in that it separates clusters of taxa where they have diverged ancestrally for more than some fixed threshold of time. Fixed-depth taxonomy (FDT; fig. 4a), approximates the chronological approaches promoted by some authors, who advocate the use of divergence times to determine rank (Avise & Johns, 1999; Avise & Mitchell, 2007). It should be expected that FDT simulations would deviate from 'real world' taxonomy because this is not how taxa are defined in practice; however, it may be successfully applied *post hoc* to a well-resolved phylogeny (Holt & Jønsson, 2014). Lineage depth is of interest in delimiting taxonomic groups, but it is not information that is generally accessible or available for most species-level taxa (Ricotta et al., 2012). Age of origin is variable in different groups—a topological phenomenon that is explored in our other taxonomic algorithms—and information that is simply not known for many. This potential problem has been well known for decades (e.g. Hennig, 1979; Avise & Liu, 2011). Our simulations demonstrate that the FDT approach is highly sensitive to permutations of speciation and extinction rates (fig. 4c),

whereas ‘real world’ taxonomy is evidentially not, at comparable sampling magnitudes. Small changes in evolutionary rates caused the FDT and IDT simulations to shift away from biologically realistic distributions. More importantly perhaps, different depth (age) thresholds actually had relatively less impact on the resulting frequency distributions. This sensitivity illustrates a significant weakness in using time of origin as a criterion for defining higher taxa.

Under the RDT model, varying rate parameters had very limited impact on frequency distributions, even less variable than in the real-world data. While RDT is also not intended to mimic genuine taxonomic practice, this pattern demonstrates that similarly shaped distributions can arise directly from different evolutionary scenarios, which is undoubtedly the case in comparing groups of real organisms. This method still uses branch lengths as well as topology to define genera (Barraclough & Humphreys, 2015), yet recovers rather different frequency distributions. The large number of bitypic genera recovered by RDT is an artefact reflecting the effects of forcing the classification to seek sister-relationships even when those taxa may be separated by deep divergences. In real species, characterised by genetic or morphological characters, deeply-separated sister taxa would probably not be considered a bitypic genus but rather two monotypic genera.

The three taxonomic algorithms we used to classify our simulated trees, usually recovered genera that had smaller maximum sizes than in ‘real world’ data. Large genera in some cases reflect the existence of ‘bucket’ para- or polyphletic genera in real-world taxonomy; these are never present in our simulations. Other very large genera in the real world are undoubtedly monophyletic and may be already subdivided into subgenera, which may in fact be more equivalent to the genus rank in other clades (e.g. the mollusc genus *Conus*, noted above). More likely, large genera may be absent in the simulations because the

643 model did not allow for synergistic effects of speciation rates and environment, which are
644 thought to underpin rapid radiations (Harmon & Harrison, 2015).

645 It is increasingly well understood that both speciation rates and extinction rates vary
646 among clades and even within clades over time (Marshall, 2017), although these rates may be
647 approximately equal (zero net diversification) across all clades over time (Ricklefs, 2007) or
648 with a narrow tendency for globally increasing diversity (Bennett, 2013). The convergence of
649 genus size-frequency distributions under our various models, and the similar convergence in
650 real-world taxonomic data, suggest that there is perhaps a long term equilibrium in
651 evolutionary rates. Recent work has highlighted the potential heritability of speciation as a
652 trait itself (e.g. Purvis et al., 2011; Rabosky & Goldberg, 2015). The constrained sizes of the
653 largest genera recovered from our simulations with fixed speciation rates provides strong
654 additional evidence that heterogeneous rates of speciation are fundamental to the origination
655 of large genera.

656 There are two significant hurdles that have been raised as potentially impeding the use
657 of higher-ranked taxa to measure species diversity: First, whether the units (genera) are
658 defined by consistent criteria that make them comparable across different groups, and second,
659 whether the genera are monophyletic units (Hendricks et al., 2014). Our simulations
660 addressed these issues by using strict algorithms to define monophyletic genera. Applying
661 these criteria highlighted the variability introduced by changing evolutionary rates, and also
662 illustrated the comparatively constrained range of distributions found in real world taxonomy.
663

664 *Conclusions*

665 Mathematical approaches are important tools to separate real excursions in speciation
666 rates, that might require special explanations, from patterns that can be predicted within a
667 well-described probability distribution. If we begin with a premise that large genera represent

668 evolutionary anomalies, then it is logical to seek an explanation for the process that generated
669 that excursion. However, as we demonstrate here, taxonomic genera arise from phylogeny in
670 a probability space that accommodates both small and large genera, with decreasing
671 frequency as genera get larger. From these simulations, one could infer that genera of sizes
672 up to around 50 species are not exceptional, genera of several hundred species are unusual
673 and perhaps deserve taxonomic scrutiny, and certainly monotypic genera are commonplace.
674 Special adaptive significance is not necessarily required to explain a monotypic genus, or a
675 large genus, or a genus with four species.
676 Our results provide novel evidence that Linnean ranks applied to groups of species
677 can have transferable meaning between unrelated clades, even though monotypic units of
678 classification are not equivalent to topological nested clades. Genus sizes should follow a
679 skew-distribution; monotypic genera are *expected* to be very common, and large genera are
680 *expected* to be very rare. The largest genera, of sizes that dramatically exceed anything
681 recovered in simulation, are probably not appropriate phylogenetic or systematic units.
682 Understanding the frequency distribution of supraspecific taxa, and their behaviour as
683 mathematical units, is crucial to a more robust understanding of taxonomic surrogacy. It is
684 essential to know how diversity, when measured in terms of genera or families, can be
685 translated into species richness. The skewed distribution of genus sizes, which is a real
686 phenomenon, precludes using a simple count of genera or higher ranked taxa to answer many
687 questions about comparative species diversity. The present study provides a foundation for a
688 new approach to quantify the error introduced by taxonomic surrogacy. Our results
689 demonstrate for the first time that determining this is an achievable target, and that
690 established systematics already holds the key to robust quantitative analyses of global
691 diversity.
692

693 **References**
694
695 Aldous D, Krikun M, Popovic L. 2008. Stochastic models for phylogenetic trees on higher-
696 order taxa. Journal of Mathetmatical Biology 56: 525–557.
697 Aldous DJ, Krikun MA, Popovic L. 2011. Five statistical questions about the tree of life.
698 Systematic Biology 60:318-28.
699 Aldous DJ. 2001. Stochastic models and descriptive statistics for phylogenetic trees, from
700 Yule to today. Statistical Science 16: 23–23.
701 Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G,
702 Harmon LJ. 2009. Nine exceptional radiations plus high turnover explain species
703 diversity in jawed vertebrates. Proceedings of the National Academy of Sciences,
704 USA 106: 13410–13414.
705 Alroy J, Marshall CR, Bambach RK, Bezusko K, Foote M, Fürsich FT, Hansen TA, Holland
706 SM, Ivany LC, Jablonski D, Jacobs DK, Jones DC, Kosnik MA, Lidgard S, Low S,
707 Miller AI, Novack-Gottshall PM, Olszewski TD, Patzkowsky ME, Raup DM, Roy K,
708 Sepkoski JJ, Sommers MG, Wagner PJ, Webber A. 2001. Effects of sampling
709 standardization on estimates of Phanerozoic marine diversification. Proceedings of the
710 National Academy of Sciences of the USA 98:6261–6266.
711 Alroy J, Aberhan M, Bottjer DJ, Foote M, Fürsich FT, Harries PJ, Hendy AJ, Holland SM,
712 Ivany LC, Kiessling W, Kosnik MA. 2008. Phanerozoic trends in the global diversity
713 of marine invertebrates. Science 321: 97–100.
714 Anderson S. 1974. Patterns of faunal evolution. The Quarterly Review of Biology 49: 311–
715 332.
716 Appeltans W, Ahyong ST, Anderson G, Angel MV, Artois T, Bailly N, Bamber R, Barber A,
717 Bartsch I, Berta A, Błażewicz-Paszkowycz M, Bock P, Boyko CB, Brandão SN, Bray
718 RA, Bruce NL, Cairns SD, Chan T-Y, Cheng L, Collins AG, Cribb T, Curini-Galletti

719 M, Dahdouh-Guebas F, Davie PJF, Dawson MN, De Clerck O, Decock W, De Grave
720 S, de Voogd NJ, Domning DP, Emig CC, Erséus C, Eschmeyer W, Fauchald K,
721 Fautin DG, Feist SW, Fransen CHJM, Furuya H, Garci-Alvarez O, Gerken S, Gibson
722 D, Gittenberger A, Gofas S, Gómez-Daglio L, Gordon D, Guiry MD, Hernandez F,
723 Hoeksema BW, Hopcroft RR, Jaume D, Kirk P, Koedam N, Koenemann S, Kolb JB,
724 Kristensen RM, Kroh A, Lambert G, Lazarus D, Lemaitre R, Longshaw M, Lowry J,
725 Macpherson E, Madin LP, Mah C, Mapstone G, McLaughlin PA, Mees J, Meland K,
726 Messing CG, Mills CE, Molodtsova TN, Mooi R, Neuhaus B, Ng PKL, Nielsen C,
727 Norenburg J, Opresko DM, Osawa M, Paulay G, Perrin W, Pilger JF, Poore GCB,
728 Pugh P, Read GB, Reimer JD, Rius M, Rocha RM, Saiz-Salinas JI, Scarabino V,
729 Schierwater B, Schmidt-Rhaesa A, Schnabel KE, Schotte M, Schuchert P, Schwabe E,
730 Segers H, Self-Sullivan C, Shenkar N, Siegel V, Sterrer W, Stohr S, Swalla B, Tasker
731 ML, Thuesen EV, Timm T, Todaro MA, Turon X, Tyler S, Uetz P, van der Land J,
732 Vanhoorne B, van Ofwegen LP, van Soest RWM, Vanaverbeke J, Walker-Smith G,
733 Walter TC, Warren A, Williams CG, Wlson SP, Costello MJ. 2012. The magnitude of
734 global marine species diversity. Current Biology 22: 2189–2202.

735 Avise JC, Johns GC. 1999. Proposal for a standardized temporal scheme of biological
736 classification for extant species. Proceedings of the National Academy of Sciences,
737 USA 96: 7358–7363.

738 Avise JC, Liu J-X. 2011. On the temporal inconsistencies of Linnean taxonomic ranks.
739 Biological Journal of the Linnean Society 102:707–714.

740 Avise JC, Mitchell D. 2007. Time to standardize taxonomies. Systematic Biology 56: 130–
741 133.

742 Barraclough TG, Humphreys AM. 2015. The evolutionary reality of species and higher taxa
743 in plants: a survey of post-modern opinion and evidence. New Phytologist 207: 291–
744 296.
745 Bass D, Richards TA. 2012. Three reasons to re-evaluate fungal diversity ‘on Earth and in the
746 ocean’. Fungal Biology Reviews 25: 159–164.
747 Bennett, KD. 2013. Is the number of species on Earth increasing or decreasing? Time, chaos
748 and the origin of species. Palaeontology 56: 1305–1325.
749 Bertrand Y, Pleijel F, Rouse GW. 2006. Taxonomic surrogacy in biodiversity assessments,
750 and the meaning of Linnaean ranks. Systematic Biodiversity 4: 149–159.
751 Bond JE, Opell BD. 1998. Testing adaptive radiation and key innovation hypotheses in
752 spiders. Evolution 52: 403–414.
753 Boxshall GA, Mees J, Costello MJ, and 253 other authors. 2014. World Register of Marine
754 Species. Available from http://www.marinespecies.org at VLIZ (accessed October
755 2014).
756 Budd GE, Jackson ISC. 2016. Ecological innovations in the Cambrian and the origins of the
757 crown group phyla. Philosophical Transactions of the Royal Society B 371:
758 20150287.
759 Costello MJ, May RM, Stork NE. 2013. Can we name Earth’s species before they go extinct?
760 Science 339: 413–416.
761 Cracraft J. 1983. Species concepts and speciation analysis. Current Ornithology 1: 159–187.
762 Ellis D. 1985. Taxonomic sufficiency in pollution assessment. Marine Pollution Bulletin 16:
763 459.
764 Fenner M, Lee WG, Wilson JB. 1997. A comparative study of the distribution of genus size
765 in twenty angiosperm floras. Biological Journal of the Linnean Society 62: 225–237.
766 Foote M. 2012. Evolutionary dynamics of taxonomic structure. Biology Letters 8: 135–138.

767 Froese R, Pauly D (editors). 2015. FishBase. www.fishbase.org (accessed August 2015).
768 Gaston KJ, Williams PH. 1993. Mapping the world's species-the higher taxon approach.
769 Biodiversity Letters 1: 2–8.
770 Gift N, Stevens PF. 1997. Vagaries in the delimitation of character states in quantitative
771 variation—an experimental study. Systematic Biology 46: 112–125
772 Gill F, Donsker D (editors). IOC World Bird List (v 4.4). doi 10.14344/IOC.ML.4.4; 2014.
773 Giribet G, Hormiga G, Edgecombe GD. 2016. The meaning of categorical ranks in
774 evolutionary biology. Organisms Diversity and Evolution 16: 427–430.
775 Harmon LJ, Harrison S. 2015. Species diversity is dynamic and unbounded at local and
776 continental scales. American Naturalist 185: 584-593.
777 Heim NA, Peters SE. 2011. Regional environmental breadth predicts geographic range and
778 longevity in fossil marine genera. PLoS One 6:1–12.
779 Heino J. 2014. Taxonomic surrogacy, numerical resolution and responses of stream
780 macroinvertebrate communities to ecological gradients: Are the inferences
781 transferable among regions? Ecological Indicators 36: 186–194.
782 Hendricks JR, Saupe EE, Myers CE, Hermsen EJ, Allmon WD. 2014. The generification of
783 the fossil record. Paleobiology 40: 511–528.
784 Hennig W. 1979. *Phylogenetic Systematics.* (reprinted) Urbana, USA: University of Illinois
785 Press.
786 Holman EW. 1985. Evolutionary and psychological effects in pre-evolutionary
787 classifications. Journal of Classification 2: 29–39.
788 Holt BG, Jønsson KA. 2014. Reconciling hierarchical taxonomy with molecular phylogenies.
789 Systematic Biology 63: 1010–1017.
790 Hopwood AT. 1959. The development of pre-Linnaean taxonomy. Proceedings of the
791 Linnean Society, London 170: 230–234.

792 Huelsenbeck JP, Lander KM. 2003. Frequent inconsistency of parsimony under a simple
793 model of cladogenesis. Systematic Biology 52: 641–648.
794 Humphreys AM, Barraclough TG. 2014. The evolutionary reality of higher taxa in mammals.
795 Proceedings of the Royal Society B 281: 20132750.
796 Jablonski D, Finarelli JA. 2009. Congruence of morphologically-defined genera with
797 molecular phylogenies. Proceedings of the National Academy of Sciences, USA. 106:
798 8262–8266.
799 Kendall DG. 1948. On the generalized "birth-and-death" process. Annals of Mathematical
800 Statistics 19: 1–15.
801 Krug AZ, Jablonski D, Valentine JW. 2008. Species-genus ratios reflect a global history of
802 diversification and range expansion in marine bivalves. Proceedings of the Royal
803 Society B 275: 1117–1123.
804 Lee MSY. 2003. Species concepts and species reality: salvaging a Linnaean rank. Journal of
805 Evolutionary Biology 16: 179–188.
806 Losos JB, Jackman TR, Larson A, de Queiroz K, Rodríguez-Schettino L. 1998. Contingency
807 and determinism in replicated adaptive radiations of island lizards. Science 279:
808 2115–2118.
809 Lu PJ, Yogo M, Marshall CR. 2006. Phanerozoic marine biodiversity dynamics in light of the
810 incompleteness of the fossil record. Proceedings of the National Academy of
811 Sciences, USA 103: 2736–2739.
812 Marshall CR, 2017. Five palaeobiological laws needed to understand the evolution of the
813 living biota. Nature Ecology & Evolution 1: s41559-017.
814 Maruvka YE, Shnerb NM, Kessler DA, Ricklefs RE. 2013. Model for macroevolutionary
815 dynamics. Proceedings of the National Academy of Sciences, USA 110: E2460–
816 E2469

817 Mayr E. 1982. The Growth of Biological Thought: Diversity, Evolution, and Inheritance.
818 Cambridge USA: Belknap Press of Harvard University Press.
819 Minelli A. 2015. Species diversity vs. morphological disparity in the light of evolutionary
820 developmental biology. Annals of Botany 117: 781-794.
821 Mooers AO, Heard SB. 1997. Inferring evolutionary process from phylogenetic tree shape.
822 Quarterly Review of Biology 72: 31–54.
823 Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B. 2011. How many species are there on
824 Earth and in the ocean? PLoS Biology 9: e1001127
825 Nee S. 2006. Birth-death models in macroevolution. Annual Review of Ecology, Evolution
826 and Systematics. 37: 1–17.
827 Patzkowsky ME. 1995. A hierarchical branching model of evolutionary radiations.
828 Paleobiology 21: 440–460.
829 Puillandre N, Duda TF, Meyer C, Olivera M, Bouchet P. 2015. One, four or 100 genera? A
830 new classification of the cone snails. Journal of Molluscan Studies 81: 1–23.
831 Purvis A, Fritz S, Rodríguez J, Harvey PH, Grenyer R. 2011. The shape of mammalian
832 phylogeny: patterns, processes and scales. Philosophical Transactions of the Royal
833 Society B 366: 2462–77.
834 Przeworski M, Wall JD. 1998. An evaluation of a hierarchical branching process as a model
835 for species diversification. Paleobiology 24: 498–511.
836 Qian H, Ricklefs RE. 2000. Large-scale processes and the Asian bias in species diversity of
837 temperate plants. Nature 66: 180–182.
838 Quental TB, Marshall CR. 2010. Diversity dynamics: molecular phylogenies need the fossil
839 record. Trends in Ecology and Evolution 25: 434–441.
840 Rabosky DL, Goldberg, EE. 2015. Model inadequacy and mistaken inferences of trait-
841 dependent speciation. Systematic Biology 2015;64, 340–355.

842 Rabosky DL, Lovette IJ. Explosive evolutionary radiations: Decreasing speciation or
843 increasing extinction through time? Evolution 62: 1866–1875.
844 Rannala B, Huelsenbeck JP, Yang Z, Nielsen R. 1998. Taxon sampling and the accuracy of
845 large phylogenies. Systematic Biology 47: 702–710.
846 Raup DM. 1978. Cohort analysis of generic survivorship. Paleobiology 4: 1–15.
847 Raup DM. 1985. Mathematical models of cladogenesis. Paleobiology 11: 42–52.
848 Raup DM, Boyajian GE. 1988. Patterns of generic extinction in the fossil record.
849 Paleobiology 14: 109–25.
850 Raup DM, Gould SJ, Schopf TJM, Simberloff DS. 1973. Stochastic models of phylogeny and
851 the evolution of diversity. Journal of Geology 81: 525–542.
852 Raup DM, Gould SJ. 1974. Stochastic simulation and evolution of morphology – towards a
853 nomothetic paleontology. Systematic Zoology 23: 305–322.
854 Raup DM, Sepkoski Jr, JJ. 1986. Periodic extinction of families and genera. Science 231:
855 833–836.
856 Reed WJ, Hughes BD. 2002. From gene families and genera to incomes and internet file
857 sizes: Why power laws are so common in nature. Physical Review E 66: 067103.
858 Reed WJ, Jorgensen M. 2004. The double Pareto-lognormal distribution — a new parametric
859 model for size distributions. Communications in Statistics—Theory and Methods 33:
860 1733–1753.
861 Ricklefs RE. 2007. Estimating diversification rates from phylogenetic information. Trends in
862 Ecology and Evolution 22: 601–610.
863 Ricotta C, Bacaro G, Marignani M, Godefroid S, Mazzoleni S. 2012. Computing diversity
864 from dated phylogenies and taxonomic hierarchies: does it make a difference to the
865 conclusions? Oecologia 170: 501–506

866 Ricotta C, Ferrari M, Avena G. 2002. Using the scaling behaviour of higher taxa for the
867 assessment of species richness. Biological Conservation 107: 131–133.
868 Robeck HE, Maley CC, Donoghue MJ. 2000. Taxonomy and temporal diversity patterns.
869 Paleobiology 26: 171–187.
870 Scheffers BR, Joppa LN, Pimm SL, Laurance WF. 2012. What we know and don't know
871 about Earth's missing biodiversity. Trends in Ecology and Evolution 27: 501–510.
872 Schorr D, Paulson D (editors). 2014. World Odonata List v 57.
873 http://www.pugetsound.edu/academics/academic-resources/slater-
874 museum/biodiversity-resources/dragonflies/world-odonata-list2/ (accessed October
875 2014).
876 Scotland RW, Sanderson MJ. 2004. The significance of few versus many in the tree of life.
877 Science 303: 643–644.
878 Seehausen O. 2006. African cichlid fish: a model system in adaptive radiation research.
879 Proceedings of the Royal Society B 273: 1987–1998.
880 Sepkoski, D. 2012. Rereading the fossil record. The growth of paleobiology as an
881 evolutionary discipline. Chicago USA: University of Chicago Press.
882 Sepkoski JJ, Kendrick DC. 1993. Numerical experiments with model monophyletic and
883 paraphyletic taxa. Paleobiology 19: 168–184.
884 Stork NE, McBroom J, Gely C, Hamilton AJ. 2015. New approaches narrow global species
885 estimates for beetles, insects, and terrestrial arthropods. Proceedings of the National
886 Academy of Sciences, USA. 112: 7519–7523.
887 Strand M, Panova M. 2014. Size of genera – biology or taxonomy? Zoologica Scripta 44,
888 106–116.

889 Symul T, Assad SM, Lam PK. 2011. Real time demonstration of high bitrate quantum
890 random number generation with coherent laser light. Applied Physics Letters 98:
891 231103. doi: 10.1063/1.3597793.
892 Thorpe RS, Losos JB. 2004. Evolutionary diversification of Caribbean *Anolis* lizards:
893 concluding comments. In: Dieckmann U, Doebeli M, Metz JAJ, Tautz D, eds.
894 *Adaptive Speciation*, Cambridge UK: Cambridge University Press, 322–344.
895 Timms LL, Bowden JJ, Summerville KS, Buddle CM. 2013. Does species-level resolution
896 matter? Taxonomic sufficiency in terrestrial arthropod biodiversity studies. Insect
897 Conservation and Diversity 6: 453–462.
898 Triantis KA, Rigal F, Parent CE, Cameron RA, Lenzner B, Parmakelis A, Yeung NW,
899 Alonso MR, Ibáñez M, de Frias Martins AM, Teixeira DN. 2016. Discordance
900 between morphological and taxonomic diversity: land snails of oceanic archipelagos.
901 Journal of Biogeography 43: 2050-2061.
902 Uetz P, Hošek J (editors). 2015. The Reptile Database. http://www.reptile-database.org
903 (accessed August 2015).
904 Watson HW. 1875. On the probability of the extinction of families. Journal of the
905 Anthropological Institute of Great Britain and Ireland 4: 138–44.
906 Wilson DE, Reeder DM (editors). 2005. Mammal Species of the World. A Taxonomic and
907 Geographic Reference (3rd ed). Baltimore USA: Johns Hopkins University Press.
908 Yule GU. 1925. A mathematical theory of evolution, based on the conclusions of Dr. J. C.
909 Willis, F.R.S. Philosophical Transactions of the Royal Society London B 213: 21–87.
910
911

Figures Captions

Fig 1. The probability space of birth-death models that generate simulated phylogenies, for rates of speciation (λ, horizontal axis) and extinction (μ, vertical axis), illustrating the main emergent properties of the model. The **probability of eventual total extinction** of the descendant clade is relative to the ratio μ/λ, the slope within this space illustrated with varying shades of grey from guaranteed extinction (μ/λ >1, black) to increasing probability of clade persistence (paler wedges correspond to ratios indicated on right vertical axis). The average number of living **descendant species** at a fixed sampling time point (*t*) is relative to the difference λ-μ, visualised as the negative intercept of a line with slope 1, and increases exponentially as $e^{t\cdot(\lambda-\mu)}$. Thus when λ-μ = 0.01, at t = 400, simulations produce an average of 55 species; a small increase to λ-μ = 0.02 would result in 3000 species per tree in the same timeframe. The parameters selected for simulations herein (coloured circles) were chosen to represent a span of model behaviours with consistent average clade size, but varying clade-extinction probabilities (shades of grey in background).

Fig 2. Schematic representation of three independent taxonomic algorithms, applied to sort simulated species trees into monophyletic genus units.In *Relative-Distance Taxonomy*, tips (species) that are relatively closer to each other than to the previous common ancestor are united in a genus. Here, the threshold is 0.5 or 50% of the relative depth. The depth between node a_1 and b_1 is more than 0.5 the depth from b_1 to its alternate descendant. Thus the two descendent lines from b_1 are split into two genera. *Internal-Depth Taxonomy* separates monophyletic of clades of tips wherever an inter-nodal distance exceeds a given threshold (paraphyletic clusters are divided into monphyletic genera). *Fixed-Depth Taxonomy* defines genera to be the monophyletic groups of descendants of nodes after a given depth threshold.

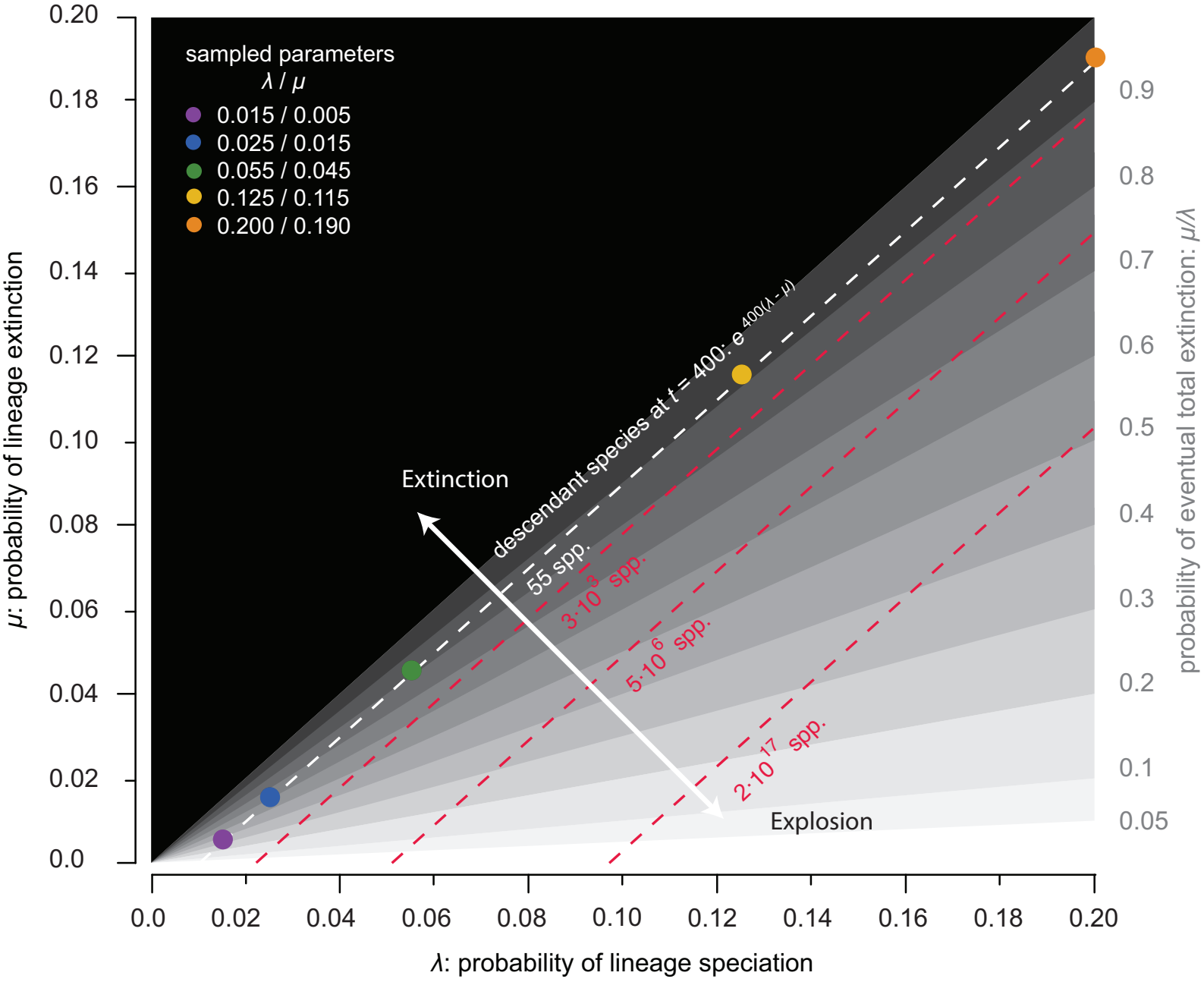
Fig 3. Size-frequency of genera in real world taxonomic data: the percentage of genera containing a set number of valid nominal species, summarised from global datasets for select groups.

Fig. 4. Size-frequency of genera in synthetic taxonomy derived from simulated data, using five parameter sets for rates of speciation (λ) and extinction (μ), shown in different colours; the size-frequency distribution of the total 'real world' dataset is included for comparison (summed from data shown in fig. 3). In each panel, solid and dotted lines indicate different thresholds for the algorithms that define synthetic genera. A) genera defined by *Relative-Difference Taxonomy*, with a threshold of 50% difference in depth (dotted lines) or 60% (solid lines). B) genera defined by *Internal-Depth Taxonomy*: defined by monophyletic clades of tips (species) within 20 generations (5% of tree depth, solid lines) or 40 generations (dotted lines) from any adjacent tips. C) genera defined by *Fixed-Depth Taxonomy*: defined by monophyletic clades of tips (species) within 20 generations (5% of tree depth, solid lines) or 40 generations (dotted lines) from the most recent common ancestor. D) frequency distributions for each algorithm, summed over all speciation and extinction rate parameters (showing six different datasets from simulations, grey, and real-world taxonomic data, black; symbols, and dotted and dashed lines correspond to algorithm thresholds as in other parts).

Table

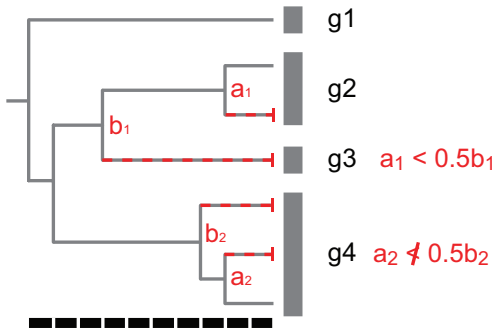
Table 1. Summary information for valid, and taxonomically accepted, non-extinct species and genera compiled from comprehensive global taxonomic datasets.

	Mammals	Marine invertebrates	Birds	Reptiles	Fish	Dragonflies	**Total**
Number of **species**	5,492	214,417	10,695	10,178	32,324	6,043	279,149
Number of **genera**	1,242	29,316	2,278	1,176	4,914	688	39,614
Maximum genus size	173	1,028	87	398	291	147	1,028
Number of monotypic genera	538	10,970	903	329	1,704	195	14,639
Species in monotypic genera	9.8%	5.1%	8.4%	3.2%	5.3%	3.2%	5.2%
Proportion of genera monotypic	43.3%	37.4%	39.6%	28.0%	34.7%	28.3%	37.0%


sampled parameters
λ / μ
0.015 / 0.005
0.025 / 0.015
0.055 / 0.045
0.125 / 0.115
0.200 / 0.190
descendant species at t = 400: e^400(λ - μ)
55 spp.
3·10^3 spp.
5·10^6 spp.
2·10^17 spp.
Extinction
Explosion
μ: probability of lineage extinction
λ: probability of lineage speciation
probability of eventual total extinction: μ/λ
0.0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
0.05
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

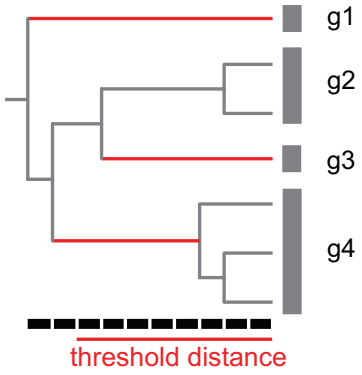
Relative-Distance Taxonomy

split where relative depth of adjacent nodes exceed threshold proportion (e.g. $a_i < 0.5b_i$)



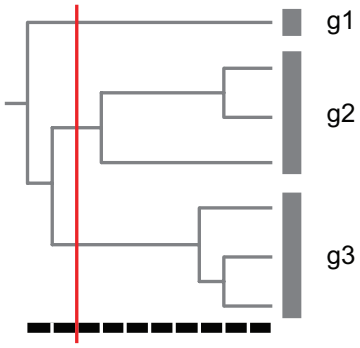
Internal-Depth Taxonomy

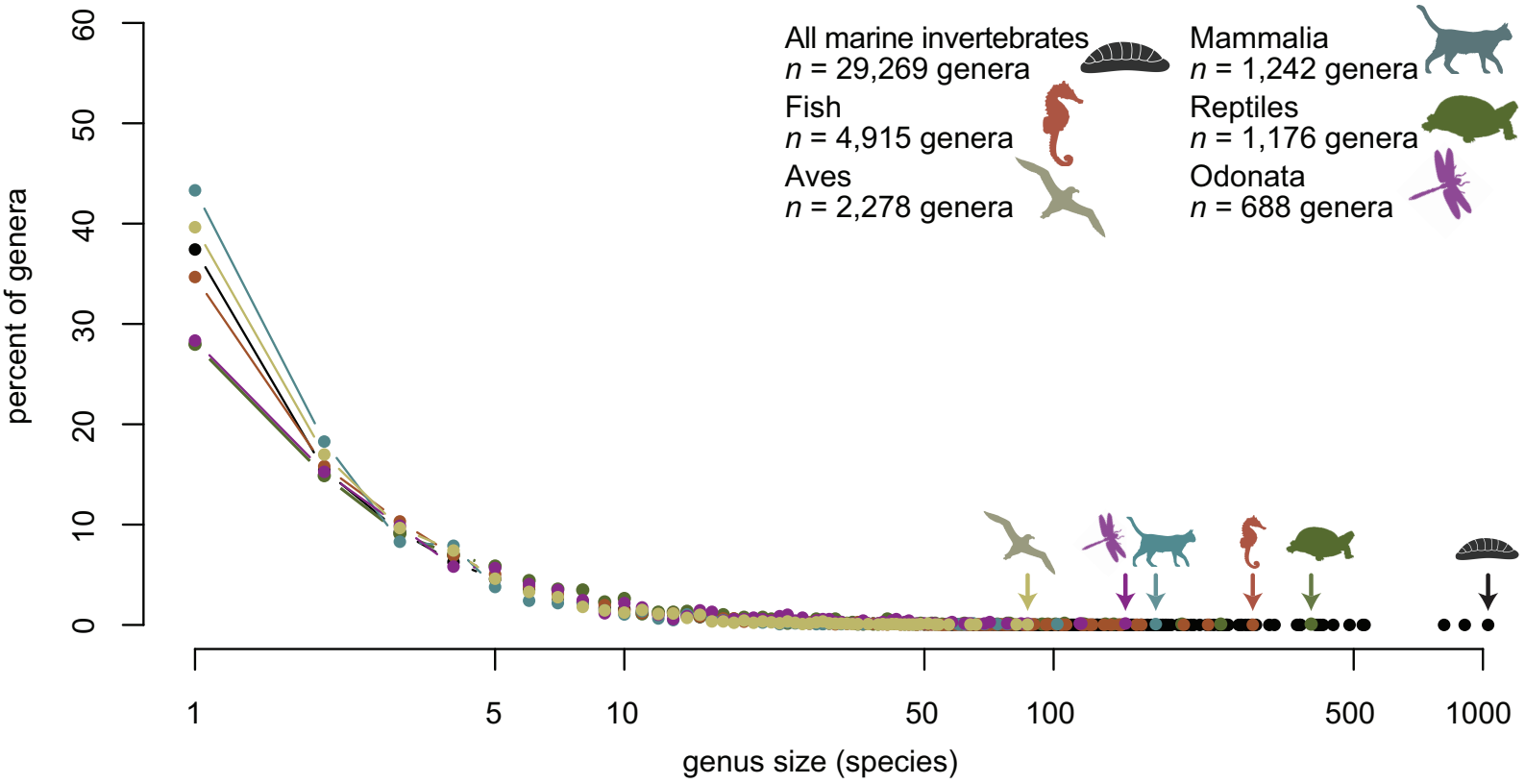
monophyletic descendents after inter-nodal distance exceeds threshold (% of tree length)

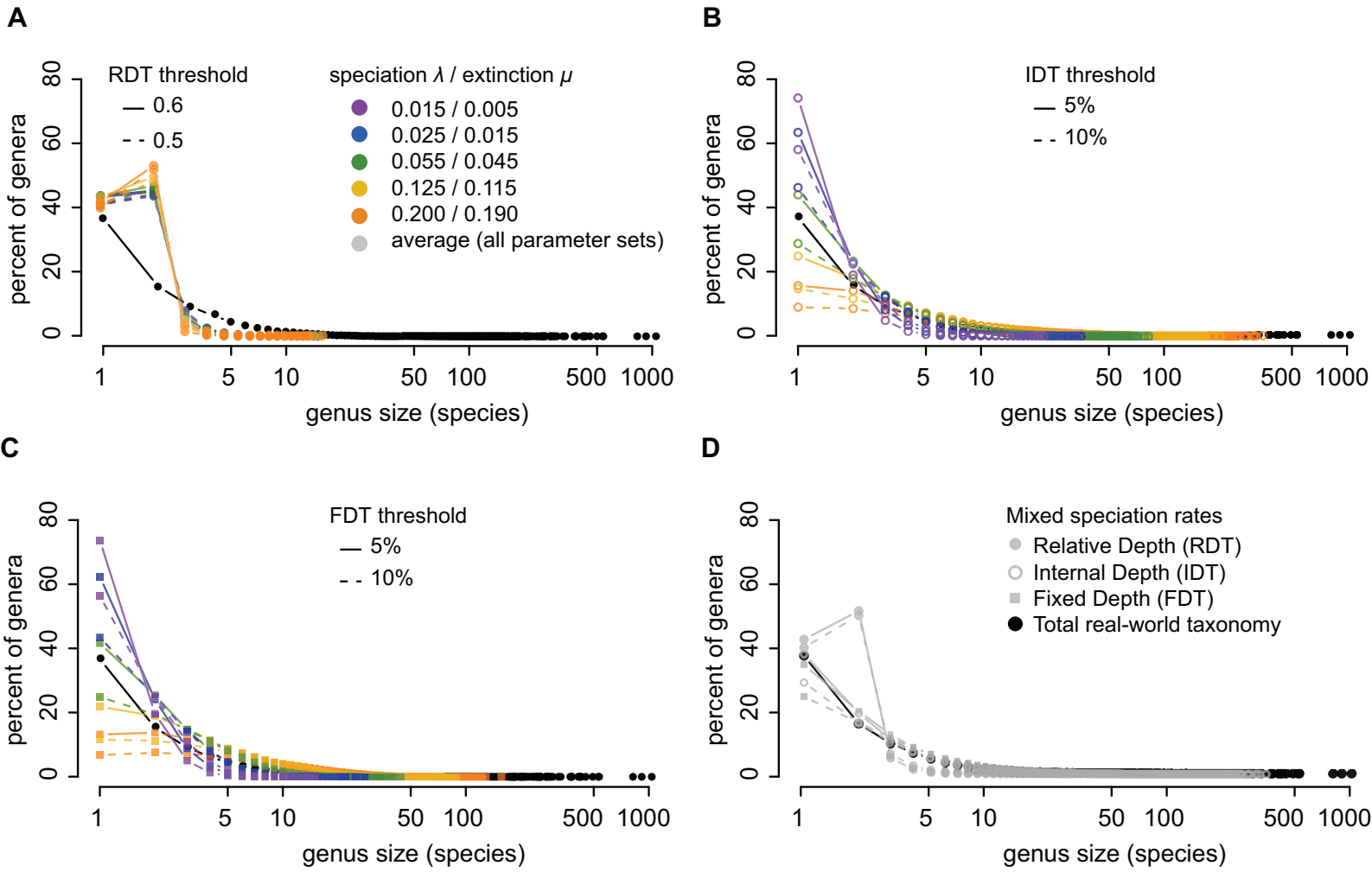


Fixed-Depth Taxonomy

clades originating after threshold (% of tree length)




All marine invertebrates
n = 29,269 genera
Fish
n = 4,915 genera
Aves
n = 2,278 genera
Mammalia
n = 1,242 genera
Reptiles
n = 1,176 genera
Odonata
n = 688 genera
percent of genera
0
10
20
30
40
50
60
1
5
10
50
100
500
1000
genus size (species)


A
RDT threshold
— 0.6
- - 0.5
speciation λ / extinction μ
0.015 / 0.005
0.025 / 0.015
0.055 / 0.045
0.125 / 0.115
0.200 / 0.190
average (all parameter sets)
percent of genera
0
20
40
60
80
1
5
10
50
100
500
1000
genus size (species)
B
IDT threshold
— 5%
- - 10%
percent of genera
genus size (species)
C
FDT threshold
— 5%
- - 10%
percent of genera
genus size (species)
D
Mixed speciation rates
Relative Depth (RDT)
Internal Depth (IDT)
Fixed Depth (FDT)
Total real-world taxonomy
percent of genera
genus size (species)