A phylogeny-informed characterization of global tetrapod traits addresses data gaps and biases

Mario R. Moura[1,2,3,4*], Karoline Ceron[1], Jhonny J. M. Guedes[5], Yanina V. Sica[2,3], Julie Hart[3,6], Wendy Dorman[3,7], Pamela Gonzalez-del-Pliego[3,8], Ajay Ranipeta[3], Alessandro Catenazzi[9], Fernanda P. Werneck[10], Luís Felipe Toledo[1], Natham S. Upham[11], João F. R. Tonini[12], Timothy J. Colston[13], Robert Guralnick[14], Rauri C. K. Bowie[15], R. Alexander Pyron[16], and Walter Jetz[2,3*]

1. Departamento de Biologia Animal, Universidade Estadual de Campinas, Campinas, SP 13083-862, Brazil
2. Department of Ecology and Evolutionary Biology, Yale University, New Haven CT, 06511, USA
3. Center for Biodiversity and Global Change, Yale University, New Haven CT, 06511, USA
4. Departamento de Biociências, Universidade Federal da Paraíba, Areia, PB 58397-000, Brazil
5. Programa de Pós-Graduação em Ecologia e Evolução, Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, GO 74690-900, Brazil
6. New York Natural Heritage Program, State University of New York College of Environmental Science and Forestry, Albany, NY 12233, USA
7. Department of Natural Resources and Environmental Sciences (NRES), University of Illinois Urbana-Champaign, Urbana, IL 61801, USA
8. Rui Nabeiro Biodiversity Chair, MED Institute, Universidade de Évora, 7000 Évora, Portugal
9. Department of Biological Sciences, Florida International University, Miami FL 33199, USA
10. Programa de Coleções Científicas Biológicas, Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM 69067-375, Brazil
11. School of Life Sciences, Arizona State University, Tempe, AZ 85284, USA
12. Department of Biology, University of Richmond, Richmond VA, 23173, USA
13. Biology Department, University of Puerto Rico at Mayagüez, Mayagüez, Puerto Rico, 00681
14. Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
15. Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, Berkeley CA 9470, USA
16. Department of Biological Sciences, The George Washington University, Washington DC, 20052, USA

*Corresponding author: Mario R. Moura (mariormoura@gmail.com), Walter Jetz (walter.jetz@yale.edu)

Abstract

Tetrapods (amphibian, reptiles, birds and mammals) have become a central model system for global ecology, conservation, and biodiversity science at large. But continuing data gaps, limited data standardisation, and ongoing flux in taxonomic nomenclature constrain integrative research on this group and potentially cause biased inference. We combined and harmonised taxonomic, spatial, phylogenetic, and attribute data with phylogeny-based imputation to provide a comprehensive data resource (TetrapodTraits) that includes values, predictions, and sources for 75 attributes. These attributes include body size, activity time, micro- and macrohabitat, geography, environmental preferences, human influence and threat status, for all 33,281 tetrapod species covered in recent fully sampled phylogenies. We assess gaps and biases across taxa and space, finding that shared data missing in attribute values increased with taxon-level completeness and richness across clades. Prediction of missing attribute values using multiple assessment methods revealed substantial changes in estimated

macroecological patterns. These results highlight particular biases incurred by non-random missingness and mechanisms to best address it. While there is an obvious need for further data collection and updates, our phylogeny- -informed database of tetrapod traits can support a better representation of undersampled species and their attributes in ecology, evolution, and conservation.

Introduction

Over the past two decades biodiversity science has seen a dramatic growth in impactful large-scale ecological, evolutionary, and conservation research enabled by near-global study systems such as terrestrial vertebrates, or Tetrapoda (amphibians, reptiles, birds, and mammals). These efforts usually rely on datasets spanning wide temporal, spatial, and taxonomic scales [1] that ideally are fully harmonized and well curated. Although terrestrial vertebrates are considered a well-known animal groups compared to invertebrates and plants [2,3], significant data gaps remain [4,5]. Consequently, trait-based research on biodiversity is often hampered by spatially and phylogenetically incomplete datasets.

Because this data missingness is usually highly non-random, i.e. overall trait space is represented unevenly [4], it can therefore have pervasive consequences for inference in trait-based research [6,7] and conservation applications [6,8]. Indeed, bias is recognized in the availability of trait data on certain taxa, regions, and traits [4], and missing values for a given variable are often associated with incompleteness in another, a phenomenon known as co-missingness or shared gaps [9]. Such congruent or aggregated (as opposed to segregated) patterns in trait missingness can arise from societal and research preferences towards either charismatic species [3] or taxonomic groups and geographical regions that are easily sampled [6,10], whereas segregated patterns may reveal traits and taxa that are difficult to sample or otherwise under-represented for a given clade. Documenting which groups show aggregated or segregated data missingness helps to inform new data acquisition strategies and reduce existing biases in trait databases.

While sampled trait data for many species attributes remain limited in Tetrapoda, new methods can help to minimise these data gaps and improve our understanding of biodiversity. Past practices have included the removal of species with missing data or replacement of missing values by observed averages, but these strategies may ultimately reduce statistical power and increase bias [7,9,11]. More recently, growth in large-scale phylogenetic analyses have boosted the development of methods to increase accuracy of imputing missing values [12–15]. Among tetrapods, recent large-scale applications of imputation methods include, for example, the use of either phylogenetic regression methods or machine-learning techniques to predict missing values in trait data for amphibians [16,17], reptiles [17,18], birds [17,19], and mammals [17,19,20], as well as to inform threat statuses for data deficient and non-assessed species [18,21].

We leveraged these new imputation approaches with additional data mobilization to provide a comprehensive database and assessment of key ecological attributes of all extant 33,281 tetrapod species covered in recent fully-sampled phylogenies, including 7,238 amphibians [22], 384 chelonians and crocodilians [18], 9,755 squamates and tuatara [23], 9,993 birds [24], and 5,911 mammals [25]. We assessed 75 standardised species-level attributes combining a range of sources on taxonomy, body size, activity time, microhabitat, macrohabitat, geography, environmental preferences, human influence, and threat status. Since not all species have yet genetic data (representing an important source of the remaining uncertainty about their placements in available phylogenetic topologies), we also evaluated completeness in genetic data, i.e. the remaining Darwinian shortfall [26]. We taxonomically harmonised data sources and filled gaps using phylogeny-based imputation methods [12–15] for which we verified the performance and remaining uncertainty through extensive validation procedures.

We use this combined dataset to assess the geographic, taxonomic, and trait-related biases and evaluate how their model-based closure supports improved information and biological inference. Because of the biodiversity knowledge paradox – the high biodiversity in the tropics [27] but the better sampling of taxa in temperate regions [4,5,28] – we expect larger unsampled fractions of the tetrapod trait space for species occurring in tropical versus temperate regions. Similarly, the high research capacity (i.e., infrastructure and expertise availability) towards birds and mammals relative to amphibians and reptiles [2,3] contribute to the uneven sampling of trait space [4]. We expect larger biases in trait data among historically undersampled taxa. Finally, species' biology and associated sampling methodologies are known to affect species detection and collection rates in the field [29–31]. For instance, species detectability is typically lower for small- than large-bodied species, and similarly so for nocturnal relative to diurnal species [32–34], whereas sampling methodologies often favour the collection and research of species living above ground compared to fossorial or arboreal ones [35–39]. We thus expect some convergent missingness across the tetrapod trait space, with undersampled species being typically small-sized, nocturnal, and fossorial or arboreal.

Methods

We curated and assembled available databases for global tetrapod groups and used the latest phylogeny-based methods to create the most comprehensive tetrapod attribute dataset to date. Our database, TetrapodTraits, includes various fields that track data sources and imputation techniques, which we refer to collectively as 'attributes', along with trait data. Our procedures can be summarised in five general steps: (i) data acquisition, (ii) taxonomic harmonisation, (iii) outlier verification, (iv) taxonomic imputation, and (v) phylogenetic imputation. Details on the validation of our predictions and imputation performance are available in the Supplementary Information (S1 Text).

Data Acquisition

We compiled 75 species-level attributes regarding taxonomy, body size, activity time, microhabitat, macrohabitat, threat status, geography, environmental preferences, and associated data sources for each tetrapod species. We gathered information from several global, continental, and regional databases, and complemented the existing data from published (articles, book chapters, field guides, and monographs) and grey literature (e.g., technical reports, government documents, monographs, theses), besides incorporating unpublished data gathered in fieldwork performed by some of us. To minimise the uneven representation of ecological attributes across clades, we initially identified genera and families whose species did not have data on body size, activity time, or microhabitat. We then used species belonging to these genera and families to carry out additional online searches on academic platforms (Google Scholar and Web of Science) and included complementary attribute data whenever possible. In our examination of the data sources, we did not use the imputations provided solely at the genus level (e.g., mean value per genus).

Briefly, taxonomic data were represented by higher-level taxonomic ranks (Class, Order, Family), species name, authority name, and year of description. Body size data consisted of information on body length (mm) and body mass (g). Additional details on body length measurements for different taxonomic groups are available in the Supporting Information. Activity time, as defined here, covers whether the species was diurnal and/or nocturnal. Cathemeral or crepuscular species were considered as both diurnal and nocturnal. Microhabitat included five categories of habitat use commonly reported in field guides and related literatures: fossorial, terrestrial, aquatic, arboreal, and aerial. Microhabitat data for birds were adapted from the EltonTraits database [40], which describes the estimated relative usage for seven types of foraging stratum. We reduced these seven categories to four by: summing the relative usage of species foraging below water surface or on the water surface in the aquatic microhabitat; summing the relative usage of species foraging on the ground and

below 2m in understorey as terrestrial; summing the relative usage of species foraging 2m upward in canopy and just above canopy as arboreal. Species with aerial microhabitat were kept as defined in EltonTraits database [40], and no fossorial bird was reported in the later source. We then made binary the relative usage of aquatic, terrestrial, arboreal, and aerial microhabitat using a threshold of 30% to consider a species as typical of a given microhabitat type. The microhabitat categories are not mutually exclusive, meaning that a species can be present in more than one category to represent intermediate microhabitats, such as semifossorial (which involves both fossorial and terrestrial categories) or semiarboreal (which combines terrestrial and arboreal).

Some attributes in the TetrapodTraits database were derived from expert-based range maps for amphibians [16,41,42], reptiles [18,27,41], mammals [41–43], and birds [24,42]. We matched the authoritative expert range maps for each of the tetrapod groups with the corresponding phylogenies and edited species ranges to ensure that they represented the species concept adopted in the corresponding phylogeny. Overall, our procedure for verification of the species range maps can be summarised in eight scenarios: i) no changes, where species range maps matched directly with binomials in the phylogeny, ii) synonyms, where species range maps were direct synonyms to binomials in the phylogeny, thus requiring therefore only an updated name, iii) no range, where extinct and domesticated species in the phylogeny did not need a range map, iv) split, where species range maps needed to be clipped from a parent species, or when parent species needed to have part of their range removed, vi) lumps, where species range maps needed to be combined with those of other species, vii) new species, where no range map was available, so we derived ranges from other sources in the literature or drew a minimum convex polygon with a buffer around type localities, and viii) species to remove, whose ranges needed to be removed because they were not in the phylogeny. We did not derive range maps for species with uncertain geographic distribution, those with reported occurrences only at the level of administrative units (e.g., country, state) without precise location information.

We measured range size as the number of 110 × 110 km equal-area grid cells intersected by each species, a spatial resolution that minimises the presence of errors related to the use of expert ranges maps [44–46]. We recorded the presence of a species in a grid cell if any part of the species distribution polygon overlapped with the grid cell. We then computed spatially derived attributes as the within-range average value of continuous variables (e.g., annual mean temperature , proportion of urban area) [47–50] weighted by species range occupancy in the cell, or as the proportion of the species range occupied by levels of categorical variables (e.g., proportion of the species range overlapped by each biogeographical realm). All within-range attributes were derived from spatial layers available online [47–50] and are individually described in the Results and Discussion section.

Species macrohabitat data followed the IUCN Habitat Classification scheme v. 3.1 [41]. Briefly, this scheme describes 17 major habitat categories in which species can occur, besides an 18th category for species with unknown major habitat. We used the *rredlist* package [51] to obtain macrohabitat data directly from IUCN for 31,464 species (94.5% of all species herein considered). Our macrohabitat variables correspond only to the first level of IUCN Habitat Classification scheme. We extracted macrohabitat data on another 157 species from the pertinent literature. We also used the *rredlist* package [51] to obtain non-DD assessed status for 28,885 tetrapod species based on IUCN red list v. 2021-3 [41]. For 286 species not available via *rredlist*, we used non-DD assessed statuses matching those described in previous works using the same taxonomy of fully-sampled trees [16,18,23,52]. Data deficient species totalled 3,476 species. We did not find an assessed status for 646 species. Overall, the data acquisition involved the inspection of more than 2,600 references (Table 1). A full description of attributes and further details on all sources consulted for each species are provided in the Results and Discussion section.

Taxonomic Harmonisation

The taxonomy of the TetrapodTraits database follows the respective taxonomies of the recent, fully-sampled phylogenies for each of the major tetrapod groups [18,22–25]. The amphibian phylogeny taxonomy [22] follows the 19 February 2014 edition of AmphibiaWeb (http://amphibiaweb.org), with 7,238 species. The phylogeny for chelonians [18] follows the Turtles of the World (8th ed.) checklist [53], with the addition of *Aldabrachelys abrupta* and *Al. grandidieri*, and the synonymisation of *Amyda ornata with Am. cartilaginea*, adding up to 357 chelonian species, whereas for crocodilians [18], the taxonomy followed [54], complemented by the revalidation of *Crocodylus suchus* [55] and *Mecistops leptorhynchus* [56], and the recognition of three *Osteolaemus* species [57,58], resulting in 27 species. The taxonomy of the squamate and tuatara phylogeny [23] follows the Reptile Database update of March 2015 (http://www.reptile-database.org) with 9,755 species. For brevity, we refer hereafter to species in the latter phylogeny as squamates, although we recognise that the tuatara, *Sphenodon punctatus*, is not a squamate. The taxonomy of the bird phylogeny [24] followed the Handbook of the Birds of the World [59], including 9,993 species. Finally, the taxonomy of the mammal phylogeny [25] follows the IUCN [60] with modifications resulting in a net addition of 398 species, and included 5,911 species.

To maximise data usage from previous compilation efforts, and to ensure coherence among species names and the multiple data sources, we built lists of synonymies and valid names based on multiple taxonomic databases [41,61,62], and extracted the unique synonyms in each taxonomic database. By unique synonym, we refer to a binomial (scientific name valid or not) applied to only one valid name. We then performed taxonomic reconciliation based on four steps:

1. ***Direct match with data sources:*** We directly paired the names of each of the 33,281 species in TetrapodTraits with the potential source of the data. Species-level attributes of closely related species could appear as identical values if the attributes had been extracted from sources in which different species were treated as synonymies. We minimised the inclusion of duplicated values by flagging each taxonomic match between TetrapodTraits and external data sources to ensure that each data entry was made only once..
2. ***Direct match with data source synonymies:*** For the species we were unable to directly match with data source in step 1, we updated the taxonomy using the list of unique synonymies, and then performed a new matching operation which allowed us to extract and flag additional data whenever possible..
3. ***Direct match with TetrapodTraits synonymies:*** Some data sources may follow more recent taxonomies than those inherited from the fully-sampled phylogenies [18,22–25]. For species without a direct match after step 2, we updated their taxonomy in the TetrapodTraits using the list of unique synonymies and repeated the data extraction and flagging procedures.
4. ***Manual verification:*** For species without a direct match after step 3, we searched specialised taxonomic databases (amphibians [61], reptiles [62], birds [63], and mammals [64]) manually for potential spelling errors and/or additional synonymies not yet included among our synonym lists. Whenever possible, we updated the taxonomy applied to data sources and then repeated the data extraction and flagging procedure. Species without data after the completion of step-4 were classified as missing data.

Outlier Verification

We implemented two approaches to detect potential inconsistencies in continuous attribute data, here body length and body mass, before applying phylogeny-based methods to impute missing values.

1. ***Interquartile Range Criterion:*** Species body length and body mass were $\log_{10}$ transformed and then flagged as outliers if their value were outside the interval defined between $[q_{0.25} - 1.5 \times IQR]$ and $[q_{0.75} + 1.5 \times IQR]$, where $q_{0.25}$ and $q_{0.75}$ are respectively the first and third quartiles, and IQR is the interquartile range $[q_{0.25} - q_{0.75}]$.
2. ***Deviation from allometric relationship:*** Although allometric escape is a phenomenon observed in nature, we used interactive scatterplots to flag species with unusual deviations from the expected allometric relationship between body length and mass. We inspected allometric relationships separately for species within each Class, Order, and Suborder.

For species flagged in steps one or two above, we checked body length and/or body mass for validity and corrected these values where necessary. Those that could not be confirmed using a reliable source, we purged from the database.

Taxonomic Imputation

The global scope of the present database inevitably includes some gaps that are hard to fill, e.g., natural history data for species known only from the holotype or a few specimens [65,66]. Previous studies have addressed this challenge by using values imputed at the level of genus or from close-relatives [40] to reduce data missingness. Although these earlier strategies of 'taxonomic imputation' might artificially reduce variability in attribute values, they are useful for filling gaps in highly conserved attributes, and can ultimately help increase the performance of phylogeny-based imputation methods applied in concert with correlated attribute data [67,68].

We used taxonomic imputations for two cases of missing data in microhabitat data: chiropterans, who were considered 'aerial' (113 species), and dolphins and whales who were considered 'aquatic' (5 species). For the remaining tetrapod species, we computed the per-genus proportion of species in each type of activity time (diurnal or nocturnal), microhabitat (fossorial, terrestrial, aquatic, arboreal, aerial), and macrohabitat (17 binary variables informing the IUCN Habitat Classification scheme). If a type of activity time, microhabitat, or macrohabitat appeared in at least 70% of species in the genus, we assumed this ecological attribute was also present among species with missing data in the respective genus. Our goal was to reduce missing values in activity time, microhabitat, and macrohabitat for groups with well-known ecologies (observed data available for at least 70% of species) before running phylogeny-based imputation methods (Table 1). We did not use taxonomic imputation for continuous attributes.

Phylogenetic Imputation

To minimise missing values in our database we used two imputation techniques demonstrated as robust in recent methodological comparisons [12,13], namely *Rphylopars* [14,15] and *missForest* [13]. The *Rphylopars* method uses the phylogenetic tree directly to impute missing values based on the phylogenetic covariances and correlated evolution between numeric attributes, whereas *missForest* is based on a random forest algorithm and allows data of mixed types.

Because of the large dimensionality of data in our analysis, we performed imputations only for body length, body mass, microhabitat, and activity time. All spatially derived attributes showed more than 99% data completeness (Table 1). We did not perform phylogeny-based imputations for species which lacked range maps and hence, spatially derived attributes. We $\log_{10}$-transformed the continuous variables (body length and body mass) to reduce skewness, and used binary variables to represent each type of microhabitat (fossorial, terrestrial, aquatic, arboreal, aerial) and activity time (diurnal and nocturnal). Note that microhabitat and activity time types are non-mutually exclusive. To account for uncertainty in fully-sampled phylogenies, we used 100 trees of each tetrapod group [18,22–25] to compute imputations.

We ran *Rphylopars* using the Brownian motion (BM) model of trait evolution and extracted the median value across the 100 trees for each imputed attribute per species. Because *Rphylopars* allows only continuous attributes [15], the values imputed for binary variables could assume values outside the range between 0–1. Thus, for each tetrapod group and binary attribute, we 7inarized imputed types of microhabitat and activity time using the 0.5 threshold.

The *missForest* algorithm does not directly include a phylogenetic tree in computations [13]. To account for phylogenetic information, we selected the first 10 eigenvectors from the phylogenetic covariance matrix of each fully-sampled tree and used these, along with the observed attribute data in the imputation routine. For each tetrapod group and across the subset of 100 trees, the selected phylogenetic eigenvectors explained on average 96.8% of the phylogenetic variation across the 100 trees of each tetrapod group (97.8% for amphibians, 99.4% for chelonians and crocodilians, 97.1% for squamates, 92.4% for birds, and 97.6% for mammals). For continuous attributes (body length and body mass), we extracted the median imputed value per species across the 100 trees. Imputed binary attributes (types of microhabitat and activity time) had their frequency computed across the 100 trees, with frequencies later made binary using the 0.5 threshold.

For birds only, we complemented the observed attribute data with 10 morphometric traits ($\log_{10}$-transformed) made recently available through the AVONET database [69], and then repeated the imputation routine using the subset of 100 fully-sampled bird trees. Given the high dimensionality of these computations, they became computationally intractable under the *Rphylopars* approach and were only performed through *missForest*. The model parameters used in *missForest* analysis for all tetrapod groups included: 500 trees to grow each decision forest (argument *ntree*), the square-root of total number of predictors as the number of variables randomly sampled at each split (default of the argument *mtry*), and a maximum of 10 iterations for *missForest* to reach the stopping criterion (default of the argument *maxiter*).

We selected the imputed values returned by the imputation approach (*Rphylopars* or *missForest*) identified as most suitable (see Supplementary Information, S1 Text) for each combination of tetrapod group (amphibians, chelonians and crocodilians, squamates, birds, and mammals) and attribute data (body length, body mass, microhabitat, and activity time). Additional fields were used to clearly label all imputations included in the TetrapodTraits database. Computations were performed in R v. 4.1.0 using the packages *Rphylopars* [15], *ape* [70], *gmodels* [71], and *missForest* [13]. Imputed values for both approaches are provided in the Data Availability section.

Table 1. Summary of the major groups of attributes, species counts, and sources in TetrapodTraits database. 'Imputed' counts indicate the number of species computed with taxonomic (Taxon.) or phylogenetic (Phylog.) imputations. References used for data acquisition sources are shown for each attribute and tetrapod group.

Attribute group	Number of species			Number of sources	Major Sources per Tetrapod group				
	Obs. Data	Taxon. Imputed	Phylog. Imputed		Amphibians	Turtles and Crocs	Squamates	Birds	Mammals
Body length	24,472	0	8,809	1559	[72–76]*	[18]*	[77]*	[78–80]*	[20,78]*
Body mass	26,613	0	6,668	460	[72,73,81–84]*	[18]*	[77]*	[40,69,80]*	[20,40,78,85–88]*
Activity time	24,512	588	8,181	347	[72,89–93]*	[94]*	[95–106]*	[40]*	[20,40,82,83,87]*
Microhabitat	31,470	1,125	686	688	[72,89,107]*	[18]*	[95–98,100–112]*	[40,107]*	[20,40,107]*
Macrohabitat	31,621	1,440	0	140	[41,42]*	[41,42]*	[41,42]*	[41,42]*	[41,42]*
Geography	33,125	0	0	7	[16,41,42]	[18,41,42]	[27]	[24,42]	[17,19, 22–26]
Environment	33,125	0	0	3	[47,48]	[47,48]	[47,48]	[47,48]	[47,48]
Human Influence	33,125	0	0	3	[49,50]	[49,50]	[49,50]	[49,50]	[49,50]
Threat Status	32,647	0	0	6	[16,41,113]	[18,41,114]	[41,115]	[41,52]	[41,116]

*For brevity, references listed above include only sources that contributed with at least 30 species; however, the raw files of TetrapodTraits include references for all species-specific sources.

Patterns of shared missing data

We assessed co-occurrence patterns in missing data across species using the 'checkerboard score' (C-score; [117]), which is less prone to type II errors than other co-occurrence metrics computed under a null model approach [118]. The C-score was based on a binary presence-absence matrix (PAM) of species (columns) by missing attributes (rows); the latter are represented by five binary variables informing the missingness of observed values in body length, body mass, activity time, microhabitat, and threat status. Missing data in threat status were represented by Data Deficient (DD) or non-assessed species. We computed the C-score using individual PAMs for each genus and family with at least 2 species, and each pairwise combination of attribute variable, following recommendations by [119].

To verify whether an observed C-score differed from the value expected by chance, we built a null distribution of C-score values using a randomization algorithm in which attribute completeness (rows sums) remained fixed while the probability of showing missing attribute values was considered equal for all species. Null distributions were built for each individual PAM using 10K iterations with a burn-in of 500. We then computed the standardised effect-size (SES) of the C-score and associated p-values and identified the pairwise attribute combinations with an aggregated (SES C-score < 0 and $p < 0.05$) or segregated (SES C-score > 0 and $p < 0.05$) co-occurrence pattern of missing data per taxa. Computations were performed using the *EcoSimR* [120] package in R.

Patterns and Biases in Imputed Traits

If certain parts of the multivariate attribute space are missing, the inclusion of imputed values may fill such 'invisible fractions' and change emergent attribute properties per taxon or geographical assemblage (e.g., average value, geometric mean, proportion or prevalence of a certain category). But changes in attribute property per taxon can also arise from the inclusion of biased imputed values [67], since imputation errors can increase either with the percentage of missing data or when multiple attributes are imputed simultaneously [12,67,68]. To verify how the inclusion of imputed values affected attribute patterns per taxa, we calculated the per-genus and per-family relative change in attribute property after adding imputed values.

For continuous attributes (body length and mass), the attribute property was represented by the geometric mean. Geometric mean is less sensitive to outliers (which greatly impact arithmetic mean) and to fluctuations in the position value of the median. For binary attributes, we scored types of microhabitat and activity time into continuous metrics of verticality (0 = strictly fossorial, 0.25 = fossorial and terrestrial, 0.5 = terrestrial or aquatic, 0.75 = terrestrial and arboreal, and 1 = strictly arboreal or aerial) and nocturnality (0 = strictly diurnal, 0.5 = cathemeral or crepuscular, 1 = strictly nocturnal). We then extracted the mean verticality and nocturnality per taxon and geographical assemblage. Relative change in attribute values was represented by the ratio of attribute property with imputations (i.e., mean computed using observed + imputed data) to attribute property without imputations (i.e., mean computed using observed data only):

$$Relative\ Change = 1 - (Attribute\ Property_{obs+imp} / Attribute\ Property_{obs})$$

Positive values of relative change indicate the proportional increase in attribute property after inclusion of imputed values, whereas negative values show the opposite. We used the R function *tree.merger* of the *RRphylo* package [121] to build a tetrapod super tree by combining the available global phylogenies for amphibians [22], chelonians and crocodilians [18], squamates [23], birds [24], and mammals [25]. The full tetrapod tree topology was only used for visualization purposes. In the presence of severe data limitations (e.g., very low

completeness, highly uneven sampled data, and lack of inter-correlated attributes), the relative changes in attribute properties could arise due to estimation errors instead of representing unsampled fractions of the attribute space. We therefore used Kruskal-Wallis tests to assess whether the relative increase or decrease of attribute properties differed with respect to the number of simultaneously imputed attributes.

Results and Discussion

Data Completeness

Our assessment of data gaps focused on five attributes with considerable missing data: body length, body mass, activity time, microhabitat, and threat status. Across these attributes, we observed complete species-level data for 40.7% of tetrapod species (n = 13,540), 31.5% of genera (n = 1,609), and 19.7% of families (n = 102). Seven tetrapod genera showed no data for these five attributes, all seven of which were monotypic rodents (*Calassomys*, *Halmaheramys*, *Musseromys*, *Paucidentomys*, and *Waiomys*) or bats (*Dryadonycteris* and *Hsunycteris*). No family showed zero completeness for these five attributes combined. The lowest completeness levels were among attributes related to body length (73.5%), activity time (73.6%), body mass (80%), and microhabitat (94.5%, Table S1-S2). While assessed threat status was available for 98% of tetrapod species, this number decreased to 87.6% when DD species were treated as missing data (Fig. 1).

At the level of geographic assemblages, the attribute completeness of body length showed large variation, with Neotropical and Afrotropical assemblages presenting the lowest completeness (Fig. 2A). The low spatial variability in attribute completeness for body mass, activity time, microhabitat, and assessed threat status (Fig. 2B–E) indicates that species with missing data are mostly narrow-ranged [4], which can limit the influence of these gaps in assemblage-level patterns. Of the 6,893 species missing at least two of the five attributes (body length, body mass, activity time, microhabitat, and threat status), 77% occurred in 10 or fewer 110 × 110 km equal-area grid cells. If we restrict missing data to three or more attributes, then 87% of species filtered occur in ≤10 grid cells. The completeness of genetic data in the fully-sampled trees of tetrapods [18,22–25] also showed high variation in spatial coverage (Fig. 2F), with Afrotropical and Indomalayan realms emerging with the highest Darwinian shortfall [26]. Overall, our findings help illustrate that geographical patterns in data coverage can be strongly influenced by data for wide-ranged species.

Patterns of shared missing data

We found non-random co-occurrence patterns in data gaps for 20% of the pairwise attribute combinations at genus-level and for 43% at family-level. Almost all non-random co-occurrence patterns showed aggregation of missing values, that is, missing data was shared among multiple attributes for the same species, particularly among squamates and amphibians (Fig. 3A). In our examination of shared gaps between threat status and other attributes, we found leaf-litter amphibians (*Craugastor*, *Eleutherodactylus*, and *Pristimantis*) standing out among tetrapod genera with highest aggregation in missing values (Fig. 3B), and closely followed by fossorial and semifossorial snakes (*Atractus* and *Tantilla*). These findings support the view of amphibians and squamates as being the least researched tetrapod classes [2,3], and also indicate the uneven sampling coverage of attribute space within certain clades.

Highly aggregated patterns of data gaps were more common among species-rich clades with a lower proportion of missing data (Figs 3 and S5). That is, shared gaps were strongest in species-rich and well-sampled taxa, particularly in those holding many species known only from their type locality or holotype [65,122]. About 86% of the genera and 74% of families within the top 50 taxa in shared gaps (Figs 3 and S5) include ‘lost taxa’, that is, species that have not been reliably observed in >50 years but are not yet declared extinct [66]. While the rediscovery of such rarely observed species is not uncommon [122], their absence in attribute databases will not be easily filled through field observations [123]. The data gaps resulting

from these 'lost taxa' prevent their inclusion in trait-based research, potentially harming biogeographical research and conservation practice.

Biases uncovered by imputation-based gap filling

We found that the use of imputations to fill gaps resulted in substantial changes in aggregate attribute properties and the insights they would support. These changes occurred in both directions and were strongest in continuous attributes (body length and body mass). About 65% of the tetrapod genera and 78% of the tetrapod families include at least one species with phylogeny-based imputations. After the inclusion of imputed values, the average attribute value per genus decreased in 5.9% of the attribute-genus combinations, and increased in 5% of these combinations. At the family-level, after the inclusion of imputed attributes, 14% of the attribute-family combinations showed an increased in the average values whereas 20% decreased (Fig. 4).

Changes in attribute properties at the level of geographic assemblages revealed substantial changes in the average body size in tropical regions (Fig. 5A-F). Such differences are largely absent in more temperate regions, reflecting the dedicated mobilization and the resulting large availability of natural history data throughout North America, Europe, and Asia [2,4,124]. Average nocturnality of tetrapod assemblages tended to strongly increase after the inclusion of imputed values, pointing to a global knowledge bias from lack of observed attribute data (Fig. 5H-I), particularly for squamates and to a lesser extent, amphibians (Figs S6-S11). We attribute this to both the low detectability of nocturnal species [125,126] and to the general lack of data for tropical regions where nocturnal species predominate [127]. Changes in average verticality were less pronounced across tetrapod assemblages (Fig. 5K-L), which is somewhat expected given the high completeness of microhabitat data. Even so, fossorial and arboreal species were responsible for 53% of imputed microhabitat data although they represented about 44% of non-imputed microhabitat categories.

Although imputation biases can increase in the absence of additional correlated attributes [67], we found no relationship between the number of imputed attributes and changes in average attributes after imputation-based gap-filing (Fig. S12). That is, changes in average attributes were not higher in genera and families that included species with multiple missing attributes. Moreover, changes in average attributes were either mostly unrelated with shared gaps (Fig. S13) or occurred in opposite directions, with low shared gaps leading to greater changes.

Broadly, we found that the magnitude of change in average attributes from imputation-based gap-filling was associated with both completeness and richness (Figs S14-S16). Those taxa and assemblages with great incompleteness and with fewest species saw the most pronounced changes in average attribute values. Lower attribute completeness creates more opportunities for variation in attribute space after imputation-based gap-filing, whereas low species richness can enhance the importance of imputed values in the sample.

TetrapodTraits – a comprehensive, imputation informed attributes database

The full phylogenetically coherent database we developed, TetrapodTraits, is being made publicly available to support a range of research applications in ecology, evolution, and conservation and to help minimize the impacts of biased data in this model system. The database includes 75 fields registering observed and imputed species-level attributes linked to their respective sources across 33,281 tetrapod species. Entries across this database show at least 98% of completeness after inclusion of imputed values and are organized in 9 major groups (Table 1). These are summarised and characterised in greater detail in Table S1.

1. **Taxonomy –** includes 8 attributes that inform scientific names and respective higher-level taxonomic ranks, authority name, and year of species description. Field names: Scientific.Name, Genus, Family, Suborder, Order, Class, Authority, and YearOfDescription.

2. **Tree –** includes 2 attributes that notify which fully-sampled phylogeny contains the species, along with whether the species placement was imputed or not in the phylogeny. Field names: TreeTaxon, TreeImputed.

3. **Body size –** includes 7 attributes that inform length, mass, and data sources on species sizes, and details on the imputation of species length or mass. Field names: BodyLength_mm, LengthMeasure, ImputedLength, SourceBodyLength, BodyMass_g, ImputedMass, SourceBodyMass.

4. **Activity time –** includes 5 attributes that describe period of activity (e.g., diurnal, fossorial) as dummy (binary) variables, data sources, details on the imputation of species activity time, and a diurnality score. Field names: Diu, Noc, ImputedActTime, SourceActTime, Nocturnality.

5. **Microhabitat –** includes 8 attributes covering habitat use (e.g., fossorial, terrestrial, aquatic, arboreal, aerial) as dummy (binary) variables, data sources, details on the imputation of microhabitat, and a verticality score. Field names: Fos, Ter, Aqu, Arb, Aer, ImputedHabitat, SourceHabitat, Verticality.

6. **Macrohabitat –** includes 19 attributes that reflect major habitat types according to the IUCN classification, the sum of major habitats, data source, and details on the imputation of macrohabitat. Field names: MajorHabitat_1 to MajorHabitat_10, MajorHabitat_12 to MajorHabitat_17, MajorHabitatSum, ImputedMajorHabitat, SourceMajorHabitat.

7. **Geography –** includes 12 attributes that present species range size, average latitude and longitude, and the proportion of species range within each biogeographical realm. Field names: RangeSize, Longitude, Latitude, Afrotropic, Australasia, IndoMalay, Neartic, Neotropic, Oceania, Paleartic, Antartic.

8. **Environment –** includes 5 attributes that characterise the average environmental condition within the species geographic range. Field names: AnnuMeanTemp, AnnuPrecip, TempSeasonality, PrecipSeasonality, Elevation.

9. **Human Influence –** includes 10 attributes that indicate the within-range average or proportion of human-related activities, assessed status and sources. Field names: ETA50K, HumanDensity, PropAnthroHyde, PropUrbanArea, PropCroplandArea, PropPastureArea, PropRangelandArea, PropGrazingArea, AssessedStatus, SourceStatus.

Conclusion

The growth in mobilized attribute data over the last several decades has enabled significant progress and increased geographic and taxonomic generality in comparative biology [82,128,129], macroevolution [130], macroecology [131], and conservation biology [132–134]. Despite the richness of available data, research outcomes and their interpretation often remained constrained by non-randomly missing data [6,8,68], and by the lack of standardisation across data dimensions (trait, phylogenetic, spatial) and data sources [135]. We show a clear latitudinal bias in the sampling of genetic data, with limited availability of genetic samples towards the Equator. In concert with taxonomic updates [136,137], these issues have imposed significant limits on multi-taxon investigations and global biodiversity synthesis [138–140]. We documented remaining global data gaps for key traits and then used imputation to fill these gaps, revealing the biases in our knowledge due to the missingness of attribute data. By using phylogeny-based imputation methods in concert with inter-related

ecological attributes, we uncovered the hitherto invisible part of attribute space. Our findings confirmed the predominance of attributes share missing data and provide a comprehensive assessment of gaps and biases across tetrapod groups.

Attributes predictions for species with missing data brought to light key correlates and allowed us to identify the significant consequences these gaps can have for inference. For instance, larger body size has been related to both high species detectability [32,36,125] and discovery probability [141–143], larger research effort and public interest [38,39,144,145], increased availability of assessed status [146,147], and better data sampling. Despite the difficulties associated with researching species of large body size – e.g., high sensitivity to disturbances [148,149], lower abundance [150,151], challenges with collection or transport – the presence of multiple data biases towards small-sized species confirm the struggle of biodiversity scientists to accurately characterise the smallest organisms in nature. In summary, different knowledge shortfalls [139] can share similar causes.

Biodiversity knowledge shortfalls are also clearly related to activity time and microhabitat. Nocturnal species often show lower detectability [125], and typically are less researched [38] and underassessed with respect to threat status [146], which helps explain why data on diurnal species are more likely to have activity time data. In addition, the higher detectability of terrestrial than arboreal or fossorial species via standard sampling methods [36,152] can lead to less biased occurrence data [153] and more research effort [28,38]. Although foraging stratum may constrain certain aspects of our biodiversity knowledge, verticality has a limited influence on the availability of microhabitat data.

Attribute imputations can be limited and biased [67,68]. For tetrapods generally and this study specifically, these concerns were minimized by the already substantial taxon-specific datasets and by the gap-oriented mobilization of additional data. We had to impute all four focal attributes (body length, body mass, activity time, and microhabitat) in only 276 tetrapod species (0.8%). No family had the four focal attributes missing in all species, and only eight genera (0.15% of total genera, which cumulatively include 13 species) were missing the four focal attributes. Two important diagnostics offered strong support for the robustness of our attribute predictions. First, there was no association between the number of imputed attributes and changes in average attributes after imputation-based gap-filling (Fig. S12). Second, there was no positive association between shared gaps in attribute data and changes in average attribute values (Fig. S13). We encourage researchers to carefully consider how assumptions and uncertainty of any imputed data might affect downstream use and interpretation, and the information we developed herein is no exception. For many tetrapod research cases; however, we expect the inclusion of species with imputed data to be preferrable to the wholesale exclusion of the data-limited species.

We have demonstrate that even for tetrapod species, a central model system in biodiversity science, our ecological knowledge attribute data are lacking consistently across certain types of traits, taxa, and regions. Through careful imputation we uncover that these gaps have brought about attribute distributions that have strong geographical and taxonomic biases. We expect that our approach and likely our general insights regarding biodiversity knowledge shortfalls will transcend taxa and systems. The consequences of biased data can percolate through trait-based metrics and comparative analyses, and may ultimately mislead research on functional diversity [8,17,154], extinction risk [6,19,155], and ecogeography [156–158]. The new phylogeny-based attribute database we have constructed can help minimise the impact of biased data and support new avenues of research in tetrapod conservation, ecology and evolution.

Supporting Information

The TetrapodTraits database is available for reviewers at vertlife.org. in a compressed (.zip) folder containing the following files, all using UTF-8 encoding:

- TetrapodTraits_v1.csv
- TetrapodTraits_v1_rphylopars_imputations.csv
- TetrapodTraits_v1_missforest_imputations.csv
- TetrapodAssemblages_110km.csv
- ReadMe_Metadata.txt

Code Availability

The Tetrapod Traits database and the R-code used for data validation will be made available on vertlife.org and GitHub upon acceptance for publication.

Acknowledgements

We deeply thank the many authors of the numerous sources consulted during the preparation of this database. This work was possible only because of their valuable efforts. We gratefully acknowledge support from the following sources: São Paulo Research Foundation (FAPESP) for grants supporting MRM (#2021/11840-6 and #2022/12231-6), LFT (#2016/25358-3) and KC (#2020/12558-0); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the fellowship to JJMG; Conselho Nacional de Desenvolvimento Científico - CNPq for research grants in support of FPW (#311504/2020-5) and LFT (#302834/2020-6); U.S. National Science Foundation (NSF) for grants supporting RAP (DEB-1441719), RCKB (DEB-1441652), and WJ (DEB-1441737 and DEB-1441719). WJ also acknowledges support from NASA grants 80NSSC17K0282 and 80NSSC18K0435.

Author Contributions

Conceptualization: Mario R. Moura and Walter Jetz

Data Curation: Mario R. Moura, Karoline Ceron, Jhonny J. M. Guedes, Yanina V. Sica, Julie Hart, Wendy Dorman, Pamela Gonzalez-del-Pliego, Ajay Ranipeta, Alessandro Catenazzi, Fernanda P. Werneck, Luís Felipe Toledo, Natham S. Upham, Timothy J. Colston

Formal Analysis: Mario R. Moura

Funding Acquisition: R. Alexander Pyron, Rauri C.K. Bowie, Walter Jetz

Investigation: Mario R. Moura

Methodology: Mario R. Moura

Software: Mario R. Moura

Supervision: Walter Jetz

Visualization: Mario R. Moura and Walter Jetz

Writing – original draft preparation: Mario R. Moura

Writing – review & editing: Mario R. Moura, Karoline Ceron, Jhonny J. M. Guedes, Yanina V. Sica, Alessandro Catenazzi, Fernanda P. Werneck, Luís Felipe Toledo, João F.R. Tonini, Timothy J. Colston, Rauri C. K. Bowie, Robert Guralnick, R. Alexander Pyron and Walter Jetz

Competing interests

The authors declare no competing financial interests.

References

1. Beck J, Ballesteros-Mejia L, Buchmann CM, Dengler J, Fritz SA, Gruber B, et al. What's on the horizon for macroecology? Ecography (Cop). 2012;35: 673–683. doi:10.1111/j.1600-0587.2012.07364.x
2. Titley MA, Snaddon JL, Turner EC. Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. Schierwater B, editor. PLoS One. 2017;12: e0189577. doi:10.1371/journal.pone.0189577
3. Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. Taxonomic bias in biodiversity data and societal preferences. Sci Rep. 2017;7: 9132. doi:10.1038/s41598-017-09084-6
4. Etard A, Morrill S, Newbold T. Global gaps in trait data for terrestrial vertebrates. Sheard C, editor. Glob Ecol Biogeogr. 2020;29: 2143–2158. doi:10.1111/geb.13184
5. Oliver RY, Meyer C, Ranipeta A, Winner K, Jetz W. Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. Moritz C, editor. PLOS Biol. 2021;19: e3001336. doi:10.1371/journal.pbio.3001336
6. González-Suárez M, Lucas PM, Revilla E. Biases in comparative analyses of extinction risk: mind the gap. J Anim Ecol. 2012;81: 1211–1222. doi:10.1111/j.1365-2656.2012.01999.x
7. Kim SW, Blomberg SP, Pandolfi JM. Transcending data gaps: a framework to reduce inferential errors in ecological analyses. Ecol Lett. 2018;21: 1200–1210. doi:10.1111/ele.13089
8. Pakeman RJ. Functional trait metrics are sensitive to the completeness of the species' trait data? Oksanen J, editor. Methods Ecol Evol. 2014;5: 9–15. doi:10.1111/2041-210X.12136
9. Nakagawa S, Freckleton RP. Missing inaction: the dangers of ignoring missing data. Trends Ecol Evol. 2008;23: 592–596. doi:10.1016/j.tree.2008.06.014
10. Sandel B, Gutiérrez AG, Reich PB, Schrodt F, Dickie J, Kattge J. Estimating the missing species bias in plant trait measurements. J Veg Sci. 2015;26: 828–838. doi:10.1111/jvs.12292
11. Taugourdeau S, Villerd J, Plantureux S, Huguenin□Elie O, Amiaud B. Filling the gap in functional trait databases: use of ecological hypotheses to replace missing data. Ecol Evol. 2014;4: 944–958. doi:10.1002/ece3.989
12. Penone C, Davidson AD, Shoemaker KT, Di Marco M, Rondinini C, Brooks TM, et al. Imputation of missing data in life-history trait datasets: which approach performs the best? Freckleton R, editor. Methods Ecol Evol. 2014;5: 961–970. doi:10.1111/2041-210X.12232
13. Stekhoven DJ, Buhlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28: 112–118. doi:10.1093/bioinformatics/btr597
14. Bruggeman J, Heringa J, Brandt BW. PhyloPars: estimation of missing parameter values using phylogeny. Nucleic Acids Res. 2009;37: W179–W184. doi:10.1093/nar/gkp370
15. Goolsby EW, Bruggeman J, Ané C. Rphylopars□: fast multivariate phylogenetic comparative methods for missing data and within-species variation. Fitzjohn R, editor. Methods Ecol Evol. 2017;8: 22–27. doi:10.1111/2041-210X.12612
16. González-del-Pliego P, Freckleton RP, Edwards DP, Koo MS, Scheffers BR, Pyron RA, et al. Phylogenetic and Trait-Based Prediction of Extinction Risk for Data-Deficient Amphibians. Curr Biol. 2019;29: 1557-1563.e3. doi:10.1016/j.cub.2019.04.005
17. Carmona CP, Tamme R, Pärtel M, de Bello F, Brosse S, Capdevila P, et al. Erosion of global functional diversity across the tree of life. Sci Adv. 2021;7. doi:10.1126/sciadv.abf2675
18. Colston TJ, Kulkarni P, Jetz W, Pyron RA. Phylogenetic and spatial distribution of evolutionary diversification, isolation, and threat in turtles and crocodilians (non-avian archosauromorphs). BMC Evol Biol. 2020;20: 81. doi:10.1186/s12862-020-01642-3

19. Cooke RSC, Eigenbrod F, Bates AE. Projected losses of global mammal and bird ecological strategies. Nat Commun. 2019;10: 2279. doi:10.1038/s41467-019-10284-z
20. Soria CD, Pacifici M, Di Marco M, Stephen SM, Rondinini C. COMBINE: a coalesced mammal database of intrinsic and extrinsic traits. Ecology. 2021;102: e03344. doi:10.1002/ECY.3344/SUPPINFO
21. Caetano GHO, Chapple DG, Grenyer R, Raz T, Rosenblatt J, Tingley R, et al. Automated assessment reveals that the extinction risk of reptiles is widely underestimated across space and phylogeny. Jordano P, editor. PLOS Biol. 2022;20: e3001544. doi:10.1371/journal.pbio.3001544
22. Jetz W, Pyron RA. The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. Nat Ecol Evol. 2018;2: 850–858. doi:10.1038/s41559-018-0515-5
23. Tonini JFR, Beard KH, Ferreira RB, Jetz W, Pyron RA. Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. Biol Conserv. 2016;204: 23–31. doi:10.1016/j.biocon.2016.03.039
24. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. The global diversity of birds in space and time. Nature. 2012;491: 444–448. doi:10.1038/nature11631
25. Upham NS, Esselstyn JA, Jetz W. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. Tanentzap AJ, editor. PLOS Biol. 2019;17: e3000494. doi:10.1371/journal.pbio.3000494
26. Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM. Darwinian shortfalls in biodiversity conservation. Trends Ecol Evol. 2013;28: 689–695. doi:10.1016/j.tree.2013.09.003
27. Roll U, Feldman A, Novosolov M, Allison A, Bauer AM, Bernard R, et al. The global distribution of tetrapods reveals a need for targeted reptile conservation. Nat Ecol Evol. 2017;1: 1677–1682. doi:10.1038/s41559-017-0332-2
28. Guedes JJM, Moura MR, Diniz□Filho JAF. Species out of sight: elucidating the determinants of research effort in global reptiles. Ecography (Cop). 2023; Early View. doi:10.1111/ecog.06491
29. Kéry M, Schmidt BR. Imperfect detection and its consequences for monitoring for conservation. Community Ecol. 2008;9: 207–216. doi:10.1556/ComEc.9.2008.2.10
30. Mazerolle MJ, Bailey LL, Kendall WL, Royle JA, Converse SJ, Nichols JD. Making great leaps forward: Accounting for detectability in herpetological field studies. J Herpetol. 2007;41: 672–689. doi:10.1670/07-061.1
31. Jarzyna MA, Jetz W. Detecting the Multiple Facets of Biodiversity. Trends Ecol Evol. 2016;31: 527–538. doi:10.1016/j.tree.2016.04.002
32. Anderson AS, Marques TA, Shoo LP, Williams SE. Detectability in Audio-Visual Surveys of Tropical Rainforest Birds: The Influence of Species, Weather and Habitat Characteristics. Ambrosini R, editor. PLoS One. 2015;10: e0128464. doi:10.1371/journal.pone.0128464
33. Einoder LD, Southwell DM, Lahoz-Monfort JJ, Gillespie GR, Fisher A, Wintle BA. Occupancy and detectability modelling of vertebrates in northern Australia using multiple sampling methods. Crowther MS, editor. PLoS One. 2018;13: e0203304. doi:10.1371/journal.pone.0203304
34. Richardson E, Nimmo DG, Avitabile S, Tworkowski L, Watson SJ, Welbourne D, et al. Camera traps and pitfalls: an evaluation of two methods for surveying reptiles in a semiarid ecosystem. Wildl Res. 2017;44: 637. doi:10.1071/WR16048
35. Ribeiro-Júnior MA, Gardner TA, Ávila-Pires TC. Evaluating the Effectiveness of Herpetofaunal Sampling Techniques across a Gradient of Habitat Change in a Tropical Forest Landscape. J Herpetol. 2008;42: 733. doi:10.1670/07-097R3.1
36. Chutipong W, Lynam A, Steinmetz R, Savini T, Gale G. Sampling mammalian carnivores in western Thailand: Issues of rarity and detectability. Raffles Bull Zool. 2014;62: 521–535.
37. Moura MR, Costa HC, Peixoto MA, Carvalho ALG, Santana DJ, Vasconcelos HL. Geographical and socioeconomic determinants of species discovery trends in a

biodiversity hotspot. Biol Conserv. 2018;220: 237–244. doi:10.1016/j.biocon.2018.01.024

38. Brooke ZM, Bielby J, Nambiar K, Carbone C. Correlates of Research Effort in Carnivores: Body Size, Range Size and Diet Matter. PLoS One. 2014;9: e93195. doi:10.1371/journal.pone.0093195

39. Guedes JJM, Moura MR, Alexandre F. Diniz Filho J. Species out of sight: elucidating the determinants of research effort in global reptiles. Ecography (Cop). 2023;2023: e06491. doi:10.1111/ecog.06491

40. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. Ecology. 2014;95: 2027–2027. doi:10.1890/13-1917.1

41. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2021.3. 2021 p. www.iucnredlist.org.

42. Jetz W, McPherson JM, Guralnick RP. Integrating biodiversity distribution knowledge: toward a global map of life. Trends Ecol Evol. 2012;27: 151–159.

43. Marsh CJ, Sica Y V., Burgin CJ, Dorman WA, Anderson RC, Toro Mijares I, et al. Expert range maps of global mammal distributions harmonised to three taxonomic authorities. J Biogeogr. 2022; 1–14. doi:10.1111/jbi.14330

44. Hurlbert AH, Jetz W. Species richness, hotspots, and the scale dependence of range maps in ecology and conservation. Proc Natl Acad Sci. 2007;104: 13384–13389.

45. Hurlbert AH, White EP. Disparity between range map-and survey-based analyses of species richness: patterns, processes and implications. Ecol Lett. 2005;8: 319–327.

46. Jetz W, Sekercioglu CH, Watson JEM. Ecological correlates and conservation implications of overestimating species geographic ranges. Conserv Biol. 2008;22: 110–119. doi:10.1111/j.1523-1739.2007.00847.x

47. Amatulli G, Domisch S, Tuanmu M-N, Parmentier B, Ranipeta A, Malczyk J, et al. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. Sci Data. 2018;5: 180040. doi:10.1038/sdata.2018.40

48. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, et al. Climatologies at high resolution for the earth's land surface areas. Sci Data. 2017;4: 170122. doi:10.1038/sdata.2017.122

49. Klein Goldewijk K, Beusen A, Doelman J, Stehfest E. Anthropogenic land use estimates for the Holocene-HYDE 3.2. Earth Syst Sci Data. 2017;9: 927–953. doi:10.5194/essd-9-927-2017

50. Nelson A. Estimated travel time to the nearest city of 50,000 or more people in year 2000. Ispra: Global Environment Monitoring Unit - Joint Research Centre of the European Commission; 2008. Available: http://forobs.jrc.ec.europa.eu/products/gam/

51. Scott Chamberlain. rredlist: "IUCN" Red List Client. 2020. p. R package version 0.7.0. Available: https://cran.r-project.org/package=rredlist

52. Jetz W, Thomas GH, Joy JB, Redding DW, Hartmann K, Mooers AO. Global Distribution and Conservation of Evolutionary Distinctness in Birds. Curr Biol. 2014;24: 919–930. doi:10.1016/j.cub.2014.03.011

53. Anders G J Rhodin, Iverson JB, Van-Dijk PP, Buhlmann KA, Pritchard PCH, Mittermeier RA. Turtles of the world: Annotated checklist and atlas of taxonomy, synonymy, distribution, and conservation status (8th Ed.). New York: Chelonian Research Foundation; 2017. Available: https://books.google.com/books/about/Turtles_of_the_World_Annotated_Checklist.html?id=9PSOswEACAAJ

54. Oaks JR. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. Evolution (N Y). 2011;65: 3285–3297. doi:10.1111/j.1558-5646.2011.01373.x

55. Meredith RW, Hekkala ER, Amato G, Gatesy J. A phylogenetic hypothesis for *Crocodylus* (Crocodylia) based on mitochondrial DNA: Evidence for a trans-Atlantic voyage from Africa to the New World. Mol Phylogenet Evol. 2011;60: 183–191. doi:10.1016/j.ympev.2011.03.026

56. Shirley MH, Carr AN, Nestler JH, Vliet KA, Brochu CA. Systematic revision of the living African Slender-snouted Crocodiles (Mecistops Gray, 1844). Zootaxa. 2018;4504: 151. doi:10.11646/zootaxa.4504.2.1
57. Franke FA, Schmidt F, Borgwardt C, Bernhard D, Bleidorn C, Engelmann W-E, et al. Genetic differentiation of the African dwarf crocodile *Osteolaemus tetraspis* Cope, 1861 (Crocodylia: Crocodylidae) and consequences for European zoos. Org Divers Evol. 2013;13: 255–266. doi:10.1007/s13127-012-0107-1
58. Shirley MH, Villanova VL, Vliet KA, Austin JD. Genetic barcoding facilitates captive and wild management of three cryptic African crocodile species complexes. Anim Conserv. 2015;18: 322–330. doi:10.1111/acv.12176
59. del Hoyo J, Elliott A, Sargatal J, Christie DA. Handbook of the Birds of the World. Barcelona: Lynx Edicions;
60. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2015.1. 2015 p. www.iucnredlist.org.
61. Frost DR. Amphibian Species of the World: an Online Reference. Version 6.1. In: American Museums of Natural History, New York, USA [Internet]. 2021 [cited 20 Jan 2022]. Available: https://amphibiansoftheworld.amnh.org/
62. Uetz P, Freed P, Hosek J. The Reptile Database. 2021 [cited 20 Jan 2022]. Available: http://www.reptile-database.org
63. Lepage D, Vaidya G, Guralnick R. Avibase – a database system for managing and organizing taxonomic concepts. Zookeys. 2014;420: 117–135. doi:10.3897/zookeys.420.7089
64. Burgin CJ, Colella JP, Kahn PL, Upham NS. How many species of mammals are there? J Mammal. 2018;99: 1–14. doi:10.1093/jmammal/gyx147
65. Meiri S, Bauer AM, Allison A, Castro-Herrera F, Chirio L, Colli G, et al. Extinct, obscure or imaginary: The lizard species with the smallest ranges. Stevens G, editor. Divers Distrib. 2018;24: 262–273. doi:10.1111/ddi.12678
66. Martin TE, Bennett GC, Fairbairn A, Mooers AO. 'Lost' taxa and their conservation implications. Anim Conserv. 2022;Early View: 1–11. doi:10.1111/acv.12788
67. Jardim L, Bini LM, Diniz-Filho JAF, Villalobos F. A Cautionary Note on Phylogenetic Signal Estimation from Imputed Databases. Evol Biol. 2021;48: 246–258. doi:10.1007/s11692-021-09534-0
68. Johnson TF, Isaac NJB, Paviolo A, González□Suárez M. Handling missing values in trait data. Schrodt F, editor. Glob Ecol Biogeogr. 2021;30: 51–62. doi:10.1111/geb.13185
69. Tobias JA, Sheard C, Pigot AL, Devenish AJM, Yang J, Sayol F, et al. AVONET: morphological, ecological and geographical data for all birds. Ecol Lett. 2022;25: 581–597. doi:10.1111/ELE.13898
70. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics. 2004;20: 289–290. doi:10.1093/bioinformatics/btg412
71. Warnes GR, Bolker B, Lumley T, Johnson RC. gmodels: Various R Programming Tools for Model Fitting. 2018. p. R package version 2.18.1. Available: https://cran.r-project.org/package=gmodels
72. Oliveira BF, São-Pedro VA, Santos-Barrera G, Penone C, Costa GC. AmphiBIO, a global database for amphibian ecological traits. Sci Data. 2017;4: 170123. doi:10.1038/sdata.2017.123
73. Zamudio KR, Bell RC, Nali RC, Haddad CFB, Prado CPA. Polyandry, Predation, and the Evolution of Frog Reproductive Modes. Am Nat. 2016;188: S41–S61. doi:10.1086/687547
74. Manamendra-Arachchi K, Pethiyagoda R. The Sri Lankan shrub-frogs of the genus Philautus Gistel, 1848 (Ranidae: Rhacophorinae), with description of 27 new species. Raffles Bull Zool. 2005;12: 163–303.
75. Liang F, Shuqin H, Changyuan Y, Yongzhao H. Fauna Sinica: Amphibia （Vol. 3）. Beijing: Science Press; 2009.
76. Channing A, Ro□del M-O. Field guide to the frogs & other amphibians of Africa.

Cape Town: Struik Nature; 2019.
77. Feldman A, Sabath N, Pyron RA, Mayrose I, Meiri S. Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. Glob Ecol Biogeogr. 2016;25: 187–197. doi:10.1111/geb.12398
78. Myhrvold NP, Baldridge E, Chan B, Sivam D, Freeman DL, Ernest SKM. An amniote life history database to perform comparative analyses with birds, mammals, and reptiles. Dryad Digit Repos. 2014;96: 3109. doi:10.5061/dryad.t6m96
79. Meyer de Schauensee R. The birds of China. Washington, DC: Smithsonian Institution Press; 1984.
80. Hutchins M, Jackson JA, Bock WJ, Olendorf D. Grzimek's Animal Life Encyclopedia, v. 8-11: Birds. 2nd editio. Farmington Hills: Gale Group; 2003.
81. Santini L, Benítez-López A, Ficetola GF, Huijbregts MAJ. Length–mass allometries in amphibians. Integr Zool. 2018;13: 36–45. doi:10.1111/1749-4877.12268
82. Stark G, Pincheira-Donoso D, Meiri S. No evidence for the 'rate-of-living' theory across the tetrapod tree of life. Glob Ecol Biogeogr. 2020;29: 857–884. doi:10.1111/GEB.13069
83. Stark G, Pincheira-Donoso D, Meiri S. CORRIGENDUM: No evidence for the 'rate of living' theory across the tetrapod tree of life. Glob Ecol Biogeogr. 2021;30: 2514–2514. doi:10.1111/geb.13400
84. Jacobsen NHG. A Herpetological survey of the Transvaal. University of Natal. 1989. Available: https://researchspace.ukzn.ac.za/xmlui/handle/10413/8461?show=full
85. Faurby S, Davis M, Pedersen RØ, Schowanek SD, Antonelli1 A, Svenning J. PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. Ecology. 2018;99: 2626–2626. doi:10.1002/ecy.2443
86. Faurby S, Svenning J-C. Resurrection of the Island Rule: Human-Driven Extinctions Have Obscured a Basic Evolutionary Pattern. Am Nat. 2016;187: 812–820. doi:10.1086/686268
87. Jones KE, Bielby J, Cardillo M, Fritz SA, O'Dell J, Orme CDL, et al. PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Michener WK, editor. Ecology. 2009;90: 2648–2648. doi:10.1890/08-1494.1
88. Kingdon J, Happold D, Butynski T, Hoffmann M, Happold M, Kalina J. Mammals of Africa v. 1-6. 1st editio. London: Bloomsbury Publishing; 2013.
89. Haddad CFB, Toledo LF, Prado CPA, Loebmann D, Gasparini JL, Sazima I. Guide to the amphibians of the Atlantic Forest: diversity and biology. São Paulo: Anolis Book; 2013.
90. Stark G, Meiri S. Cold and dark captivity: Drivers of amphibian longevity. Glob Ecol Biogeogr. 2018;27: 1384–1397. doi:10.1111/geb.12804
91. Stark G, Meiri S. CORRIGENDUM: Cold and dark captivity: Drivers of amphibian longevity. Glob Ecol Biogeogr. 2021;30: 2513–2513. doi:10.1111/geb.13399
92. Thomas KN, Gower DJ, Bell RC, Fujita MK, Schott RK, Streicher JW. Eye size and investment in frogs and toads correlate with adult habitat, activity pattern and breeding ecology. Proc R Soc B Biol Sci. 2020;287: 20201393. doi:10.1098/rspb.2020.1393
93. Neves MO, Cabral H, Pedrozo M, Ferreira VL, Moura MR, Santana DJ. Dataset of occurrences and ecological traits of amphibians from Upper Paraguay River Basin, central South America. Nat Conserv. 2020;41: 71–89. doi:10.3897/natureconservation.41.54265
94. Lovich JE, Gibbons W. Turtles of the World: A guide to every family. Princeton: Princeton University Press; 2021.
95. Meiri S. Traits of lizards of the world: Variation around a successful evolutionary design. Glob Ecol Biogeogr. 2018;27: 1168–1172. doi:10.1111/geb.12773
96. Harrington SM, de Haan JM, Shapiro L, Ruane S. Habits and characteristics of arboreal snakes worldwide: arboreality constrains body size but does not affect lineage diversification. Biol J Linn Soc. 2018;125: 61–71. doi:10.1093/biolinnean/bly097
97. O'Shea M. The Book of Snakes: A life-size guide to six hundred species from around

the world. Chicago: University of Chicago Press; 2018.
98. Pietersen D, Verburgt L, Davies J. Snakes and other reptiles of Zambia and Malawi. Cape Town: Struik Nature; 2021.
99. Scharf I, Feldman A, Novosolov M, Pincheira-Donoso D, Das I, Böhm M, et al. Late bloomers and baby boomers: ecological drivers of longevity in squamates and the tuatara. Glob Ecol Biogeogr. 2015;24: 396–405. doi:10.1111/geb.12244
100. Feldman A, Meiri S. Australian snakes do not follow Bergmann's rule. Evol Biol. 2014;41: 327–335. doi:10.1007/s11692-014-9271-x
101. Guedes TB, Nogueira C, Marques OAV. Diversity, natural history, and geographic distribution of snakes in the Caatinga, Northeastern Brazil. Zootaxa. 2014;3863: 1–93. doi:10.11646/zootaxa.3863.1.1
102. Fraga R, Lima AP, Prudente ALC, Magnusson WE. Guide to the Snakes of the Manaus region - Central Amazonia. Manaus: Editora INPA; 2013.
103. Marques OA V, Eterovic A, Sazima I. Snakes of the Brazilian Atlantic Forest: An illustrated field guide for the Serra do Mar range. Ribeirão Preto: Holos; 2004.
104. Marques OA V, Eterovic A, Nogueira CC, Sazima I. Serpentes do Cerrado: Guia ilustrado. Ribeirão Preto: Holos; 2015.
105. Chippaux J-P, Jackson K. Snakes of Central and Western Africa. Baltimore Maryland: Johns Hopkins University Press; 2019.
106. Vogel G. Venomous snakes of Asia. Frankfurt: Edition Chimaira; 2006.
107. Munstermann MJ, Heim NA, McCauley DJ, Payne JL, Upham NS, Wang SC, et al. A global ecological signal of extinction risk in terrestrial vertebrates. Conserv Biol. 2021; 1–13. doi:10.1111/cobi.13852
108. Bars-Closel M, Kohlsdorf T, Moen DS, Wiens JJ. Diversification rates are more strongly related to microhabitat than climate in squamate reptiles (lizards and snakes). Evolution (N Y). 2017;71: 2243–2261. doi:10.1111/evo.13305
109. Branch B. Photographic guide to snakes, other reptiles and amphibians of East Africa. Cape Town: Random House Struik; 2014.
110. Branch B. A photographic guide to snakes and other reptiles of Southern Africa. Struik Nature; 2002.
111. Figueroa A. Phylogenetic Relationships and Evolution of Snakes. University of New Orleans. 2016. Available: https://scholarworks.uno.edu/td/2222/
112. Campbell JA, Lamar WW. The venomous reptiles of the Western Hemisphere - 2 volumes. Ithaca: Comstock Pub. Associates; 2004.
113. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2016.3. 2016 p. www.iucnredlist.org.
114. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2018.2. 2018 p. www.iucnredlist.org.
115. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2019.3. 2020 p. www.iucnredlist.org.
116. IUCN - International Union for Conservation of Nature. IUCN Red List of Threatened Species. In: Version 2019.2. 2019 p. www.iucnredlist.org.
117. Stone L, Roberts A. The checkerboard score and species distributions. Oecologia. 1990;85: 74–79. doi:10.1007/BF00317345
118. Gotelli NJ. Null model analysis of species co-occurrence patterns. Ecology. 2000;81: 2606–2621. doi:10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2
119. Ulrich W, Gotelli NJ. Pattern detection in null model analysis. Oikos. 2013;122: 2–18. doi:10.1111/j.1600-0706.2012.20325.x
120. Gotelli NJ, Hart E, Ellison A. Null Model Analysis for Ecological Data. R package version 0.1.0. 2015. pp. 1–53.
121. Castiglione S, Tesone G, Piccolo M, Melchionna M, Mondanaro A, Serio C, et al. A new method for testing evolutionary rate variation and shifts in phenotypic evolution. Cooper N, editor. Methods Ecol Evol. 2018;9: 974–983. doi:10.1111/2041-210X.12954
122. Scheffers BR, Yong DL, Harris JBC, Giam X, Sodhi NS. The World's Rediscovered

Species: Back from the Brink? Gratwicke B, editor. PLoS One. 2011;6: e22531. doi:10.1371/journal.pone.0022531

123. Ladle R, Jepson P, Malhado ACM, Jennings S, Barua M. The causes and biogeographical significance of species' rediscovery. Front Biogeogr. 2012;3: 111–118. doi:10.21425/F5FBG12432

124. Meyer C, Kreft H, Guralnick R, Jetz W. Global priorities for an effective information basis of biodiversity distributions. Nat Commun. 2015;6: 8221. doi:10.1038/ncomms9221

125. Einoder LD, Southwell DM, Lahoz-Monfort JJ, Gillespie GR, Fisher A, Wintle BA. Occupancy and detectability modelling of vertebrates in northern Australia using multiple sampling methods. Crowther MS, editor. PLoS One. 2018;13: e0203304. doi:10.1371/journal.pone.0203304

126. Richardson E, Nimmo DG, Avitabile S, Tworkowski L, Watson SJ, Welbourne D, et al. Camera traps and pitfalls: an evaluation of two methods for surveying reptiles in a semiarid ecosystem. Wildl Res. 2017;44: 637. doi:10.1071/WR16048

127. Vidan E, Roll U, Bauer A, Grismer L, Guo P, Maza E, et al. The Eurasian hot nightlife: Environmental forces associated with nocturnality in lizards. Glob Ecol Biogeogr. 2017;26: 1316–1325. doi:10.1111/geb.12643

128. Atwood TB, Valentine SA, Hammill E, McCauley DJ, Madin EMP, Beard KH, et al. Herbivores at the highest risk of extinction among mammals, birds, and reptiles. Sci Adv. 2020;6. doi:10.1126/sciadv.abb8458

129. Anderson SR, Wiens JJ. Out of the dark: 350 million years of conservatism and evolution in diel activity patterns in vertebrates. Evolution (N Y). 2017; 1–16. doi:10.1111/evo.13284

130. Pigot AL, Sheard C, Miller ET, Bregman TP, Freeman BG, Roll U, et al. Macroevolutionary convergence connects morphological form to ecological function in birds. Nat Ecol Evol. 2020;4: 230–239. doi:10.1038/s41559-019-1070-4

131. Violle C, Reich PB, Pacala SW, Enquist BJ, Kattge J. The emergence and promise of functional biogeography. Proc Natl Acad Sci. 2014;111: 13690–13696. doi:10.1073/pnas.1415442111

132. Cardillo M, Mace GM, Gittleman JL, Jones KE, Bielby J, Purvis A. The predictability of extinction: biological and external correlates of decline in mammals. Proc R Soc B Biol Sci. 2008;275: 1441–1448. doi:10.1098/rspb.2008.0179

133. Pollock LJ, Thuiller W, Jetz W. Large conservation gains possible for global biodiversity facets. Nature. 2017;546: 141–144. doi:10.1038/nature22368

134. Lee TM, Jetz W. Unravelling the structure of species extinction risk for predictive conservation science. Proc R Soc B Biol Sci. 2011;278: 1329–1338. doi:10.1098/rspb.2010.1877

135. Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König□Ries B, et al. Towards an ecological trait□data standard. Orme D, editor. Methods Ecol Evol. 2019;10: 2006–2019. doi:10.1111/2041-210X.13288

136. Bremer K, Bremer B, Karis P, Källersjö M. Time for change in taxonomy. Nature. 1990;343: 202–202. doi:10.1038/343202a0

137. Garnett ST, Christidis L. Taxonomy anarchy hampers conservation. Nature. 2017;546: 25–27. doi:10.1038/546025a

138. Wüest RO, Zimmermann NE, Zurell D, Alexander JM, Fritz SA, Hof C, et al. Macroecology in the age of Big Data – Where to go from here? J Biogeogr. 2019; jbi.13633. doi:10.1111/jbi.13633

139. Hortal J, de Bello F, Diniz-Filho JAF, Lewinsohn TM, Lobo JM, Ladle RJ. Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annu Rev Ecol Evol Syst. 2015;46: annurev-ecolsys-112414-054400. doi:10.1146/annurev-ecolsys-112414-054400

140. Heberling JM, Miller JT, Noesgaard D, Weingart SB, Schigel D. Data integration enables global biodiversity synthesis. Proc Natl Acad Sci. 2021;118. doi:10.1073/pnas.2018093118

141. Moura MR, Jetz W. Shortfalls and opportunities in terrestrial vertebrate species discovery. Nat Ecol Evol. 2021;5: 631–639. doi:10.1038/s41559-021-01411-5
142. Diniz-Filho JAF, Bastos RP, Rangel TFLVB, Bini LM, Carvalho P, Silva RJ. Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. Glob Ecol Biogeogr. 2005;14: 469–477. doi:10.1111/j.1466-822x.2005.00165.x
143. Meiri S. Small, rare and trendy: traits and biogeography of lizards described in the 21st century. J Zool. 2016;299: 251–261. doi:10.1111/jzo.12356
144. Roll U, Mittermeier JC, Diaz GI, Novosolov M, Feldman A, Itescu Y, et al. Using Wikipedia page views to explore the cultural importance of global reptiles. Biol Conserv. 2016;204: 42–50. doi:10.1016/j.biocon.2016.03.037
145. Tyler EHM, Somerfield PJ, Berghe E Vanden, Bremner J, Jackson E, Langmead O, et al. Extensive gaps and biases in our knowledge of a well-known fauna: implications for integrating biological traits into macroecology. Glob Ecol Biogeogr. 2012;21: 922–934. doi:10.1111/j.1466-8238.2011.00726.x
146. Meiri S, Chapple DG. Biases in the current knowledge of threat status in lizards, and bridging the 'assessment gap.' Biol Conserv. 2016;204: 6–15. doi:10.1016/j.biocon.2016.03.009
147. Miqueleiz I, Bohm M, Ariño AH, Miranda R. Assessment gaps and biases in knowledge of conservation status of fishes. Aquat Conserv Mar Freshw Ecosyst. 2020;30: 225–236. doi:10.1002/aqc.3282
148. Blumstein DT. Developing an evolutionary ecology of fear: how life history and natural history traits affect disturbance tolerance in birds. Anim Behav. 2006;71: 389–399. doi:10.1016/j.anbehav.2005.05.010
149. Henle K, Davies KF, Kleyer M, Margules C, Settele J. Predictors of Species Sensitivity to Fragmentation. Biodivers Conserv. 2004;13: 207–251. doi:10.1023/B:BIOC.0000004319.91643.9e
150. Robinson JG, Redford KH. Body Size, Diet, and Population Density of Neotropical Forest Mammals. Am Nat. 1986;128: 665–680. doi:10.1086/284596
151. White EP, Ernest SKM, Kerkhoff AJ, Enquist BJ. Relationships between body size and abundance in ecology. Trends Ecol Evol. 2007;22: 323–330. doi:10.1016/j.tree.2007.03.007
152. Ribeiro-Júnior MA, Gardner TA, Ávila-Pires TC. Evaluating the Effectiveness of Herpetofaunal Sampling Techniques across a Gradient of Habitat Change in a Tropical Forest Landscape. J Herpetol. 2008;42: 733. doi:10.1670/07-097R3.1
153. Meyer C, Jetz W, Guralnick RP, Fritz SA, Kreft H. Range geometry and socio-economics dominate species-level biases in occurrence information. Glob Ecol Biogeogr. 2016. doi:10.1111/geb.12483
154. Cooke RSC, Bates AE, Eigenbrod F. Global trade-offs of functional redundancy and functional dispersion for birds and mammals. Schrodt F, editor. Glob Ecol Biogeogr. 2019;28: 484–495. doi:10.1111/geb.12869
155. Vilela B, Villalobos F, Rodríguez MÁ, Terribile LC. Body Size, Extinction Risk and Knowledge Bias in New World Snakes. Lambertucci SA, editor. PLoS One. 2014;9: e113429. doi:10.1371/journal.pone.0113429
156. Vidan E, Novosolov M, Bauer AM, Herrera FC, Chirio L, Campos CN, et al. The global biogeography of lizard functional groups. J Biogeogr. 2019;46: 2147–2158. doi:10.1111/jbi.13667
157. Oliveira BF, Scheffers BR. Vertical stratification influences global patterns of biodiversity. Ecography (Cop). 2019;42: 249–249. doi:10.1111/ecog.03636
158. Gaston KJ, Chown SL, Evans KL. Ecogeographical rules: elements of a synthesis. J Biogeogr. 2008;35: 483–500. doi:10.1111/j.1365-2699.2007.01772.x

Figure Legends

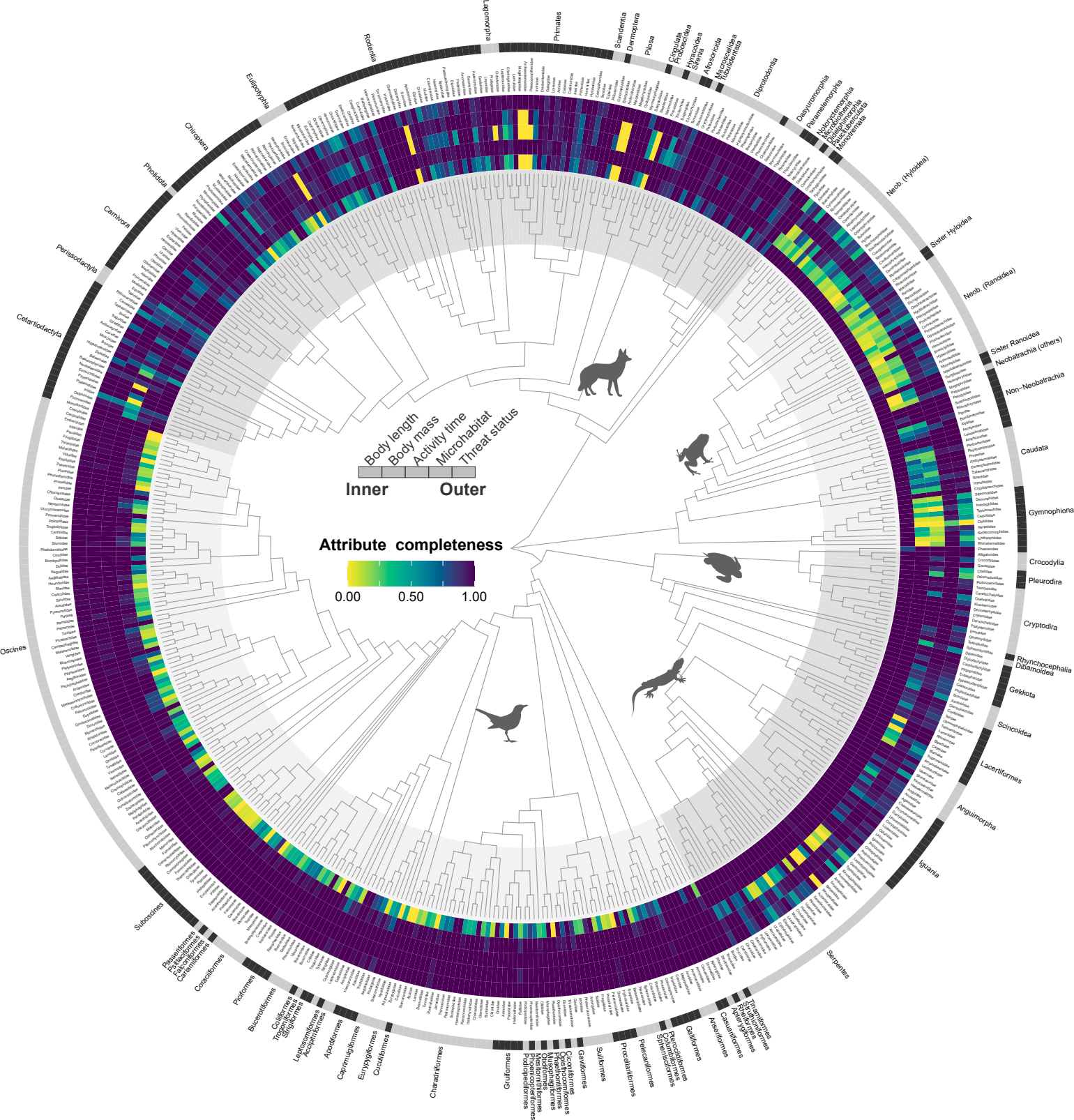
Figure 1. Attribute data coverage (completeness) of tetrapod species by family. Colour rings show the proportion of species with observed attributes in each tetrapod family for the five attributes with lowest completeness. Phylogenetic relationships shown are simple the family-level topology of tetrapod families.

Figure 2. Average data completeness across tetrapod assemblages. Proportion of species with observed values for: (A) body length, (B) body mass, (D) activity time, (E) microhabitat, (G) assessed threat status, (H) phylogenetic data, (J) attribute set, i.e., the average pattern for maps depicted in ABDE, and (K) the complete database, i.e., average pattern for maps depicted in ABDEGH. Maps show grid cell assemblages of 110 × 110 km size in an equal-area projection. Latitudinal plots (C, F, I, L) show the average values of these cells across latitudes. See Fig. S4 for spatial patterns of attribute completeness for each tetrapod group. Colour ramps followed Jenks' natural breaks classification.

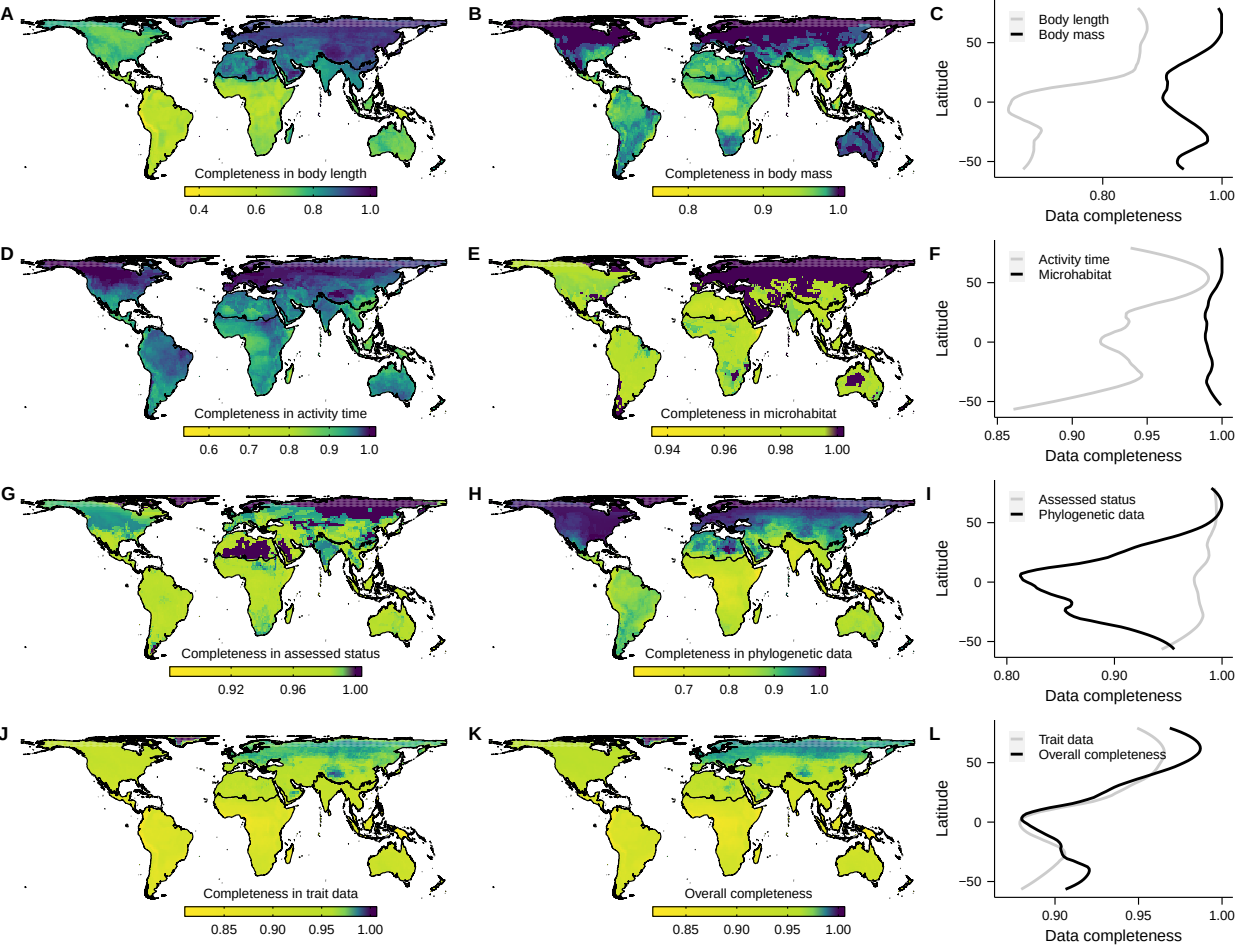
Figure 3. The top 50 tetrapod genera with most pronounced patterns of shared data gaps. For each genus, the aggregation metric equals the median value of the standardised effect-size (SES) of the C-score metric computed across (A) all pairwise attribute combinations among the five focal traits; and (B) all attribute pairs mandatorily involving threat status. Gray numbers on the left side of each panel indicate the species richness per genus. See Fig. S5 for aggregation patterns computed at the family-level.

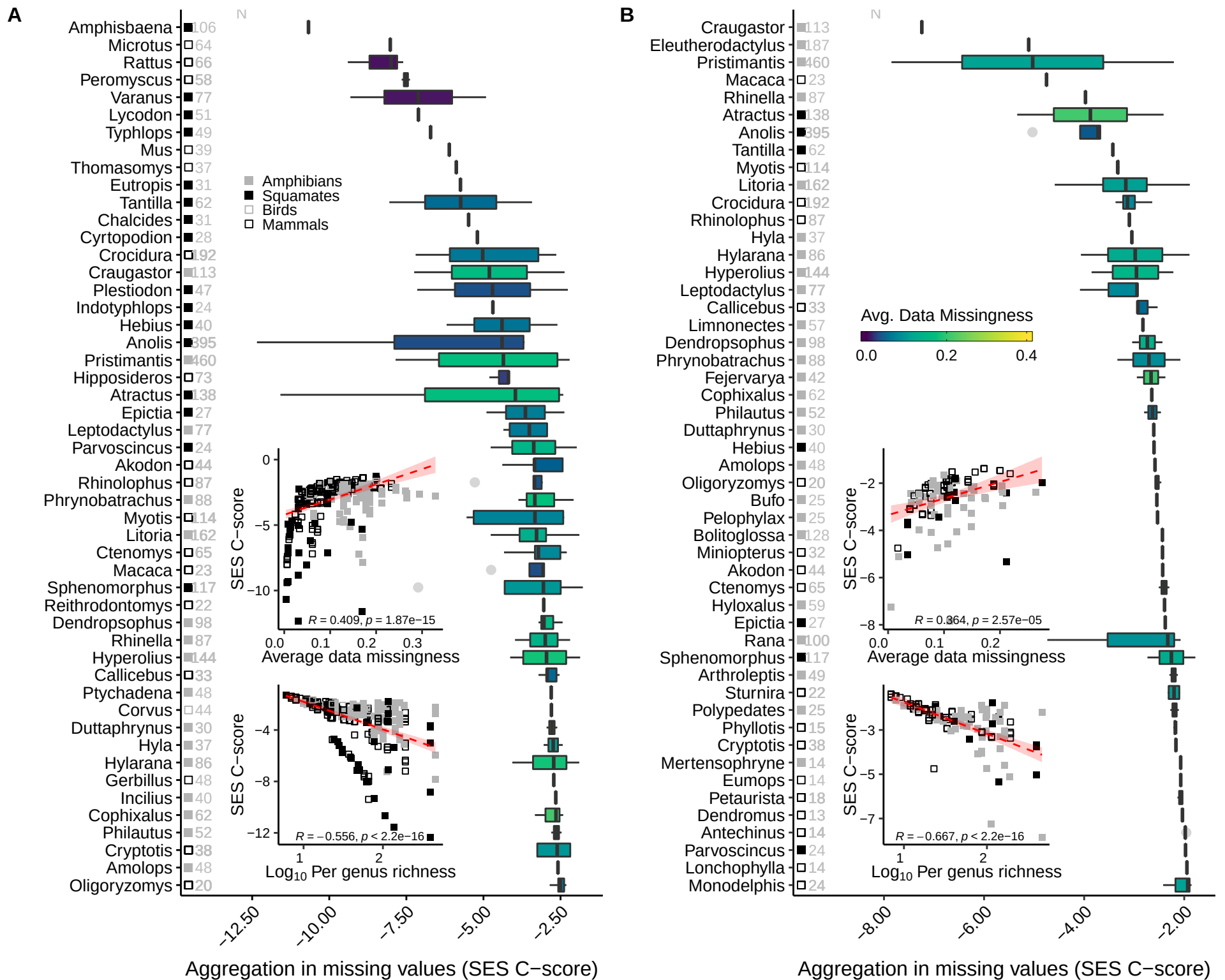
Figure 4. Changes in family-level attributes average after imputation-based gap-filling. For continuous attributes (BL, body length; BM, body mass), relative change in attribute property quantifies the proportional change in the geometric mean attribute per genus. For binary attributes (types of activity time and microhabitat), relative changes in attribute property measures the change in mean species nocturnality (Noct) and verticality (Vert) scores. The inset donut chart denotes the proportion of genera (inner donut) and families (outer donut) by category of attribute change. Fig. S11 contains individual charts of each tetrapod group.

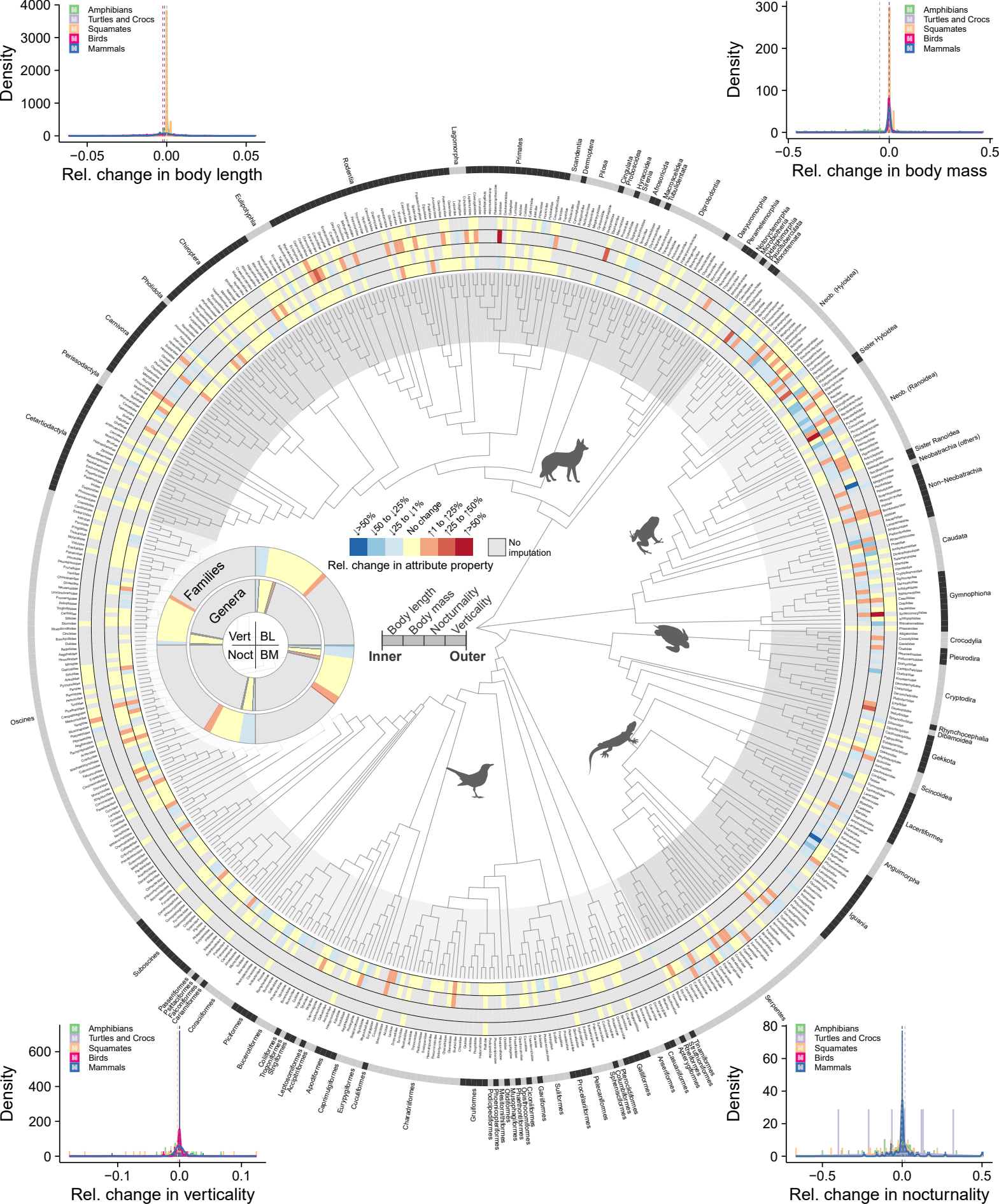
Figure 5. Changes in average attribute per tetrapod assemblage after imputation-based gap-filing. For each grid cell, maps show the average species attribute value and respective relative change in attribute value after inclusion of imputed values for (A-C) body length, (D-F) body mass, (G-I) nocturnality, (J-L) verticality. Body length and body mass were $\log_{10}$ transformed before computations. Gray cells indicate assemblages without species with imputed values. See Figs S6-S10 for patterns of attribute change for each tetrapod group.



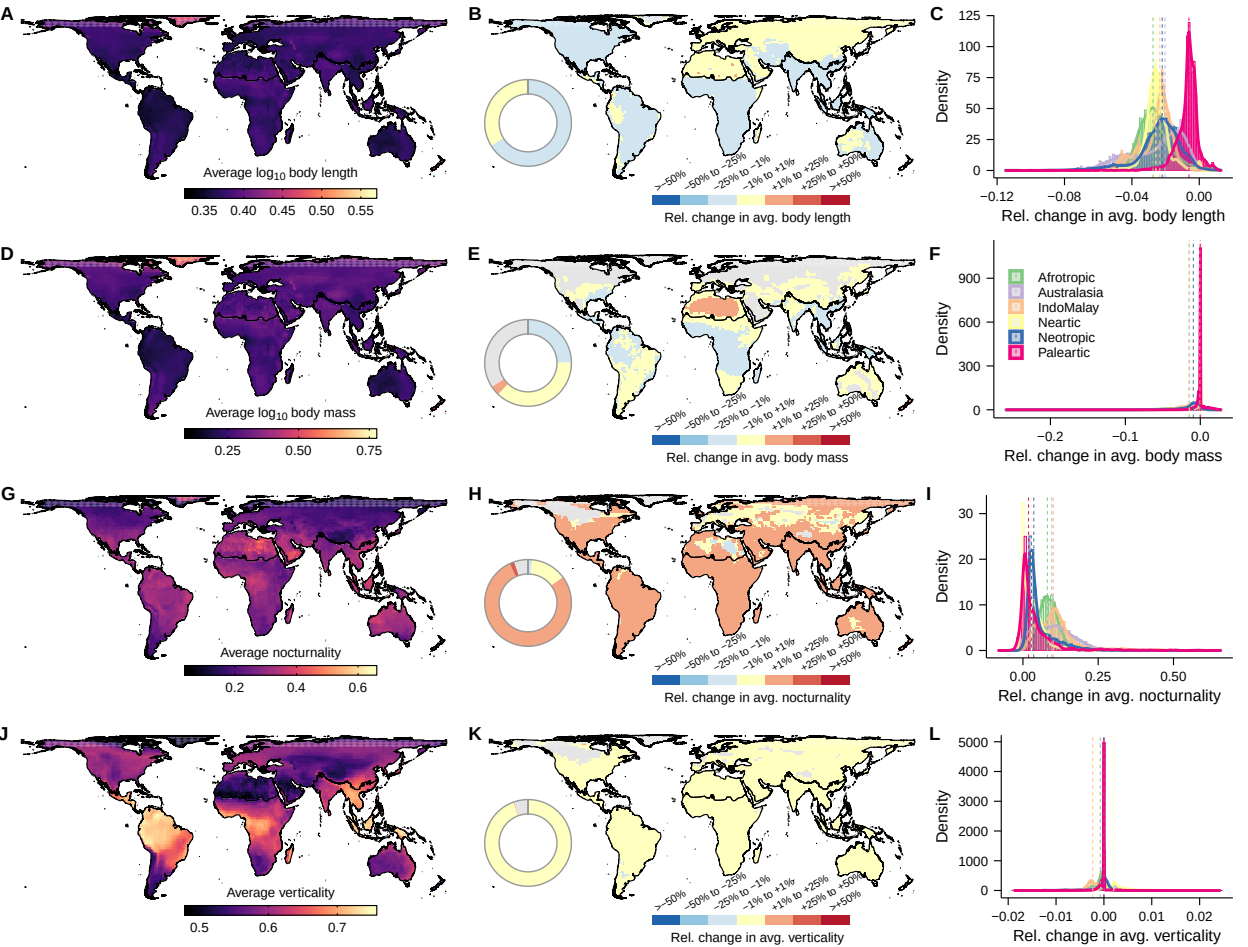
Body length
Body mass
Activity time
Microhabitat
Threat status
Inner
Outer
Attribute completeness
0.00
0.50
1.00
Lagomorpha
Primates
Scandentia
Dermoptera
Pilosa
Cingulata
Proboscidea
Hyracoidea
Sirenia
Afrosoricida
Macroscelidea
Tubulidentata
Diprotodontia
Dasyuromorphia
Peramelemorphia
Notoryctemorphia
Microbiotheria
Didelphimorphia
Paucituberculata
Monotremata
Neob. (Hyloidea)
Sister Hyloidea
Neob. (Ranoidea)
Sister Ranoidea
Neobatrachia (others)
Non−Neobatrachia
Caudata
Gymnophiona
Crocodylia
Pleurodira
Cryptodira
Rhynchocephalia
Dibamoidea
Gekkota
Scincoidea
Lacertiformes
Anguimorpha
Iguania
Serpentes
Thamiformes
Strutiformes
Rheiformes
Apterygiformes
Casuariformes
Anseriformes
Galliformes
Cerrocliformes
Columbiformes
Sphenisciformes
Pelecaniformes
Procellariiformes
Suliformes
Gaviiformes
Ciconiiformes
Phaethontiformes
Otidiformes
Mesitornithiformes
Phoenicopteriformes
Podicipediformes
Gruiformes
Charadriiformes
Cuculiformes
Eurypygiformes
Caprimulgiformes
Apodiformes
Leptosomiformes
Accipitriformes
Strigiformes
Trogoniformes
Coliiformes
Bucerotiformes
Piciformes
Coraciiformes
Cariamiformes
Falconiformes
Psittaciformes
Passeriformes
Suboscines
Oscines
Cetartiodactyla
Perissodactyla
Carnivora
Pholidota
Chiroptera
Eulipotyphla
Rodentia


A
Completeness in body length
0.4
0.6
0.8
1.0
B
Completeness in body mass
0.8
0.9
1.0
C
Body length
Body mass
Latitude
50
0
−50
0.80
1.00
Data completeness
D
Completeness in activity time
0.6
0.7
0.8
0.9
1.0
E
Completeness in microhabitat
0.94
0.96
0.98
1.00
F
Activity time
Microhabitat
Latitude
50
0
−50
0.85
0.90
0.95
1.00
Data completeness
G
Completeness in assessed status
0.92
0.96
1.00
H
Completeness in phylogenetic data
0.7
0.8
0.9
1.0
I
Assessed status
Phylogenetic data
Latitude
50
0
−50
0.80
0.90
1.00
Data completeness
J
Completeness in trait data
0.85
0.90
0.95
1.00
K
Overall completeness
0.85
0.90
0.95
1.00
L
Trait data
Overall completeness
Latitude
50
0
−50
0.90
0.95
1.00
Data completeness


A
N
Amphisbaena 106
Microtus 64
Rattus 66
Peromyscus 58
Varanus 77
Lycodon 51
Typhlops 49
Mus 39
Thomasomys 37
Eutropis 31
Tantilla 62
Chalcides 31
Cyrtopodion 28
Crocidura 192
Craugastor 113
Plestiodon 47
Indotyphlops 24
Hebius 40
Anolis 395
Pristimantis 460
Hipposideros 73
Atractus 138
Epictia 27
Leptodactylus 77
Parvoscincus 24
Akodon 44
Rhinolophus 87
Phrynobatrachus 88
Myotis 114
Litoria 162
Ctenomys 65
Macaca 23
Sphenomorphus 117
Reithrodontomys 22
Dendropsophus 98
Rhinella 87
Hyperolius 144
Callicebus 33
Ptychadena 48
Corvus 44
Duttaphrynus 30
Hyla 37
Hylarana 86
Gerbillus 48
Incilius 40
Cophixalus 62
Philautus 52
Cryptotis 38
Amolops 48
Oligoryzomys 20
Amphibians
Squamates
Birds
Mammals
SES C−score
R = 0.409, p = 1.87e−15
Average data missingness
R = −0.556, p < 2.2e−16
Log10 Per genus richness
−12.50
−10.00
−7.50
−5.00
−2.50
Aggregation in missing values (SES C−score)
B
N
Craugastor 113
Eleutherodactylus 187
Pristimantis 460
Macaca 23
Rhinella 87
Atractus 138
Anolis 395
Tantilla 62
Myotis 114
Litoria 162
Crocidura 192
Rhinolophus 87
Hyla 37
Hylarana 86
Hyperolius 144
Leptodactylus 77
Callicebus 33
Limnonectes 57
Dendropsophus 98
Phrynobatrachus 88
Fejervarya 42
Cophixalus 62
Philautus 52
Duttaphrynus 30
Hebius 40
Amolops 48
Oligoryzomys 20
Bufo 25
Pelophylax 25
Bolitoglossa 128
Miniopterus 32
Akodon 44
Ctenomys 65
Hyloxalus 59
Epictia 27
Rana 100
Sphenomorphus 117
Arthroleptis 49
Sturnira 22
Polypedates 25
Phyllotis 15
Cryptotis 38
Mertensophryne 14
Eumops 14
Petaurista 18
Dendromus 13
Antechinus 14
Parvoscincus 24
Lonchophylla 14
Monodelphis 24
Avg. Data Missingness
0.0
0.2
0.4
SES C−score
R = 0.364, p = 2.57e−05
Average data missingness
R = −0.667, p < 2.2e−16
Log10 Per genus richness
−8.00
−6.00
−4.00
−2.00
Aggregation in missing values (SES C−score)



Amphibians
Turtles and Crocs
Squamates
Birds
Mammals
Density
Rel. change in body length
Rel. change in body mass
Rel. change in verticality
Rel. change in nocturnality
Rel. change in attribute property
No imputation
Families
Genera
Vert
BL
Noct
BM
Body length
Body mass
Nocturnality
Verticality
Inner
Outer
Lagomorpha
Primates
Scandentia
Dermoptera
Pilosa
Cingulata
Proboscidea
Hyracoidea
Sirenia
Afrosoricida
Macroscelidea
Tubulidentata
Diprotodontia
Dasyuromorphia
Peramelemorphia
Notoryctemorphia
Microbiotheria
Dactylopsilia
Paucituberculata
Monotremata
Neob. (Hyloidea)
Sister Hyloidea
Neob. (Ranoidea)
Sister Ranoidea
Neobatrachia (others)
Non-Neobatrachia
Caudata
Gymnophiona
Crocodylia
Pleurodira
Cryptodira
Rhynchocephalia
Dibamoidea
Gekkota
Scincoidea
Lacertiformes
Anguimorpha
Iguania
Serpentes
Rodentia
Eulipotyphla
Chiroptera
Pholidota
Carnivora
Perissodactyla
Cetartiodactyla
Oscines
Suboscines
Passeriformes
Psittaciformes
Falconiformes
Cariamiformes
Coraciformes
Piciformes
Bucerotiformes
Coliiformes
Trogoniformes
Strigiformes
Leptosomiformes
Accipitriformes
Apodiformes
Caprimulgiformes
Eurypygiformes
Cuculiformes
Charadriiformes
Gruiformes
Podicipediformes
Phoenicopteriformes
Mesitornithiformes
Otidiformes
Musophagiformes
Phaethontiformes
Ciconiiformes
Ciconiiformes
Gaviiformes
Suliformes
Procellariiformes
Pelecaniformes
Pteroclidiformes
Columbiformes
Sphenisciformes
Galliformes
Anseriformes
Casuariiformes
Apterygiformes
Rheiformes
Struthioniformes
Tinamiformes


A
Average log10 body length
0.35 0.40 0.45 0.50 0.55
B
>−50% −50% to −25% −25% to −1% −1% to +1% +1% to +25% +25% to +50% >+50%
Rel. change in avg. body length
C
Density
125 100 75 50 25 0
−0.12 −0.08 −0.04 0.00
Rel. change in avg. body length
D
Average log10 body mass
0.25 0.50 0.75
E
>−50% −50% to −25% −25% to −1% −1% to +1% +1% to +25% +25% to +50% >+50%
Rel. change in avg. body mass
F
Afrotropic
Australasia
IndoMalay
Neartic
Neotropic
Paleartic
Density
900 600 300 0
−0.2 −0.1 0.0
Rel. change in avg. body mass
G
Average nocturnality
0.2 0.4 0.6
H
>−50% −50% to −25% −25% to −1% −1% to +1% +1% to +25% +25% to +50% >+50%
Rel. change in avg. nocturnality
I
Density
30 20 10 0
0.00 0.25 0.50
Rel. change in avg. nocturnality
J
Average verticality
0.5 0.6 0.7
K
>−50% −50% to −25% −25% to −1% −1% to +1% +1% to +25% +25% to +50% >+50%
Rel. change in avg. verticality
L
Density
5000 4000 3000 2000 1000 0
−0.02 −0.01 0.00 0.01 0.02
Rel. change in avg. verticality