Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Phylogenetic Diversity Across the Complete Tree of Life

James Rosindell

j.rosindell@imperial.ac.uk

Imperial College London https://orcid.org/0000-0002-5060-9346

Jialiang Guo

Peking University

Emily Jane McTavish

University of California, Merced

Article

Keywords: Conservation, Evolutionary Distinctiveness, Open Tree of Life, Phylogenetic Diversity

Posted Date: May 20th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4397985/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Phylogenetic Diversity Across the Complete Tree of Life

Jialiang Guo, Emily Jane McTavish and James Rosindell

Abstract

In the face of rapid biodiversity loss, many approaches have been developed to measure biodiversity in ways that go beyond species richness. One prominent example is Phylogenetic Diversity (PD), which measures evolutionary history by summing the branches required to connect a set of species on a dated phylogenetic tree. PD may also capture other biodiversity measures by proxy such as the richness of biological features and their potential future benefits for humanity, sometimes known as 'future options'. The total global PD is known for some well-studied groups, such as most vertebrates, but PD estimates are lacking for the majority of the tree of life. Here, we characterize the distribution of PD across the complete tree of life with over 2.2 million species. To do this we use data from the Open Tree of Life and a smoothing method to interpolate between nodes without date information. We estimate that the PD represented by all described species together is between 29 and 33 trillion years. We characterize the distribution of evolutionary distinctiveness, a measure of the fair share of PD captured by individual species, across all life and within selected clades. Many clades have bimodal distributions of evolutionary distinctiveness across species which may be due to changes in diversification rate within subclades. PD has previously been used as the basis for conservation prioritization schemes such as EDGE (Evolutionary Distinct and Globally Endangered) which synthesizes phylogenetic tree data with extinction risk data from the IUCN Red List of threatened species. Here we estimate EDGE scores for over 130,000 species, many more than have been done previously. The top EDGE species is *Latimeria chalumnae*, the critically endangered West Indian Ocean coelacanth. We hope this work will pave the way for more complete and automated analyses of PD and EDGE scores across the complete tree of life.

Keywords

Conservation, Evolutionary Distinctiveness, Open Tree of Life, Phylogenetic Diversity

1. Introduction.

We are currently experiencing a massive decline in global biodiversity. Anthropogenic activities such as livestock grazing, urbanization, and emission of greenhouse gasses are reshaping the land, driving what is being described as the sixth mass extinction on our planet[1–3]. This biodiversity decline threatens the stability of many ecosystem services which are crucial for the quality of human life now and in the future[4]. Biodiversity refers broadly to the variety of life on earth. There are over 2.2 million described species of life on earth, but species richness is not the only dimension of biodiversity. Other measures of biodiversity such as trait and functional diversity use details about individual species that are currently unknown for the vast majority of described species. A complete tree of life (phylogeny) describing the evolutionary connections between species is available however, and could be used to measure deeper aspects of biodiversity at a global scale that are not captured by species richness.

The most commonly used measure of biodiversity based on data from the tree of life is Phylogenetic Diversity (PD), which measures the evolutionary history of a group of species. PD is calculated by summing the phylogenetic branch lengths required to span a given group of species[5]. This biodiversity measure can be regarded as a proxy for other biodiversity measures, especially feature diversity (the diversity of biological features or traits in the most general sense that can be detected in their structure or biochemistry). PD may potentially also capture "future options for humanity": features of species with beneficial utility for humans, which are currently unknown, but which may manifest in the future[6–9] .

The importance of preserving PD has been increasingly recognized, the Intergovernmental Science-policy Platform for Biodiversity and Ecosystem Services (IPBES) has adopted PD as an indicator of the overall capacity of biodiversity to support a good quality of life into the future[6]. This indicator, along with the related 'EDGE' (Evolutionary Distinct and Globally

Endangered) Index[10], have been included as indicators for the Convention on Biological Diversity's draft post-2020 Global Biodiversity Framework[11]. The IUCN has also established a Phylogenetic Diversity Task Force to provide expertise on the inclusion of PD in conservation strategies for practitioners, decision-makers, and the public[12]. Another important measure related to PD is the Evolutionary Distinctiveness (ED) of a species, also known as 'fair proportion'[10,13]. This measure represents a fair share of the total PD of a given clade that is attributed to each individual one of its species. ED can therefore be used to measure the contribution made by a specific individual species to phylogenetic diversity of a larger clade[13]. In the calculation of ED for a given species on a phylogeny, we consider separately all its ancestral branches back to the root. For each branch, the branch length, which comprises a contribution to PD, is divided equally (fairly) between all its descendent extant species (tips). The total PD across the complete tree of life is equal to the sum of ED across all species. We can, therefore, use ED calculation or approximation as a way to find total PD through summation. However, most often, PD and ED are calculated together using a phylogenetic tree with branch lengths.

One application of PD and ED is in conservation prioritization. Many studies have used the idea of preserving PD in preference to metrics based only on species richness as a way to guide species-based conservation priorities. For example, the EDGE score of a species is calculated by combining its ED score and the risk of extinction provided by the IUCN Red List. The EDGE metric enables researchers to identify the most evolutionarily unique and threatened species as potential priorities[10,14,15] in the context of limited resources for practical conservation[16]. This ranking has been done for many clades including gymnosperms[14], chondrichthyes[17], reptiles[18], mammals and birds[10,19–24]. However, most of the tree of life has not been explored in this way.

Here we make use of data from the Open Tree of Life (OpenTree)[25] and OneZoom tree of life explorer[26,27] to provide a study of PD and EDGE scores across all described species of life. The OpenTree project builds and maintains a comprehensive tree of life by integrating published phylogenies with taxonomic information[28]. This synthetic tree is updated using automated pipelines[29]. The current OpenTree synthetic tree contains more than 2.3 million species with each species represented by a name or set of names, and a unique identifier in the 'Open Tree Taxonomy' (OTT). We use date estimates from published phylogenetic studies in the OpenTree data store mapped to internal nodes on the OpenTree synthetic phylogeny[28]. OneZoom synthesizes the OpenTree phylogeny data with a curated backbone and other published phylogenies to provide a way to visualize the tree of life for broader audiences, and combines the phylogenetic data with the risk of extinction provided by the IUCN Red List. Together, these tools provide an opportunity to study phylogenetic diversity and evolutionary distinctiveness at the scale of all described life on earth.

We explore how the ED scores are distributed between species across the complete tree of life, focusing on selected major taxonomic groups. We estimate the total PD, and proportion of threatened PD for different clades of species, and contrast this to species richness. Finally, we estimate EDGE scores for all species that are evaluated by the IUCN red list of threatened species[30]. Whilst EDGE-based studies of increasingly large clades have been carried out[10,14,15], to the best of our knowledge, this is the first time that all described species with IUCN Red List data have been given EDGE scores.

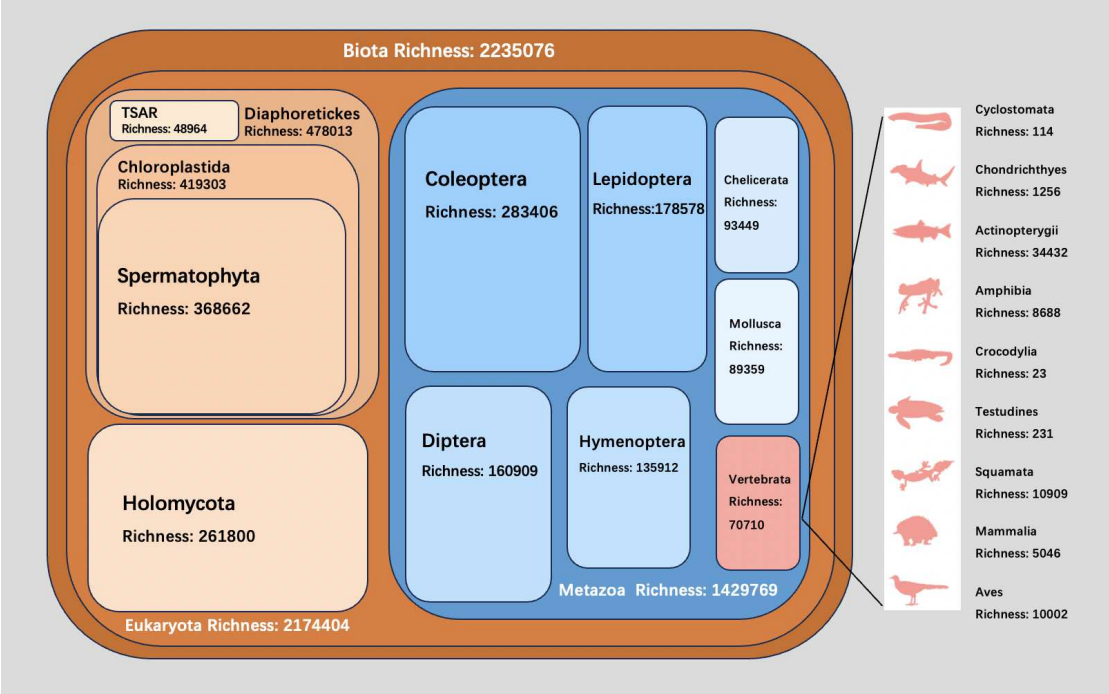


Fig 1. The relationship between the selected monophyletic clades for this study. We chose to include well known groups to aid comparison with other studies, as well as some more obscure ones based on them being monophyletic groups with distinctive positions in the tree of life. This meant accepting a known bias in existing work with our choice of taxa, whilst also including all life and selected understudied areas. In this figure we illustrate all our choices of taxa and the hierarchical relationships between them. The three colors (red, blue and orange) correspond to the nested categories of vertebrates, metazoan and eukaryotes respectively and will appear consistently in other figures. The size of each rectangle and the depth of its color qualitatively reflect the species richness of each clade to help with readability.

2. Results

The median ED score across all described species of life is 6.04 Myr (5^{th} and 95^{th} centiles are 0.36 - 47.91 Myr). We analyzed all described Eukaryotic life across 23 monophyletic clades (Fig 1). Species in the TSAR (Telonemia, Stramenopile, Alveolates and Rhizaria) clade have a median ED score of 39.59 Myr in the context of the complete tree of life. For Mollusca species median ED is 22.93 Myr and for the Cyclostomata clade median ED is 20.22 Myr. Among the

9 selected Vertebrata clades, there is relative consistency between the median ED scores of Actinopterygii (7.9 Myr), Amphibia (8.1 Myr) and Mammalia (8.55 Myr).

We observe a bimodal distribution of ED scores in the TSAR, Diaphoretickes and Hymenoptera clades. However large subclades of these groups usually have a single mode. For example, in the TSAR clade, the Rhizaria and Alveolata subclades, account for 61% of the species richness and have a single mode in their ED distribution. (Supplementary Fig 2).

The ED scores across all life appear log-normal but only superficially as they fail the (Anderson-Darling) test of log-normality (Fig 2). The Quantile-Quantile (Q-Q) plot of these groups is also available in supplementary material and illustrates that the departure from log-normality is primarily in the tails of the distribution (Supplementary Fig 1).

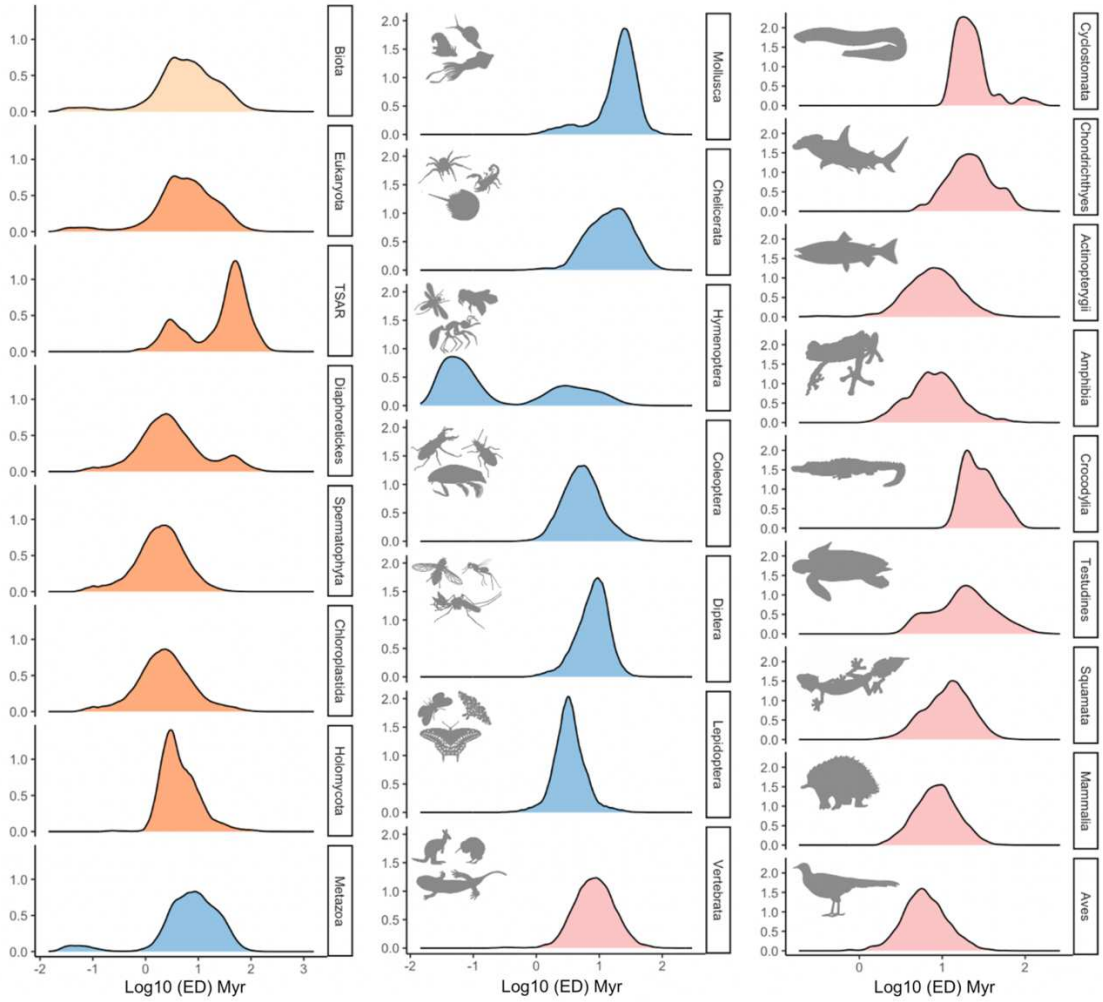


Fig 2. The distribution of ED scores across all species on log-scale, shown separately for selected monophyletic clades of species. Each species' ED score incorporated into these distributions represents the median of 100 replicates of a bootstrapping process - see methods. The 10 selected vertebrata clades are shown in red, the 7 selected metazoan clades (other than vertebrata) are shown in blue, and the 6 selected eukaryote clades (other than metazoan) are shown in orange. All life is shown in light orange. Silhouette image credits are given in the acknowledgements.

The median total PD (across bootstrap replicates) for all life calculated in our study was 30.84 trillion years. Our results show that 19 out of the 23 selected clades have robust estimates of total PD (which includes a stem back to the origin of life to enable comparability) in response to the bootstrapping whilst the Hymenoptera, Diptera and Cyclostomata clades showed a relatively high degree of uncertainty (Fig 3).

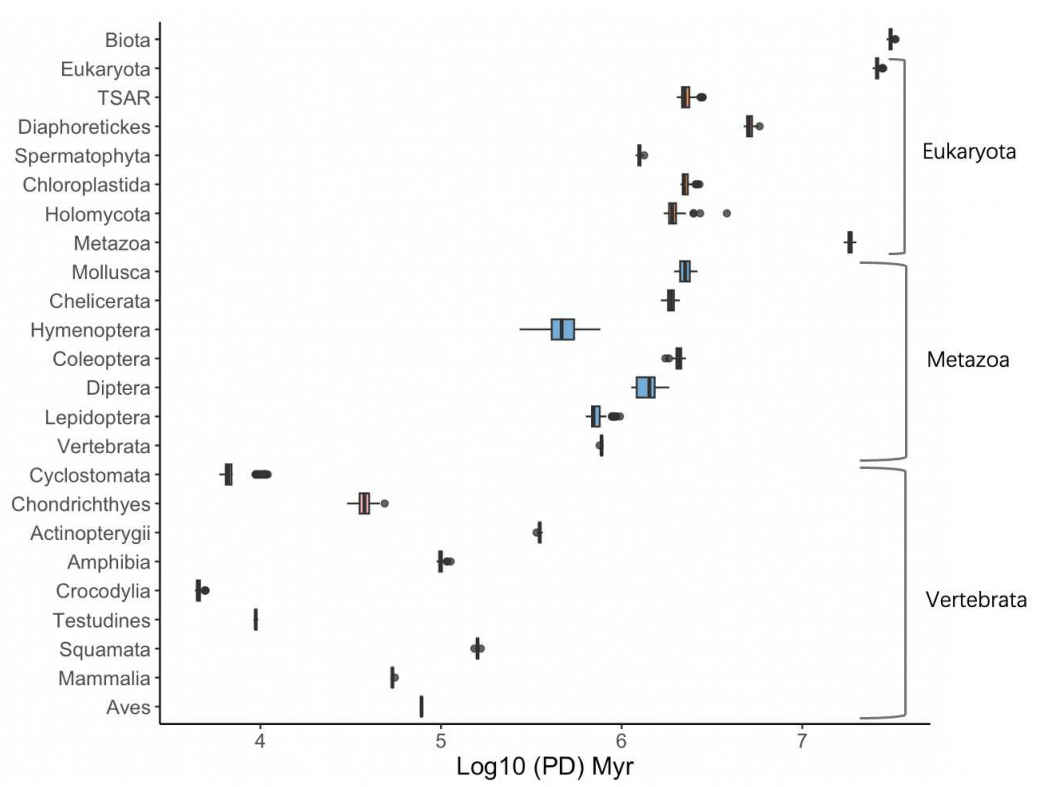


Fig 3. The distribution of total PD scores across selected monophyletic clades. Boxplots with median (solid line), 25th and 75th percentiles (box edges), and 5th and 95th centiles

(whiskers). Colors follow the same scheme as figure 2. The total PD values shown include a stem back to the origin of life to enable comparability and thus are slightly larger than the sum of the ED scores of all species in the clade because the sum of ED scores would only include a fair share of the stem.

Our results for all clades showed that the total PD increases approximately linearly with the species richness in both named and unnamed clades. (Fig 4. Named clades: Rs = 0.811, $p \ll 0.0001$; unnamed clades: Rs = 0.841, $p \ll 0.0001$; Spearman's rank correlation).

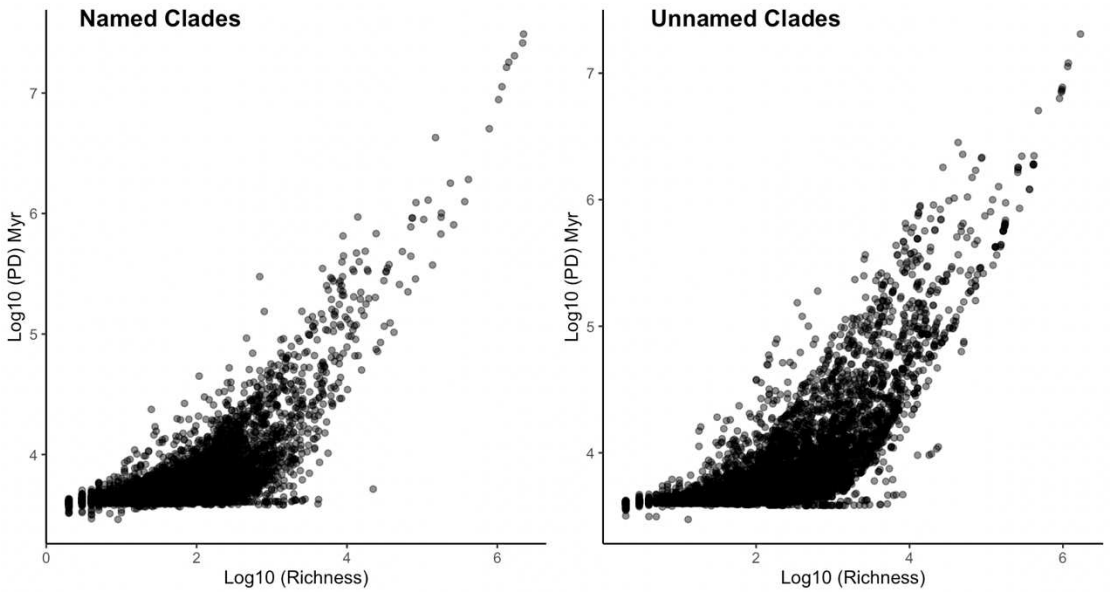


Fig 4. The relationship between species richness and total PD among randomly sampled named and unnamed monophyletic clades. Each point represents the median PD of a certain clade.

Among selected clades, TSAR, Mollusca, Chelicerata and Cyclostomata contribute disproportionately to the total PD given their total number of species (species richness). The Spermatophyta, Lepidoptera, Hymenoptera and Aves clades represent low PD for their species richness (Fig 5). Among all the selected Eukaryota clades (except Eukaryota as a whole), Metazoa has the highest median total PD (18.4 Tyr - trillion years), which can be attributed to

the significantly higher species richness compared with other Eukaryota clades. Actinopterygii is the Vertebrata clade that has both the highest species richness and the highest median total PD (351,885 Myr).

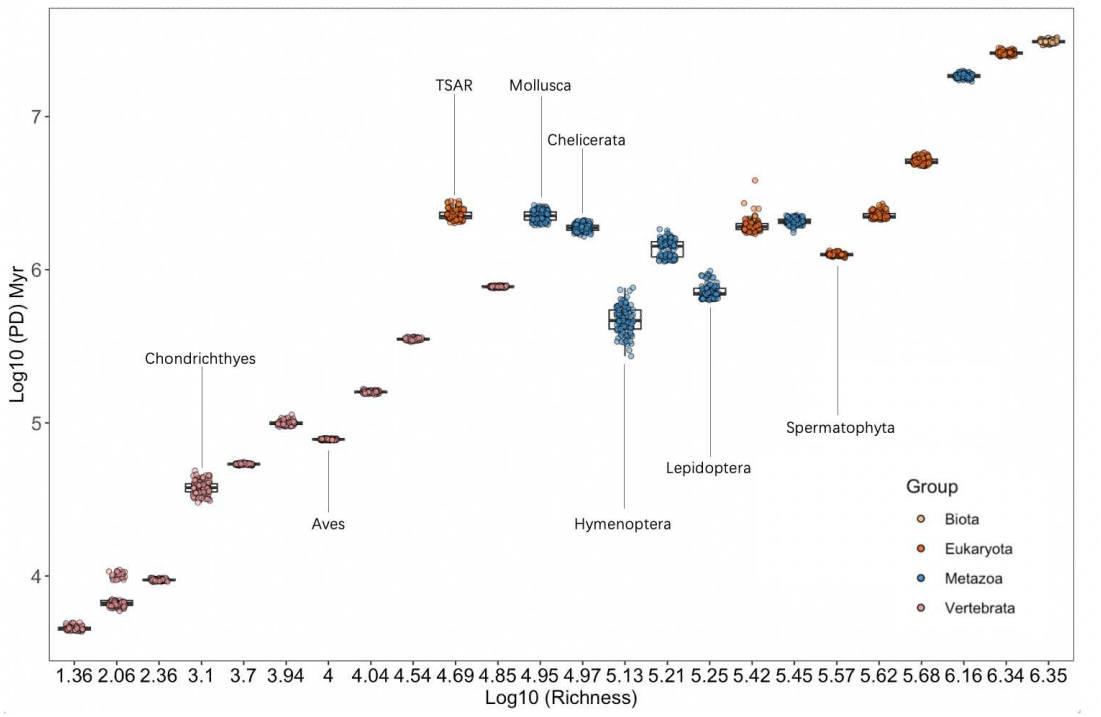


Fig 5. The relationship between species richness and total PD among selected clades. Clades with particularly high or low contribution to PD for their species richness are labeled with names to facilitate further discussion.

The ratio of PD to species richness varied between 3.4 and 196.29 Myr per species. Seven taxa including TSAR, Mollusca, Chelicerata, Cyclostomata, Chondrichthyes, Crocodylia and Testudines had a ratio greater than 20 (see Table 1).

Table 1. The relationship between species richness and total PD.

Group of taxa	Species richness	Proportion of dated nodes	Total PD (Myr) including stem back to origin of life	Ratio of total PD to species richness	Total unique PD (Myr)	Threatened unique PD (Myr)	Proportion of unique PD that is known to be threatened	Proportion of IUCN assessed species	Threatened species richness (Myr)
Eukaryota	2174404	2.15%	26005538.9	11.96	26004213.9	174339.3	0.67%	6.09%	1.67%
TSAR	**48964**	**0.14%**	**2233050**	**45.6**	**2230925**	**33.7**	**0.002%**	**0.01%**	**0.01%**
Chloroplastida	419303	3.48%	2217127.9	5.29	2214903.	54189.2	2.45%	13.49%	5.25%
Spermatophyta	368662	3.01%	1255387.9	3.4	1251769.4	50718	4.05%	15.08%	5.86%
Diaphoretickes	478013	3.07%	5057951.7	10.58	5056026.7	54442.7	1.08%	11.85%	4.61%
Holomycota	261800	0.11%	1907050.4	7.28	1904375.4	595.6	0.03%	0.19%	0.09%
Metazoa	1429769	2.3%	18409248.5	12.88	18406023.5	119322.5	0.65%	5.26%	1.04%
Mollusca	**89359**	**0.16%**	**2247533.6**	**25.15**	**2244016.4**	**21640**	**0.96%**	**7.79%**	**2.20%**
Chelicerata	**93449**	**0.2%**	**1876277.8**	**20.08**	**1872723.7**	**2836.1**	**0.15%**	**0.44%**	**0.26%**
Hymenoptera	135912	0.52%	465652.5	3.43	461987.1	553.4	0.12%	0.44%	0.12%
Coleoptera	283406	0.18%	2066422.1	7.29	2062724.6	1391.6	0.07%	0.41%	0.10%
Diptera	160909	0.06%	1423737.1	8.85	1419967	804.5	0.06%	0.19%	0.09%
Lepidoptera	178578	0.3%	698591.9	3.91	694770.9	536.5	0.08%	0.71%	0.12%
Vertebrata	70710	42.47%	774290.9	10.95	770798.3	68304.1	8.86%	74.55%	13.66%
Cyclostomata	**114**	**32.74%**	**6640.6**	**58.25**	**3140.7**	**94.8**	**3.02%**	**89.5%**	**13.16%**
Chondrichthyes	**1256**	**11.24%**	**37705.4**	**30.02**	**34140.4**	**5298.7**	**15.52%**	**84.47%**	**28.9%**
Actinopterygii	34432	27%	351885	10.22	348288.7	17327.6	4.98%	59.52%	8.13%
Amphibia	8688	20.81	99420.5	11.44	95735.2	15289.9	15.97%	87.71%	29.14%
Crocodylia	**23**	**100%**	**4514.6**	**196.29**	**729.6**	**207.95**	**28.51%**	**100%**	**47.83%**
Testudines	**231**	**100%**	**9422**	**40.79**	**5662**	**2191.1**	**38.7%**	**79.65%**	**56.71%**
Squamata	10909	31.54%	159180.1	14.59	155429.8	15293	9.84%	86.47%	14.61%
Mammalia	5046	99.94%	53781.1	10.66	50075.9	6941.1	13.86%	94.79%	20.75%
Aves	10002	100%	77963	7.79	74177.7	5394.3	7.27%	63.26%	11.64%

Threatened PD of Vertebrates. The relationship between total PD and species richness for the 23 selected clades is shown in table 1. The columns of the table have the following meanings: i) Group of taxa - this is a monophyletic group matching those introduced in figure 1; ii) Species richness - a count of the total number of species in the group; iii) Proportion of dated nodes - based on available data from OpenTree; iv) Total PD, which includes the stem back to the origin of life in order to facilitate comparison between clades with different crown ages; v) Ratio of PD to species richness - based on columns ii) and iv); vi) total unique PD - this excludes

193 the stem back to the origin of life but includes the shorter stem which connects to the group to
194 its closest related outgroup; vii) Threatened unique PD - a sum of branches within the clade
195 where all descendants are threatened (CR, EN or VU in the red list, DD species were not
196 considered threatened here); viii) Proportion of unique PD that is known to be threatened -
197 based on column vii) as a proportion of column vi) and expressed as a percentage; ix) proportion
198 of IUCN assessed species - the proportion of species richness that is covered by a valid IUCN
199 assessment i.e. excluding those which are not evaluated. The 7 clades highlighted in boldface
200 have a ratio of total PD / species richness of at least 20 and will be discussed further (Table
201 1).
202
203 We calculated the total PD and the proportion of threatened PD for Actinopterygii, Amphibia,
204 Testudines, Crocodylia, Squamata, Mammalia and Aves clades by repurposing our ED
205 estimation method to count branches based on the IUCN Red List categories of all descendant
206 species to each branch (See Methods). In terms of the proportion of evolutionary history at risk
207 in each clade, Testudines has the highest proportion of threatened PD (38.7%, 44.05% when
208 Data Deficient species were included). The least speciose clade, Crocodylia, has the second
209 largest proportion of threatened PD (28.51%, the same as when Data Deficient species were
210 considered threatened). Amphibians have the third largest proportion of threatened PD (15.97%,
211 24.7% when Data Deficient species were included). As the largest vertebrate clade in terms of
212 species richness and PD, Actinopterygii has 17.3 billion years (giga-years, Gyr) of threatened
213 PD (43.1 Gyr when Data Deficient species were considered threatened). The vertebrate clade
214 with the second largest PD was Squamata which has 15.3 Gyr of threatened PD (26.8 Gyr when
215 Data Deficient species were considered threatened). All estimates of threatened evolutionary
216 history are also given in Supplementary Data 1.
217
218 We calculated EDGE scores for those species that have an IUCN Red List assessment and are
219 aligned to the Open Tree Taxonomy by OneZoom. We ranked the EDGE scores in descending
220 order as a basis for assessing possible conservation priorities. In the top 20 EDGE species there

were 3 Malacostraca, 2 Actinopterygii, 2 Amphibia, 2 Copepoda, 2 Diplopoda, 1 Actinistia, 1 Dipino, 1 Florideophyceae, 1 Insecta, 1 Marchantiopsida, 1 Mollusca, 1 Polypodiopsida, 1 Peripatus and 1 Reptilia, no Aves or Mammals were in the top 20 despite a much larger proportion of those groups having Red List assessments. The list of the top 100 EDGE species identified in this study is available in Supplementary Table 1.

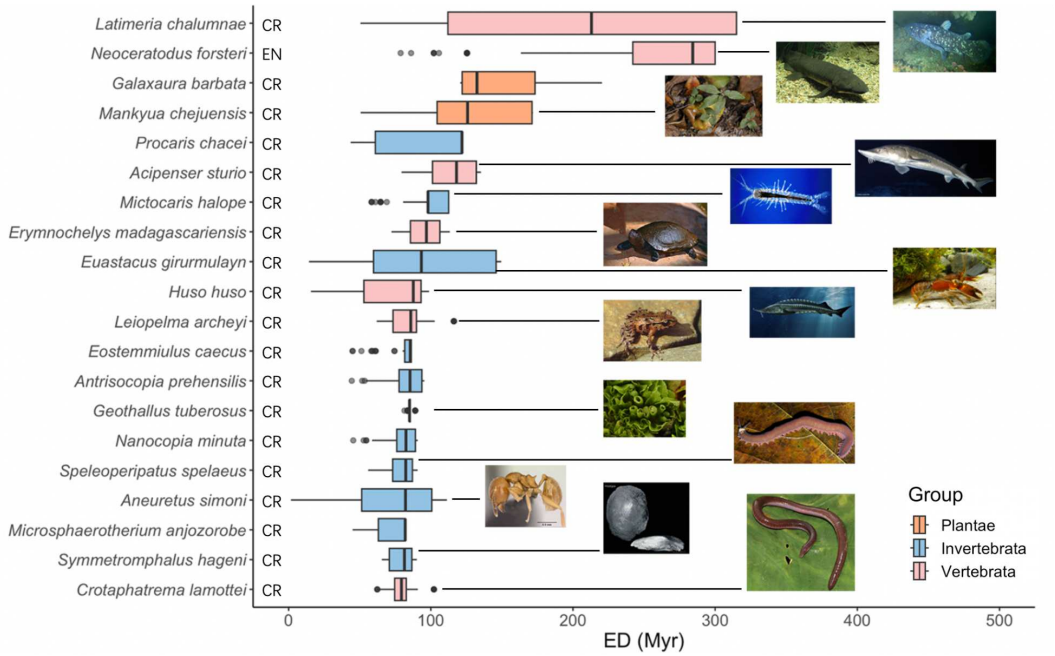


Fig 6. The ED scores of top 20 EDGE species. The images of species which have the highest EDGE score in its class are displayed with their conservation status (CR: Critically Endangered; EN: Endangered). The distribution of ED scores for each species here is generated through a bootstrap analysis to capture uncertainty in the data. This is done by sampling a subset of the date information and recalculating the result based on extrapolation from only that subset.

3. Discussion

The aim of our study was to characterize phylogenetic diversity and evolutionary distinctiveness for all described life on earth. This has only recently been made possible by the arrival of new phylogenetic datasets adding dates to some nodes on the Open Tree of Life[25,28]. We estimate the total PD of all described life to be between 29 and 33 trillion years.

Whilst most clades exhibited unimodal log-normal-like distributions of ED across all their species, the TSAR, Diaphoretickes and Hymenoptera clades instead had bimodal distributions. We hypothesize that the bimodality is caused by heterogeneity among smaller groups within those clades. Many of the clades displaying bimodality can be disaggregated into smaller groups with unimodal distributions of different medians that combine to give an overall bimodal distribution (Supplementary Fig 2). This finding is consistent with observations that patterns of diversification can shift substantively between related monophyletic clades[31,32]. It seems feasible that distributions with more than two modes could occur, but we did not observe this in our groups, possibly because subgroups with extreme values of ED in their species also contain very few species and thus are insignificant in terms of their contribution to the overall ED distribution of the larger clades. A unimodal distribution of ED scores was found in some extremely large clades such as Lepidoptera (178,578 species) and Mollusca (89,359 species). This suggests that even though evolutionary processes are likely heterogeneous at some scale within these clades, they are likely less so compared to the other clades we have studied.

Our estimation of PD scores for Actinopterygii, Testudines, Crocodylia, Lepidosaurs and Aves are fairly close to earlier estimations[33,34] except Amphibia, which is largely underestimated in our study. This inconsistency could be attributed to the fact that Amphibia is a large vertebrate clade with only 20% of nodes having date estimates. Our estimations for the proportion of threatened PD among major tetrapod clades including Amphibian, Mammalia and Aves are largely congruent with an early estimate[34] (less than 5% difference, see Supplementary Data 1), the proportion of threatened PD for Actinopterygii is underestimated in our study (12% vs. 4.9%). The estimation of mean ED scores across different clades in our study are also comparable to previously published estimates (See Supplementary Data 1). However, larger incongruencies can be observed in the comparison of ED scores for certain species, which indicates that lacking phylogenetic information may have a greater impact on both ED and EDGE score estimation at species level (Supplementary Fig 3).

Among all the eukaryotic clades, the TSAR was identified as the clade containing the largest mean evolutionary distinctiveness across all its individual species. However, even though there are some advocates for conserving the biodiversity represented by microorganisms[35,36], the TSAR group still falls outside of the mainstream of conservation and there is almost no information about the extinction risk of TSAR species on the IUCN Red List. According to our estimation, the TSAR clade has the highest proportion of species whose ED score is higher than 95% of all described species (39.2%, 19194 species among 48964 TSAR species, see Supplementary Fig 4). There is very little data informing ED estimates in TSAR. There are just a few phylogenetic studies informing the relationships of these taxa, and only 0.14% of nodes in this clade have any date estimates at all. Nonetheless, under the assumption that PD does indeed capture important elements of biodiversity such as future options for humanity, our results suggest that species within the TSAR clade should be given some attention. Our result does also reveal a potential overfocus on the conservation of mammals and aves, because the average ED score in TSAR is 45.6 Myr of evolutionary history (Supplementary Fig 5), which is 4 times higher than the top ED mammals (9.93 Myr) and about 6 times higher than the top ED aves (7.43 Myr). Whilst these numbers are heavily dependent on extremely sparse data for TSAR, they do at least suggest from the topology of the tree that there are either many unknown species in the group, or the known species are of extremely high value for capturing evolutionary history (PD). This large quantitative mismatch between TSAR, mammals and birds, may also be a symptom of limitations in the concept of PD at the scale of the complete tree of life. An alternative in the form of Evolutionary Heritage (EvoHeritage) would be a fruitful direction for future work[37].

Although the TSAR clade accounts for only 2.25% of eukaryotic species richness, this clade contributes 7.2% of the eukaryotic phylogenetic diversity with an average of 45.6 Myr of ED across species (see supplementary Fig 5), which is a disproportionately high contribution given their low species richness. However, this result might be biased by the absence of many species expected to exist but not yet described. These species, if added, would increase species richness

293 and PD disproportionately, likely bringing the ratio of PD to species richness closer to that of
294 other groups. The same will likely be true for many of the other less well studied branches of
295 the tree of life. In addition, in our analyses, the TSAR clade also has date estimates for only 69
296 out of 48963 nodes, and lacks phylogenetic information for 88.3% of nodes in that clade. The
297 high uncertainty in dates and relationships for this clade could be elevating PD scores. With the
298 development of new sequencing techniques, including higher-throughput methods, new species
299 are now being rapidly added to the tree of life[38] and it seems likely that older and less diverse
300 clades will continue to be updated in terms of both species and topology as new data comes in.
301 If these clades are understudied they may be updated to a greater degree than other clades that
302 are already rich in described species. Future work could model the effects of missing species
303 on the ED distributions of such clades as well as seek to add new data. Whilst this manuscript
304 presents an estimate for the total PD of all life and many selected clades (Fig 3), our estimation
305 of some clades in particular e.g. Hymenoptera, Diptera and Cyclostomata still show substantial
306 uncertainty in terms of estimated total PD. The highest and lowest estimates for PD of these
307 clades differ by approximately a factor of 3 (Hymenoptera: from 272,530.2 Myr to 762907.2
308 Myr; Diptera: from 1,130,409 Myr to 1,836,640 Myr; Cyclostomata: from 5,923.5 Myr to
309 10,977.4 Myr). These uncertainties arise from a combination of low proportion of dated nodes
310 and resolved nodes (Hymenoptera) and low species richness which accentuates the amount of
311 uncertainty captured by the fixed bootstrapping process (Cyclostomata). Furthermore, there are
312 other sources of uncertainty not captured by the bootstrap analysis, specifically the random
313 polytomy resolution method for the tree used[26]. We hope that efforts to generate and collate
314 more dated nodes for large and understudied groups, as well as improved data analysis pipelines
315 will both reduce uncertainty and capture uncertainty more completely in future iterations of this
316 work.
317
318 Scientists and conservationists alike have long advocated the idea of incorporating phylogenetic
319 diversity into conservation through use of EDGE-like metrics[39–42]. Here, we combined the ED

scores estimated in this study with the extinction risk data provided by the IUCN Red List in order to calculate EDGE scores. This analysis identified the critically endangered *Latimeria chalumnae* as the top EDGE species. As one of the only two living coelacanth species that are sisters to both lungfish and tetrapod, *L. chalumnae* may help reveal important facts about the adaptation of vertebrate life to land[43]. The coelacanth species also exhibit unique traits such as the intracranial joint and the rostral organ. Given the precarious status of *L. chalumnae* in the wild, the fact that it was for a long time thought extinct, and its unique features, it is a natural leader on our EDGE list and should be considered a priority for conservation and study, in order to preserve biodiversity in a broad sense[44,45]. The other coelacanth species *Latimeria menadoensis* is vulnerable and thus appears much lower down the EDGE ranking, despite its phylogenetic position. Among the top EDGE species in our analysis, the EDGE score estimates for *L. chalumnae*, *N. forsteri*, *M. chejuensis*, *A. sturio*, *E. madagascariensis*, *E. girurmulayn* can be considered reliable because most of their ED score comes from branches of the tree bounded by dated nodes. However, the ED scores of *G. barbata*, *P. chacei*, *M. halope* and *H. huso* are supported by interpolations and random polytomy resolution so whilst they rank highly here they may well rank lower in future analyses (see Supplementary Fig 6).

The original EDGE metric we use in our study is not the only metric of interest in its class. For example, Heighted EDGE (HEDGE) is a matrix to quantify how much a species is likely to contribute future biodiversity in terms of PD[42] based on potential extinctions. EDGE2 is another more recent advance that incorporates the core principles of HEDGE together with a protocol for its practical use. In EDGE2, the avertable loss of phylogenetic diversity through the conservation of individual species is cast in terms of ED2 (an updated version of ED) and GE2 (a flexible global endangerment weighting between 0 and 1 based by default on the IUCN Red List of Threatened species[30,40]). The EDGE2 and HEDGE metrics take the extinction risk of closely-related species into account, which makes them a more accurate reflection of the contribution each species is expected to make to overall PD in the future, however, evaluating them is more involved. It was not practical to estimate ED2 scores using our method across the

entire tree of life because we never dated nodes individually and instead estimated ED directly from the incomplete data. An EDGE2 analysis of the species with IUCN red list data also requires extinction risk for close relatives that will in most cases be unknown. EDGE2 has a mechanism for dealing with this, but the calculation of it on a tree with such few dated nodes would require a different method to that which we use here for getting dates for each of the true nodes in the tree. Overall, EDGE2 is likely to produce similar priorities to the EDGE approach[40]. However, there can be important differences for individual species when the ED scores are driven more by the internal branches (Supplementary Fig 6).

The priority of conserving *Latimeria chalumnae* according to EDGE2 is likely to drop because its closely-related species, *Latimeria menadoensis* (Indonesian Coelacanth), shares a lot of evolutionary history with *Latimeria chalumnae,* and is only Vulnerable which places *Latimeria menadoensis* at the 94th top EDGE species from our rankings. With a closely related species facing a lower risk of extinction, most of the evolutionary history represented by *Latimeria chalumnae* has more than one chance to survive, and the EDGE score is likely to overestimate the loss of evolutionary history that would be caused by the extinction of *Latimeria chalumnae*. In the case of *M. chejuensis* and *N. forsteri*, the high number of species in their sister clade, means that only the terminal branch will contribute significantly to both its EDGE score and EDGE2 score so the difference between EDGE and EDGE2 will not be large (Supplementary Fig 6). As with other species-poor clades, the discovery of a new and possibly less endangered sister species would have a big effect. In the context of an accelerated biodiversity decline, it is possible that much of biodiversity loss, and PD loss is occurring in microorganisms invisible to the human eye. This highlights the importance of large-scale estimation of phylogenetic diversity to better identify those clades of microorganisms that represent large amounts of biodiversity at the scale of all described life.

To the best of our knowledge this is the first attempt to estimate total PD at the scale of all described life, but see PhyloNext (https://phylonext.github.io/citations/, DOI:

10.5281/zenodo.7974081) for a promising integration of Open Tree of Life with GBIF. This comes with some sources of error due to the limited availability of data at the present time. As phylogenies capturing relationships and date estimates continue to be added to the Open Tree database the calculations produced by our method will be incrementally improved. We hope that this work will inspire the generation of a more automated pipeline for large scale PD calculations that will be broadly applicable in the future studies of biodiversity and streamline conservation biology research and practice.

Material and methods.

This study calculated ED scores for more than 2.23 million species and total PD for selected monophyletic clades. The methods consist of the following components: data synthesis, estimating total PD for selected clades of species, studying the distribution of ED scores, calculating EDGE scores for species with extinction risk assessment, linear regression model to test the relationship between total PD and species richness and calculation of total threatened PD.

Data. We made use of two data sources in our method (Fig 7). First, a tree structure linked to with IUCN Red List data provided by OneZoom in the form of two data tables: one for leaves the other for nodes in the tree[26]. Second, a JSON data file giving dates for certain nodes identifiable in the Open Tree of life. The OneZoom data tables were originally generated by the OneZoom tree of life explorer from the Open Tree of Life[26,27] synthesized with a curated backbone and other phylogenetic sources. In the nodes table, each row has a node ID and associated information including a parent field to give the node ID of that node's parent node. The "leaf_left" and "leaf_right" fields give rapid access to the set of the descendants in the leaf table[25]. In the leaves table each leaf species also has a field giving its 'Open Tree Taxonomy' (OTT) ID, which is a unique id assigned by OpenTree to taxonomically named nodes as a canonical taxon identifier[46]. Both the leaves table and nodes table have a "real_parent" field, which allows us to identify the parent node of this node or leaf in the original data source.

These “real_parent” identifiers are node ids in the OneZoom Tree of Life, which are not stable, but are always mapped wherever possible to node IDs in the Open tree of life synthetic tree. In contrast the parent field may point to a node generated only through a process of random polytomy splitting done as part of the synthesis described in dynamic visualization of million-tip trees: the OneZoom project[26,27].

The dates used are generated from phylogenetic estimates of dates for internal nodes in published studies and are provided in JSON format by McTavish et al.[28] These dates are mapped from published phylogenies to internal nodes of the OpenTree Synthetic tree using the OpenTree phylogenetic conflict APIs[29] (https://github.com/OpenTreeOfLife/chronosynth). This is an extension of the concepts published in DateLife[47]. In this data, some dates correspond to nodes with their own OTT ID whilst others are mapped to unnamed nodes in the Open Tree synthetic tree. These unnamed nodes are given an identifier based on an arbitrary selection of two leaves for which that node is the most recent common ancestor. We linked the nodes table from OneZoom[26] with the JSON data from Open Tree of Life[28] into a single table with node dates wherever known. In doing this node dates are only given to real nodes (i.e. not to nodes that only exist because of random polytomy resolution) (see Fig 7).

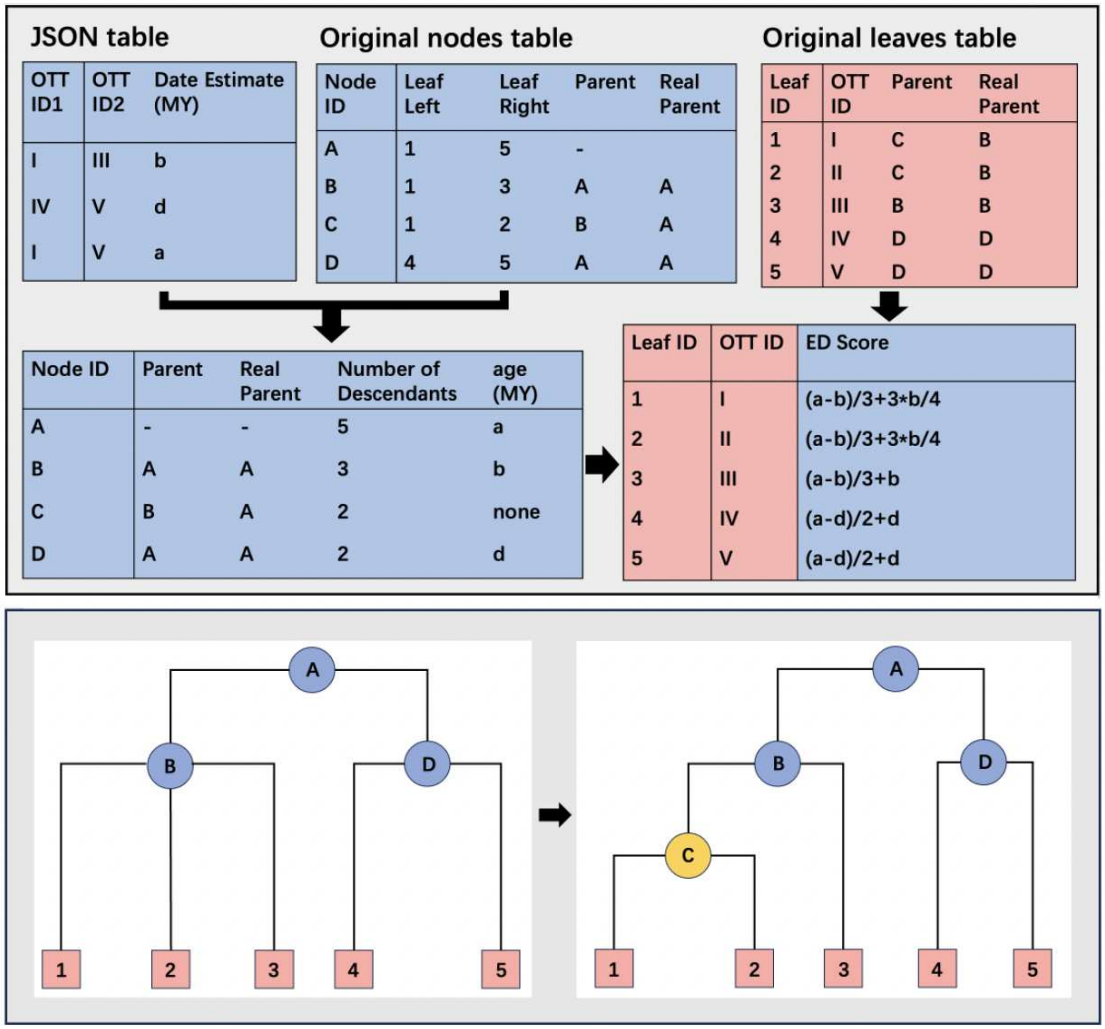


Fig 7. Overview schematic with an example tree showing the data we used and the process of calculating ED scores. Node IDs are the ones in the OneZoom database and are not stable but enable a neat structure for the tree being analyzed here. OTT taxon IDs are stable to updates in the tree data so are also included in our data tables (shown as roman numerals to differentiate them from the table IDs) and are used for most purposes. All leaves in the leaf table will be connected to a corresponding node in the node table. By finding date estimates and number of descendants for each parent from the re-assembled nodes table, the ED score for each leaf can be calculated.

Calculation of ED scores. We used the ED measure within which each branch length is weighted by the number of species which are descended from it. The ED score for a species is equal to the total of all the weighted lengths of the branches on the path from the root of the

tree to this species[10]. The ED of species i, written ED_i, can thus be calculated by the following formula:

$$ED_i = \sum_{j=1}^{n_i} \frac{L_{i,j}}{N_{i,j}}$$

Where n_i gives the total number of ancestral branches to species i, $L_{i,1}$ refers to the terminal branch length of species i, $L_{i,j}$ for $2 \leq j \leq n_i$ gives the total length of interior branch j ancestral to species i and $N_{i,j}$ gives the number of descendants for each of these same branches. The length of each branch can be calculated by the differences between two adjacent date estimates of nodes and the number of descendants for a node can be calculated by the difference between its "leaf_left" and "leaf_right" fields plus one[25]. In the calculation of ED scores for the entire tree of life, one major problem is that not all the nodes have a date estimate (For example, the node C in Figure 7). In this study, for the node without a date estimate, the branch length was estimated as the average length between the closest 2 interior nodes that have date estimates. For instance, when calculating ED score for leaf 1 in Figure 7, the length of the branch connecting node B and node C is equal to the length of the branch that connects node C to leaf 1. This can be represented as b/2. Using this method the estimated ED score of leaf 1 (or leaf 2) can be calculated as:

$$\frac{(a-b)}{3} + \frac{b}{4} + \frac{b}{2}$$

In order to capture a measure of uncertainty based on data deficiency, we conducted bootstrap analysis by randomly removing 50% of node date estimates and using the rest of the data to perform all the analyses - repeating the process 100 times. For the nodes that have multiple conflicting date estimates, we will randomly select one of the conflicting date estimates during the bootstrap analysis and discard the other(s). All the code used for the calculation of ED are available on GitHub (https://github.com/AlexGuojl/Phylogenetic-diversity-across-the-complete-tree-of-life).

Distribution of evolutionary distinctiveness. The ED score reported for each species is the median ED score among the 100 ED scores generated in the bootstrap analysis. The ED score was log transformed with the base 10. All the species included in this study were divided into these following monophyletic clades: Telonemia, Stramenopile, Alveolates and Rhizaria (TSAR), Diaphoretickes, Spermatophyta, Chloroplastida, Holomycota, Metazoa, Mollusca, Chelicerata, Hymenoptera, Coleoptera, Diptera, Lepidoptera and Vertebrata. We further separated the Vertebrata clade into 9 clades: Cyclostomata, Chondrichthyes, Actinopterygii, Amphibia, Crocodylia, Testudines, Squamata, Mammalia and Aves. For each clade, we also compared the ED score of each species within this clade to the 0.99 quantile, 0.95 quantile and median value of the ED scores across the complete tree of life. We produced a stacked bar plot to see the proportion of high ED species across these clades (see Supplementary Fig 4).

Total PD across major monophyletic clades. Total PD of a given clade can be represented as the sum of the branches within this clade. With all the ED scores calculated, the total PD of a clade in the tree of life can be calculated by the sum of ED scores minus the ED scores contributed by the root of this clade and plus the long stem which connects the origin of life. By summing the branches included by a given clade and plus a short stem that connects the most recent common ancestor of this clade and its parent node, we can also calculate the unique PD of the clade. We chose to then add this stem in its entirety in order to enable an unbiased comparison between the PD values of any pair of clades[37]. For example, the unique PD of clade C in figure 7 can be represented as: $2 * \frac{b}{2} + \frac{b}{2}$, and the total PD of this clade is: $2 * \frac{b}{2} + \frac{b}{2} + (a - b)$.

We use the ED scores generated in the 100 bootstrap replicates to calculate total PD for the monophyletic clades selected in this study. Each clade therefore has 100 bootstrapped values of total PD. We produced box plots to see the distribution of total PD among the selected clades and all the PD scores were log transformed. To test the relationship between median PD and

487 species richness for the selected clades, we conducted stratified sampling from both named and
488 unnamed interior nodes. We then calculated the total PD of 40631 named nodes and 40447
489 unnamed nodes and produced scatter plots of the results. By dropping all threatened species
490 (VU, EN and CR species) from a given tree and subtracting the resulting PD from the original
491 full PD, we estimated the threatened PD of Actinopterygii, Amphibian, Testudines, Crocodylia,
492 Lepidosauria, Mammalia and Aves. We also incorporated a "worse" situation in which the Data
493 Deficient (DD) species are regarded as threatened and therefore are dropped from the tree as
494 well.
495
496 **Priority table of EDGE species.** As a critical indicator of the health of the world's biodiversity,
497 the IUCN Red List gives a list of species and their risk of extinction[30,48]. By combining the
498 calculated ED scores with the red list, we calculated EDGE scores using following formula[13]:
499 $$EDGE_i = ln\ (1 + ED_i) + GE_i \times ln\ (2)$$
500 Where GE_i is the extinction risk of species *i*, based on the IUCN red list (Least Concern=0,
501 Near Threatened and Conservation Dependent = 1, Vulnerable=2, Endangered=3, Critically
502 Endangered=4). All species with a GE score were sorted by their EDGE scores in descending
503 order, both the top 20 and top 100 EDGE species were made into a priority table.
504
505 To quantify the contribution of each node to the ED of these top EDGE species, we further
506 calculated a cumulative ED value for each interior node, which is equal to the total weighted
507 lengths of the branches on the path from the root to this node. The branch lengths are weighted
508 in the same way as when calculating ED scores for a certain species. For example, in figure 7,
509 the cumulative ED value of node B can be represented as $\frac{(a-b)}{3}$. We did this estimation for all
510 the top10 EDGE species.
511
512 **Robustness test.** To test the robustness of the ED scores estimated in this study, we compared
513 the published mean ED scores of class Chondrichthyes[17], Amphibia[49], Squamata[50], Mammalia[51]

514 and Aves[21] with the mean ED scores generated in our bootstrap analysis (Supplementary Data
515 1). We also compared published data of ED and EDGE scores downloaded from the website of
516 EDGE of Existence program (https://www.edgeofexistence.org/species/) (see Supplementary
517 Fig 3). We produced scatter plots to characterize the deviation between our estimated values
518 and the previously calculated values for all EDGE species. All the ED scores for scatter plot
519 were log transformed.
520

521 ## Acknowledgments

522 The authors thank Rikki Gumbs for useful feedback on the manuscript. JR thanks the
523 leverhulme trust for supporting his contribution to this work through a Research Fellowship
524 (RF-2022-497). The silhouette images of organisms used in Figure 1 and Figure 2 were
525 collected from Phylopic (http://phylopic.org/). *Yochelcionella Cyrano*: Stanton F. Fink
526 (vectorized by T. Michael Keesey); *Helix aspersa*: Gareth Monger; *Gyrinus convexiusculus*:
527 Desmond W. Helmore; *Micropezidae*: T. Michael Keesey; *Ectoedemia coscoja*: E. J. Van
528 Nieukerken, A. Laštůvka, and Z. Laštůvka (vectorized by T. Michael Keesey); *Macropus rufus*:
529 Cathy, *Balanerpeton woodi*: Scott Hartman. To view a copy of the license, visit
530 https://creativecommons.org/licenses/by/3.0/.
531

532 The portraits used in Figure 6 were collected from different resource: *Latimeria chalumnae,*
533 *Neoceratodus forsteri, Mictocaris halope, Erymnochelys madagascariensis, Huso huso,*
534 *Leiopelma archeyi* and *Aneuretus simoni* were collected from Wikipedia; *Mankyua chejuensis*:
535 B.Y.Sun, M.H.Kim & C.H.Kim; *Acipenser sturio*: Photo by © Stefan Ludwig; *Euastacus*
536 *girurmulayn*: Copyright Rob McCormack (McCormack 2020); *Geothallus tuberosus*: Photo by
537 © 2023 Nathan Taylor CC-BY-NC 4.0; *Speleoperipatus spelaeus*: Ubirajara de Oliveira;
538 *Symmetromphalus hageni*: Ann. Naturhist. Mus. Wien - Vol. 93, B:253-257. Photo by L.A.
539 Beck; *Crotaphatrema lamottei:* BMNH 2008.275 in life. Photo by T. M. Doherty-Bone.
540

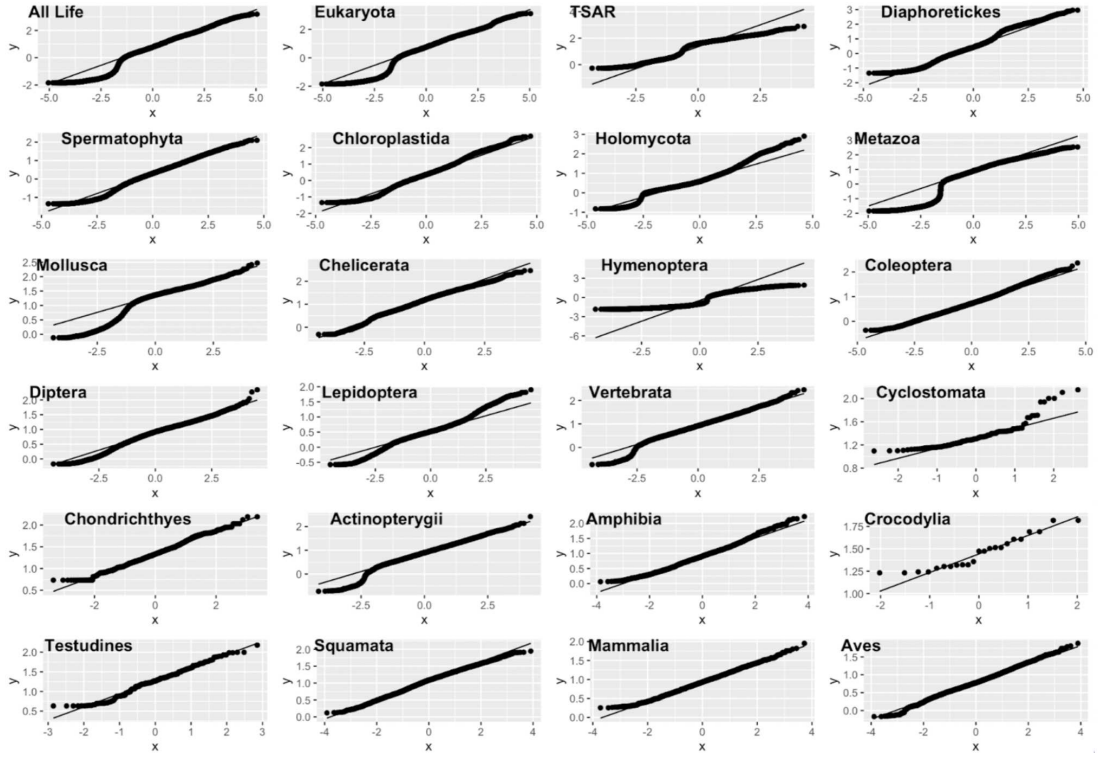
Reference

1. Barnosky, A. D. *et al.* Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57 (2011).
2. Pimm, S L et al. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* **344**, 1246752 (2014).
3. Ceballos, G. *et al.* Accelerated modern human-induced species losses: Entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253 (2015).
4. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science (80-.).* 366, eaax3100 (2019).
5. Faith, D. P. Conservation Evaluation and Phylogenetic Diversity. *Biological Conservation* **61**,1–10 (1992).
6. IPBES. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (eds. Brondízio ES, Settele J, Diaz S, Ngo HT.) (IPBES, Bonn, 2019).
7. Forest, F. *et al.* Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature* **445**, 757–760 (2007).
8. Molina-Venegas R. Conserving evolutionarily distinct species is critical to safeguard human well-being. *Sci Rep* **11**, 24187, (2021).
9. Molina-Venegas R, R. M., Pardo-de-Santayana M, Ronquillo C, Mabberley DJ. Maximum levels of global phylogenetic diversity efficiently capture plant services for humankind. *Nat Ecol Evol* **5**, 583–588 (2021).
10. Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. Mammals on the EDGE: Conservation priorities based on threat and phylogeny. *PLoS ONE* **2**, e296 (2007).
11. Secretariat of the Convention on Biological Diversity. Indicators for the post-2020 Global Biodiversity Framework. CBD/SBSTTA/24/INF/16 57 https://www.cbd.int/doc/c/a6d3/3108/88518eab9c9d12b1c418398d/sbstta-24-inf-16-en.pdf (2021).
12. IUCN. Phylogenetic Diversity Task Force. https://www.pdtf.org/ (2019).
13. Redding, D. W. & Mooers, A. O. Incorporating evolutionary measures into conservation prioritization. *Conserv. Biol.* **20**, 1670–1678 (2006).
14. Forest, F. *et al.* Gymnosperms on the EDGE. *Sci. Rep.* **8**, 6053 (2018).
15. Gumbs, R., Gray, C. L. ,. Wearn, O. R. ,. Owen, N. R. Tetrapods on the EDGE: overcoming data limitations to identify phylogenetic conservation priorities. *PLoS One* **13**, e0194680 (2018).
16. Mccarthy, D. P. *et al.* Financial costs of meeting global biodiversity conservation targets: current spending and unmet needs. *Science* **338**, 946–949 (2012).
17. Stein, R. W. *et al.* Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nat. Ecol. Evol.* **2**, 288–298 (2018).
18. Gumbs, R. *et al.* Global priorities for conservation of reptilian phylogenetic diversity in the face of human impacts. *Nat. Commun.* **11**, 2616 (2020).
19. Rosauer, D. F., Pollock, L. J., Linke, S. & Jetz, W. Phylogenetically informed spatial planning is required to conserve the mammalian tree of life. *Proc. R. Soc. B Biol. Sci.* **284**, 20170627 (2017).
20. Pollock, L. J., Thuiller, W. & Jetz, W. Large conservation gains possible for global biodiversity facets. *Nature* **546**, 141–144 (2017).
21. Jetz, W. *et al.* Global Distribution and Conservation of Evolutionary Distinctness in Birds. *Curr. Biol.* **24**, 919–930 (2014).
22. Rosauer, D., Laffan, S. W., Crisp, M. D., Donnellan, S. C. & Cook, L. G. Phylogenetic endemism: A new approach for identifying geographical concentrations of evolutionary history. *Mol. Ecol.* **18**, 4061–4072 (2009).
23. Davis, M., Faurby, S. &. Svenning, J. C. Mammal diversity will take millions of years to recover from the current biodiversity crisis. *Proc Natl Acad Sci* **115**, 11262–11267

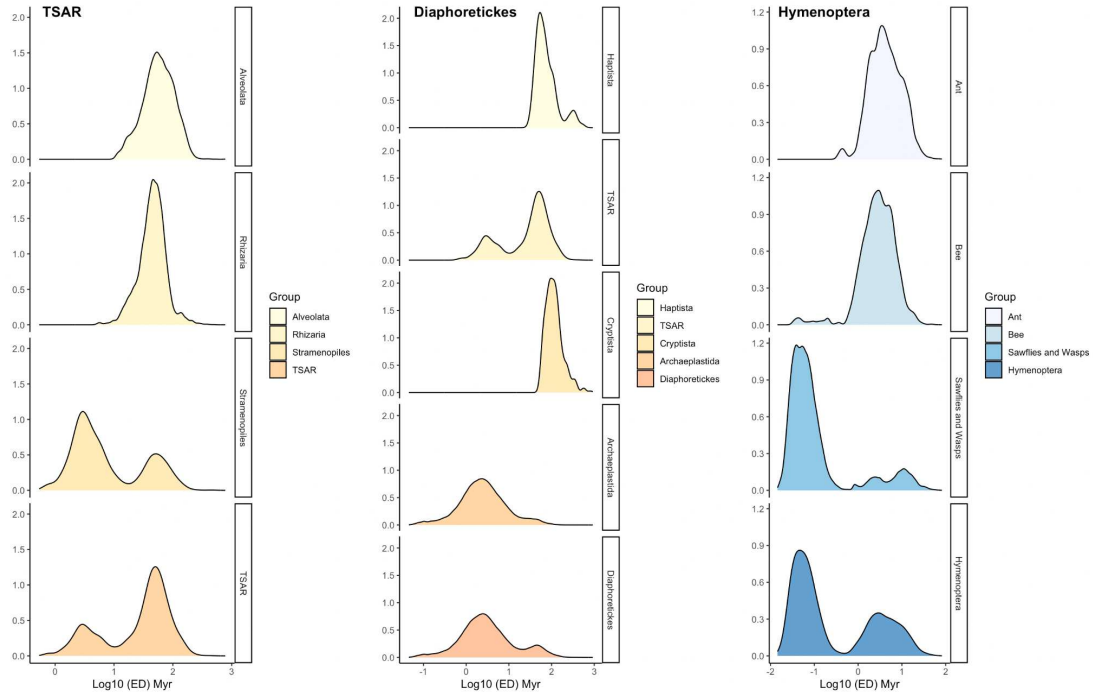
(2018).
24. Daru, B. H. *et al.* Spatial overlaps between the global protected areas network and terrestrial hotspots of evolutionary diversity. *Glob. Ecol. Biogeogr.* **28**, 757–766 (2019).
25. Hinchliff, C. E. *et al.* Synthesis of phylogeny and taxonomy into a comprehensive tree of life. **112**, 12764–12769 (2015).
26. Wong, Y. & Rosindell, J. Dynamic visualisation of million-tip trees: The OneZoom project. *Methods Ecol. Evol.* **13**, 303–313 (2022).
27. Rosindell, J. & Harmon, L. J. OneZoom: A Fractal Explorer for the Tree of Life. *PLoS Biol.* **10**, e1001406 (2012).
28. Mctavish, E. J. *et al.* Phylesystem: A git-based data store for community-curated phylogenetic estimates. *Bioinformatics* **31**, 2794–2800 (2015).
29. Redelings, B. D. & Holder, M. T. A supertree pipeline for summarizing phylogenetic and taxonomic information for millions of species. *PeerJ* **5**, e3058 (2017).
30. IUCN. IUCN red list of threatened species. www.iucnredlist.org (2022).
31. Magallo´n, S., Magallo´n, M. & Sanderson, M. J. ABSOLUTE DIVERSIFICATION RATES IN ANGIOSPERM CLADES. *Evolution* **55** 1762–1780 (2001).
32. Maliet, O., Hartig, F. & Morlon, H. A Model with Many Small Shifts for Estimating Species-Specific Diversification Rates. *Nat Ecol Evol* **3**, 1086–1092 (2019).
33. Gumbs, R. S. Setting global priorities for the conservation of tetrapod phylogenetic diversity. *Imperial College London*, (2021).
34. Gumbs, R., Scott, O., Bates, R. et al. Global conservation status of the jawed vertebrate Tree of Life. *Nat Commun* **15**, 1101 (2024). https://doi.org/10.1038/s41467-024-45119-z
35. Cotterill, F. P. D., Al-Rasheid, K. & Foissner, W. Conservation of protists: is it needed at all? *Biodiversity and Conservation* **17**, 427-443 (2007).
36. Sharma, N. K. & Rai, A. K. Biodiversity and biogeography of microalgae: Progress and pitfalls. *Environ. Rev.* **19**, 1–15 (2011).
37. Rosindell, J., Manson, K., Gumbs, R., Pearse, W. D. & Steel, M. Phylogenetic Biodiversity Metrics Should Account for Both Accumulation and Attrition of Evolutionary Heritage.*Systematic Biology*. syad072, (2023).
38. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The New Tree of Eukaryotes. *Trends Ecol. Evol.* **35**, 43–55 (2020).
39. Faith, D. P. Threatened species and the potential loss of phylogenetic diversity: Conservation scenarios based on estimated extinction probabilities and phylogenetic risk analysis. *Conserv. Biol.* **22**, 1461–1470 (2008).
40. Gumbs, R. *et al.* The EDGE2 protocol: Advancing the prioritisation of Evolutionarily Distinct and Globally Endangered species for practical conservation action. *PLoS Biol*. **21**(2), e3001991 (2023).
41. Nunes, L. A., Turvey, S. T. & Rosindell, J. The price of conserving avian phylogenetic diversity: A global prioritization approach. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 1–11 (2015).
42. Steel, M., Mimoto, A. & Mooers, A. Ø. Hedging Our Bets: The Expected Contribution of Species to Future Phylogenetic Diversity. *Evol Bioinform Online.* **3** 237–244 (2007).
43. Amemiya, C. T. *et al.* The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496**, 311–316 (2013).
44. Tucker, C. M. *et al.* Assessing the utility of conserving evolutionary history. *Biol. Rev.* **94**, 1740–1760 (2019).
45. Curnick, D. J. *et al.* Setting evolutionary-based conservation priorities for a phylogenetically data-poor taxonomic group (Scleractinia). *Anim. Conserv.* **18**, 303–312 (2015).
46. Rees, J. & Cranston, K. Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodivers. Data J.* **5**, e12581 (2017).
47. Reyes, L. L. S. DateLife: Leveraging Databases and Analytical Tools to Reveal the Dated Tree of Life. *Syst. Biol.* syae015 (2024).
48. Butchart, S. H. M. *et al.* Measuring global trends in the status of biodiversity: Red list indices for birds. *PLoS Biol.* **2**, e383 (2004).

49. Isaac, N. J. B., Redding, D. W., Meredith, H. M. & Safi, K. Phylogenetically-Informed Priorities for Amphibian Conservation. *PLoS ONE* **7**, e43912 (2012).
50. Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W. & Pyron, R. A. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biol. Conserv.* **204**, 23–31 (2016).
51. Martyn, I., Kuhn, T. S., Mooers, A. O., Moulton, V. & Spillner, A. Computing Evolutionary Distinctiveness Indices in Large Scale Analysis. *Algorithms Mol Biol.* **7**, 6 (2012).

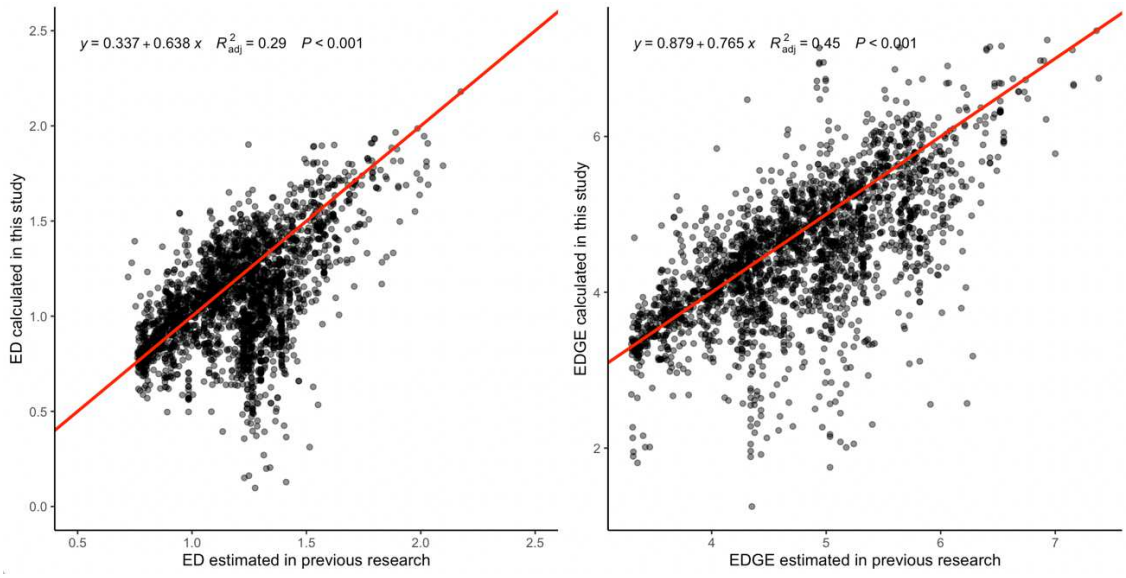
Supplementary materials



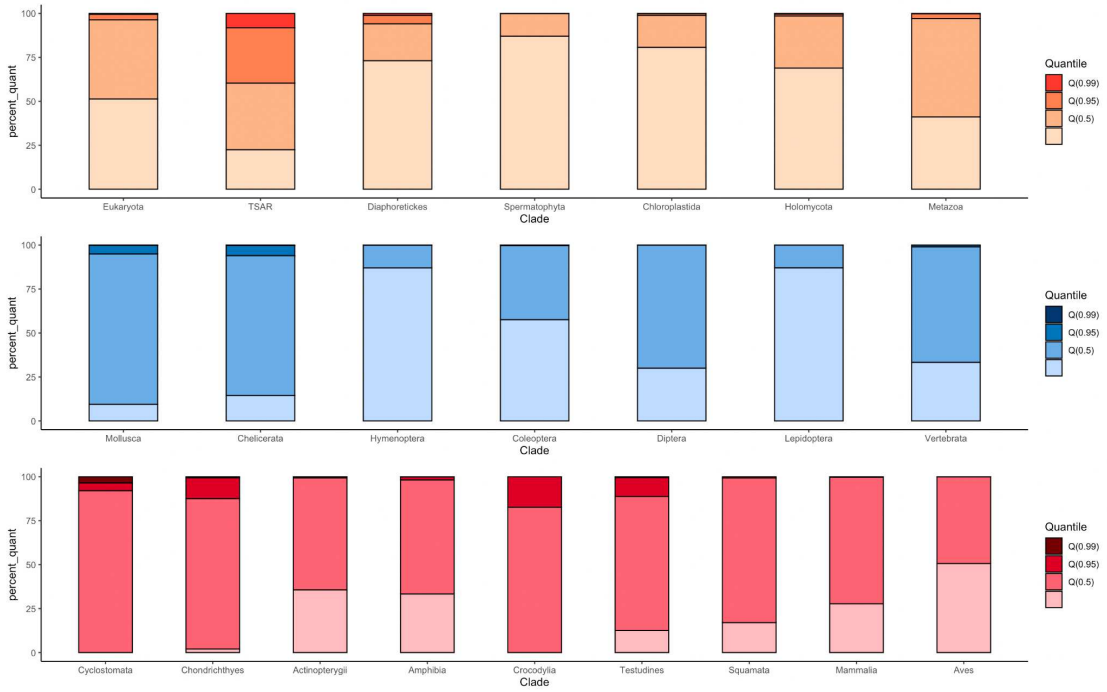
Supplementary Figure 1. The Q-Q plot of ED scores for the 24 selected groups. All of the selected groups have points that deviate from the straight diagonal line. This indicates that none of those groups has log-normally distributed ED scores.



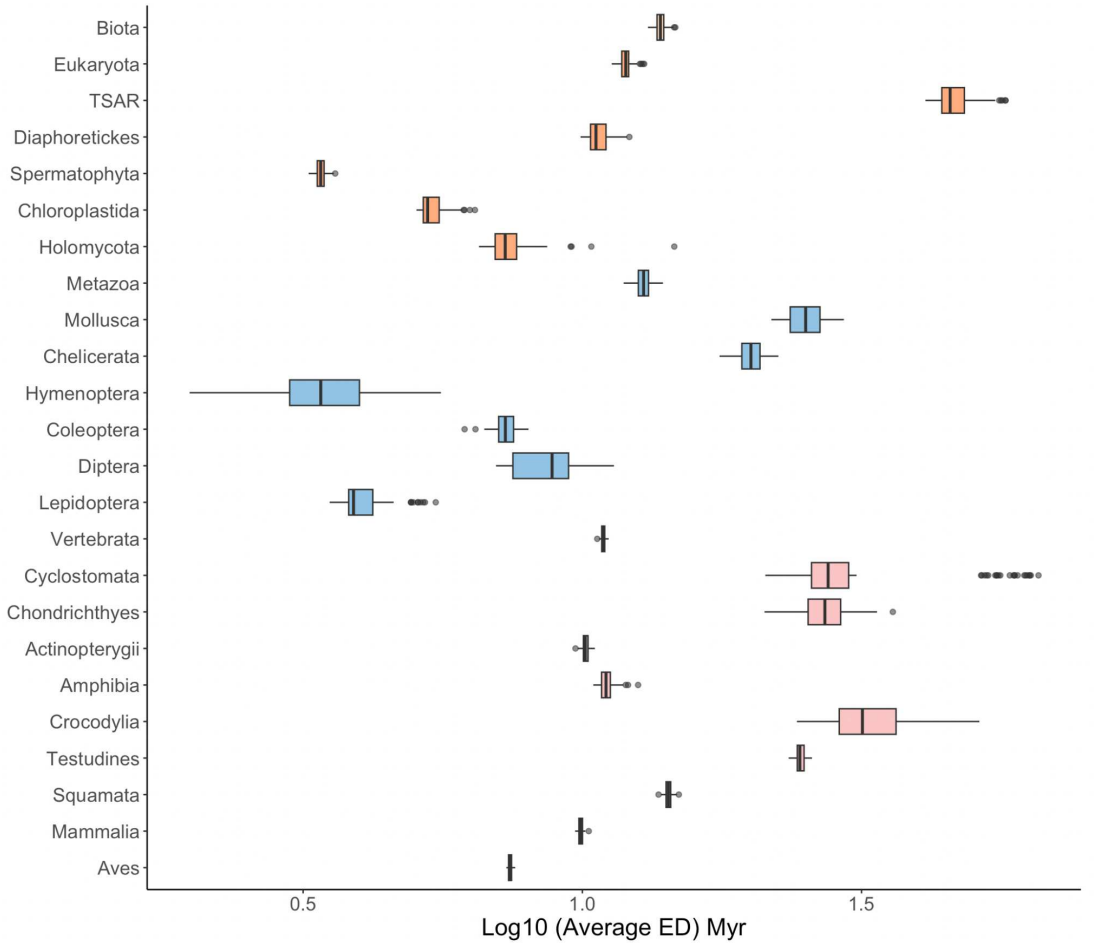
Supplementary Figure 2. The distribution of ED scores in Eukaryota subgroups. The selected eukaryota groups which show bimodal distribution of ED scores were further divided into subgroups. The depth of the color refers to the species richness of each subgroup and the color refers to broader categories using the same scheme as for other figures in the manuscript main text: orange for eukaryota without metazoa and blue for invertebrates.



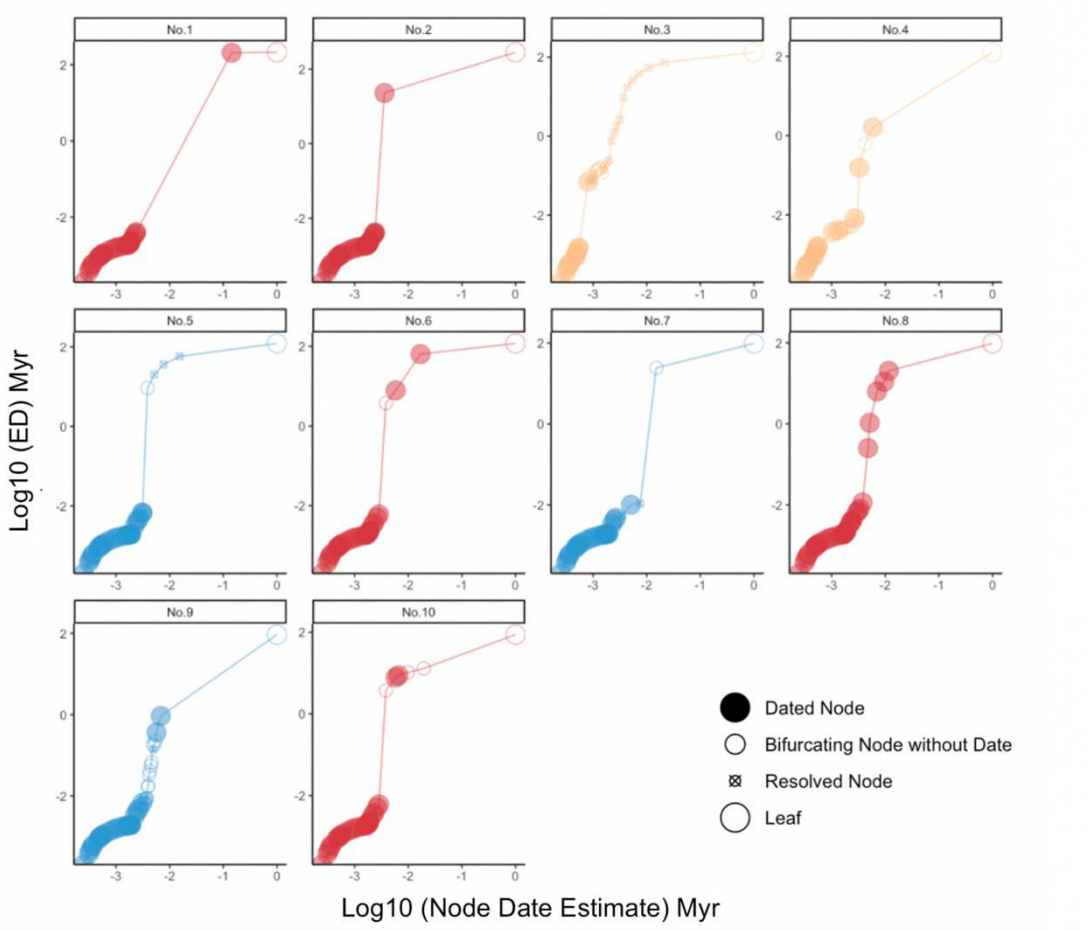
Supplementary Figure 3. Comparison of ED and EDGE scores. The x-axis shows the existing published values (Gumbs et al. 2018), and the y-axis is the values calculated in our study. The left panel gives a comparison for ED scores only and the right panel shows a comparison for EDGE scores. The red line is shown for illustrative purposes with an intercept of 0 and a slope of 1.



Supplementary Figure 4. The proportion of high ED species among selected clades. The stacked bar shows the proportion of species whose ED score is above the 0.99 quantile, 0.95 quantile and 0.5 quantile (median) of ED score of all described life on earth.



Supplementary Figure 5. The distribution of average ED scores for the selected groups. The average values are calculated by summing the ED scores of its corresponding group and divided by the species richness of that group. Each group therefore has 100 average ED scores.



690

691 **Supplementary Figure 6. Contribution of different nodes to the ED score of top 10 EDGE**
692 **species.** The different colors of points correspond to different species, the shape and size of
693 different points indicates the type of this node. “Dated nodes” have the strongest quantitative
694 backing with both a date and a topology known in the source data. “Bifurcating nodes” have a
695 known topology without polytomy, the date was only interpolated based on other dated nodes.
696 “Resolved nodes” appear in the data only through random polytomy resolution.

697

698

699

700

701

702

703
704

705 **Supplementary Table 1. The top 100 EDGE species.**
706

Rank	Name	ED (Myr)	EDGE	Conservation Status	95% Confidence Interval of ED (Myr)
1	*Latimeria chalumnae*	213.0	8.14	CR	70.89-315.03
2	*Neoceratodus forsteri*	284.2	7.73	EN	102.1-300.01
3	*Galaxaura barbata*	132.5	7.67	CR	120.69-179.2
4	*Mankyua chejuensis*	125.8	7.62	CR	50.63-171.69
5	*Procaris chacei*	121.9	7.58	CR	45.16-121.91
6	*Acipenser sturio*	118.1	7.55	CR	95.18-135.28
7	*Mictocaris halope*	98.0	7.37	CR	59.7-112.55
8	*Erymnochelys madagascariensis*	96.9	7.36	CR	75.41-113.35
9	*Euastacus girurmulayn*	93.3	7.32	CR	14.88-149.38
10	*Huso huso*	87.7	7.26	CR	16.23-98.43
11	*Leiopelma archeyi*	85.9	7.24	CR	67.52-116.22
12	*Eostemmiulus caecus*	85.5	7.23	CR	54.46-86.56
13	*Antrisocopia prehensilis*	85.3	7.23	CR	53.78-95.47
14	*Geothallus tuberosus*	85.0	7.23	CR	83.67-89.02
15	*Nanocopia minuta*	82.5	7.2	CR	54.59-90.86
16	*Speleoperipatus spelaeus*	82.4	7.2	CR	63.4-90.7
17	*Aneuretus simoni*	82.2	7.19	CR	1.6-111.32
18	*Microsphaerotherium anjozorobe*	82.0	7.19	CR	55.81-82.01
19	*Symmetromphalus hageni*	81.5	7.18	CR	69.37-90.15
20	*Crotaphatrema lamottei*	79.3	7.16	CR	62.35-102.17
21	*Latonia nigriventer*	78.6	7.15	CR	53.72-87.86
22	*Rhynchobatus luebberti*	78.2	7.14	CR	51.55-78.47
23	*Rhynchobatus immaculatus*	78.2	7.14	CR	51.55-78.47
24	*Rhynchobatus australiae*	77.6	7.14	CR	45.59-77.91
25	*Microcycas calocoma*	77.0	7.13	CR	12.73-96.94
26	*Platysternon megacephalum*	76.7	7.13	CR	38.43-81.68
27	*Dermatemys mawii*	75.5	7.11	CR	64.73-111.02
28	*Carettochelys insculpta*	151.3	7.11	EN	55.34-152.27
29	*Planorbidella depressa*	74.3	7.09	CR	63.16-82.2
30	*Pentaphlebia gamblesi*	72.6	7.07	CR	53.23-73.05
31	*Sphaeromimus lavasoa*	72.4	7.07	CR	60.78-72.36
32	*Spelaeoecia bermudensis*	72.4	7.07	CR	64.38-88.86
33	*Colossobolus pseudoaculeatus*	71.9	7.06	CR	57.12-71.94
34	*Opisthopatus roseus*	71.2	7.05	CR	59.82-75.3
35	*Rhynchobatus cooki*	70.6	7.04	CR	41.5-70.85

36	*Sphaeromimus saintelucei*	69.3	7.03	CR	53.31-69.3
37	*Granitobolus endemicus*	69.1	7.02	CR	47.03-69.11
38	*Riotintobolus minutus*	69.1	7.02	CR	58-69.05
39	*Wollemia nobilis*	68.1	7.01	CR	33.04-107.97
40	*Margaritifera marocana*	67.0	6.99	CR	56.95-74.12
41	*Margaritifera hembeli*	66.4	6.98	CR	56.36-73.48
42	*Alligator sinensis*	65.7	6.97	CR	50.68-100.12
43	*Andrias davidianus*	65.6	6.97	CR	49.64-119.56
44	*Gigantopelta aegis*	65.4	6.97	CR	55.58-72.33
45	*Acroteriobatus variegatus*	64.7	6.96	CR	34.24-64.97
46	*Colossobolus litoralis*	64.2	6.95	CR	47.66-64.2
47	*Rhynchobatus laevis*	64.1	6.95	CR	48.39-64.45
48	*Rhynchobatus springeri*	64.1	6.95	CR	48.39-64.45
49	*Hymenophyllum helicoideum*	64.0	6.95	CR	54.41-71.42
50	*Spatoglossum schmittii*	64.0	6.95	CR	19.85-99.03
51	*Isoetes hewitsonii*	63.7	6.94	CR	54.6-89.85
52	*Pleiodon ovatus*	63.6	6.94	CR	54.36-70.4
53	*Phycolepidozia exigua*	63.2	6.93	CR	61.49-65.55
54	*Sphaeromimus andrahomana*	63.0	6.93	CR	52.44-63.01
55	*Acipenser nudiventris*	62.9	6.93	CR	17.69-81.6
56	*Spiromimus albipes*	62.9	6.93	CR	53.71-62.94
57	*Rhynchobatus djiddensis*	62.9	6.93	CR	41.62-63.23
58	*Isoetes sinensis*	62.9	6.93	CR	60.14-105.48
59	*Acipenser stellatus*	62.7	6.93	CR	18.3-74.42
60	*Pseudoscaphirhynchus kaufmanni*	62.7	6.93	CR	18.3-74.42
61	*Mahezomus apicoporus*	62.2	6.92	CR	48.16-68.24
62	*Sphaeromimus splendidus*	62.2	6.92	CR	47.84-62.18
63	*Leptogyra inflata*	61.5	6.91	CR	52.17-68.01
64	*Rhina ancylostoma*	61.3	6.9	CR	40.55-61.59
65	*Rhynchorhina mauritaniensis*	61.3	6.9	CR	40.55-61.59
66	*Araucaria angustifolia*	60.7	6.9	CR	20.62-100.22
67	*Euastacus yigara*	60.6	6.89	CR	11.92-68.93
68	*Vandiemenia ratkowskiana*	60.5	6.89	CR	50.28-61.22
69	*Colossobolus minor*	60.4	6.89	CR	46.47-60.4
70	*Perissolestes remus*	60.3	6.89	CR	42.22-73.2
71	*Austropotamobius pallipes*	120.9	6.88	EN	15.75-122.02
72	*Riotintobolus mandenensis*	59.5	6.88	CR	49.51-59.49
73	*Engaewa pseudoreducta*	58.9	6.86	CR	21.57-99.63
74	*Lirapex politus*	58.7	6.86	CR	49.83-64.96

75	*Galeorhinus galeus*	58.6	6.86	CR	1.51-71.86
76	*Branchinecta belki*	58.5	6.86	CR	53.63-67.38
77	*Garypus titanius*	57.7	6.85	CR	44.93-62.54
78	*Spiromimus litoralis*	57.5	6.84	CR	45.68-57.54
79	*Margaritifera auricularia*	57.5	6.84	CR	48.95-63.62
80	*Atya brachyrhinus*	57.2	6.84	CR	18.37-60.77
81	*Triops baeticus*	113.1	6.82	EN	24.74-133.91
82	*Lasmigona decorata*	55.4	6.81	CR	48.22-61.94
83	*Lamellomphalus manusensis*	55.3	6.8	CR	46.77-61.22
84	*Barbouria cubensis*	55.1	6.8	CR	32.28-55.15
85	*Alasmidonta raveneliana*	55.0	6.8	CR	47.83-61.44
86	*Spiromimus scapularis*	54.1	6.78	CR	45.82-54.1
87	*Rhombuniopsis tauriformis*	54.0	6.78	CR	47.21-60.71
88	*Hylekobolus andasibensis*	53.6	6.77	CR	46.61-53.6
89	*Sphaeromimus titanus*	53.5	6.77	CR	42.49-53.51
90	*Dictyota galapagensis*	53.5	6.77	CR	15.77-88.97
91	*Afrogarypus seychellesensis*	53.2	6.76	CR	40.85-59.06
92	*Branchinecta mexicana*	53.1	6.76	CR	51-59.13
93	*Scaphirhynchus suttkusi*	52.7	6.76	CR	18.66-99.63
94	*Latimeria menadoensis*	213.0	6.75	VU	70.89-315.03
95	*Pristis pristis*	52.2	6.75	CR	42.12-52.47
96	*Zaglossus attenboroughi*	51.8	6.74	CR	38.79-81.76
97	*Glaucostegus cemiculus*	51.6	6.74	CR	30.08-51.85
98	*Glaucostegus halavi*	51.6	6.74	CR	30.08-51.85
99	*Zaglossus bruijni*	51.6	6.73	CR	38.09-83.73
100	*Ostinobolus subterraneus*	51.3	6.73	CR	44.64-51.33

707

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.png
- FigureS5.png
- FigureS6.png
- SupplementaryTable1.pdf
- SupplementaryData1.xlsx