



Global gaps in trait data for terrestrial vertebrates

Adrienne Etard | Sophie Morrill | Tim Newbold

Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, London, UK

Correspondence

Adrienne Etard, Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, Gower Street, London, WC1E 6BT, UK.
Email: adrienne.etard.17@ucl.ac.uk

Funding information

Royal Society

Editor: Catherine Sheard

Abstract

Aim: Trait data are increasingly being used in studies investigating the impacts of global changes on the structure and functioning of ecological communities. Despite a growing number of trait data collations for terrestrial vertebrates, there is to date no global assessment of the gaps and biases the data present. Here, we assess whether terrestrial vertebrate trait data are taxonomically, spatially and phylogenetically biased.

Location: Global.

Time period: Present.

Major taxa studied: Terrestrial vertebrates.

Methods: We compile seven ecological traits and quantify coverage as the proportion of species for which an estimate is available. For a species, we define completeness as the proportion of non-missing values across traits. We assess whether coverage and completeness differ across classes and examine phylogenetic biases in trait data. To investigate spatial biases, we test whether wider-ranging species have more complete trait data than narrow-ranging species. Additionally, we test whether species-rich regions, which are of most concern for conservation, are less well sampled than species-poor regions.

Results: Mammals and birds are well sampled even in species-rich regions. For reptiles and amphibians (herptiles), only body size presents a high coverage (>80%), in addition to habitat-related variables (amphibians). Herptiles are poorly sampled for other traits. The shortfalls are particularly acute in some species-rich regions and for certain clades. Across all classes, geographically rarer species have less complete trait information.

Main conclusions: Trait information is less available on average in some of the most diverse areas and in geographically rarer species, both of which crucial for biodiversity conservation. Gaps in trait data might impede our ability to conduct large-scale analyses, whereas biases can impact the validity of extrapolations. A short-term solution to the problem is to estimate missing trait data using imputation techniques, whereas a longer-term and more robust filling of existing gaps requires continued data-collection efforts.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Global Ecology and Biogeography published by John Wiley & Sons Ltd

KEYWORDS

completeness, coverage, phylogenetic biases, spatial biases, taxonomic biases, terrestrial vertebrates, traits

1 | INTRODUCTION

Species traits are fundamental to ecological and evolutionary research. Comparative studies regularly use trait data across organisms to understand evolutionary processes and species coexistence (Escudero & Valladares, 2016; Zamudio et al., 2016), to investigate global patterns of life-forms and functions (Díaz et al., 2016) or to assess the vulnerability of species to environmental changes (Bohm et al., 2016; Pacifici et al., 2015; Pearson et al., 2014). Given that traits influence the ability of species to cope with environmental changes (Newbold et al., 2013) and underpin the contributions of species to ecosystem processes (Lavorel & Garnier, 2002; Violle et al., 2007; Wong et al., 2018), they play an increasingly important role in functional and conservation ecology.

Past and recent efforts to collate and release trait data in the public domain have facilitated the development of trait-based research. For instance, a global trait database has been published for plants (Kattge et al., 2011). As of May 2020, data from this database had been used in 305 publications since its release (activity report, 15 September 2020, <https://www.try-db.org/TryWeb/Home.php>). Such databases constitute invaluable research tools and have the potential to advance the field greatly.

Vertebrates are one of the most studied taxa (Tittley et al., 2017). There are now diverse sources of ecological traits for vertebrate groups (primates: Galán-Acedo et al., 2019; mammals: "PanTHERIA", Jones et al., 2009; amniotes: Myhrvold et al., 2015; amphibians: "AmphiBIO", Oliveira et al., 2017). These datasets stem from important efforts to collate published estimates of trait data and make them readily available. Trait data have also been made available on online platforms [for instance, the Global Assessment of Reptile Distribution initiative: <http://www.gardinitiative.org/>; the International Union for Conservation of Nature (IUCN) Red List of Threatened Species: <https://www.iucnredlist.org/>; BirdLife data zone: <http://datazone.birdlife.org/home>].

Nevertheless, despite the importance of vertebrate species in global research outputs, there is no single source for vertebrate ecological traits. Consequently, researchers wishing to conduct comparative studies across vertebrate groups might have to collate trait data from a range of sources (such as in the studies by Cooke, Bates, & Eigenbrod, 2019; Cooke, Eigenbrod, & Bates, 2019; González-Suárez et al., 2018), a time-consuming prerequisite that might be a limiting step of the research process. Indeed, collating data from heterogeneously formatted sources presents many challenges (Schneider et al., 2019), particularly when working across a large number of species. For instance, traits might be measured differently across datasets; units might be inconsistent; and taxonomic resolution and nomenclature might vary.

The lack of a curated, readily available global database for vertebrate ecological traits impedes our ability to conduct cross-taxon

BOX 1 Definitions

Trait: *Sensu stricto*, a characteristic measurable at the level of an individual and that influences organismal fitness or performance (Violle et al., 2007). In this paper, we broaden this definition to include "ecological" traits (e.g., number of habitats used by a species), where the relationship of a species to the surrounding environment needs to be considered. Ecological traits are estimated by aggregating data across multiple individuals.

Trait completeness: For a given species, the proportion of traits for which an estimate is available.

Trait coverage: For a given trait, the proportion of species for which an estimate is available.

comparative studies at global scales. However, efforts to collate data into a single database are limited by the availability of underlying data. Given that there are important gaps in biodiversity knowledge (Hortal et al., 2015), trait datasets are often incomplete, with many species lacking estimates for many traits. The incompleteness of ecological trait data at the species level has been termed the "Raunkiaeran shortfall" by Hortal et al. (2015). Furthermore, incomplete trait data are likely to be biased. Biases in trait data can be the consequence of uneven taxonomic and spatial collection effort, with a set of charismatic or easily detectable species being more completely sampled. For instance, González-Suárez et al. (2012) investigated biases in global trait information in mammals. Notably, they found that the availability of mammalian trait data were geographically and phylogenetically biased, with larger and more widely distributed species being better sampled. In addition, data availability also differed across IUCN Red List extinction risk categories, with threatened species (Critically Endangered, Endangered or Vulnerable) being less well sampled for traits than non-threatened species (Least Concern or Near Threatened).

A major issue with incomplete, biased data is the introduction of bias in subsequent analyses. Assessing the amount of missing data in addition to the so-called "missingness mechanism" (whether missing data are missing at random, as opposed to there being systematic biases in the way missing values are distributed; see Baraldi & Enders, 2010) is an important prerequisite. Indeed, there exist diverse techniques to deal with data missingness. The simplest one consists of retaining complete cases only by filtering out missing values (case deletion; see Nakagawa & Freckleton (2008)). Nevertheless, case deletion can lead to biased parameter estimates and erroneous

conclusions when values are not missing at random (González-Suárez et al., 2012). Therefore, it is crucial to determine the most appropriate way to deal with data incompleteness. For instance, previous studies using terrestrial vertebrate trait data have implemented multiple imputation techniques to fill in the gaps (Cooke, Bates, et al., 2019; González-Suárez et al., 2012). Nevertheless, imputation techniques could be sensitive to non-randomness in trait data. Phylogenetic biases (where some clades are undersampled compared with other clades) could notably impact the performance of several imputation approaches. It is thus vital to characterize the gaps in trait data before any analysis. However, there has been no study to date investigating global patterns in the availability of trait data across terrestrial vertebrates.

Here, we aim to assess the global state of trait data in terrestrial vertebrates. We focus on a set of traits that are available across the four classes and that are commonly used by ecologists: body size; litter or clutch size; longevity; trophic level; activity time; habitat breath; and a measure of habitat specialization. We quantify and compare the gaps in trait data across classes by calculating the coverage of each trait across species and the completeness of trait estimates for each species (Box 1). We investigate taxonomic, spatial and phylogenetic biases in trait coverage and completeness.

Given that biodiversity research is biased globally towards birds and mammals (Titley et al., 2017), we hypothesize that herptiles are less well sampled for traits than mammals and birds, having both lower coverage and completeness.

Furthermore, building upon previous studies conducted on mammals (González-Suárez et al., 2012), we hypothesize that species rarity influences completeness, focusing on the geographical range size of species as one aspect of rarity. Widely distributed species could be better sampled than narrowly distributed species because their ranges overlap with more study sites, regardless of their abundance. As such, we test whether the geographical range size of species explains trait completeness.

It is well established that global research effort is distributed unequally (United Nations Educational Scientific and Cultural & Organization, 2015), with patterns underpinned by various geographical and socio-economic factors. For instance, countries with higher gross domestic product tend to host a larger number of research institutions (Martin, Blossey, & Ellis, 2012). The proximity of research infrastructures and the accessibility of survey sites play an important part in explaining the global distribution of knowledge (Hortal et al., 2015). As a result of these factors, biodiversity data gaps tend to be greater in tropical areas (Collen et al., 2008). Tropical areas have the greatest species richness, and therefore these data biases are of great concern for biodiversity conservation. It is thus important to assess whether species-rich regions are systematically undersampled for traits compared with species-poor regions, given the significance of species-rich areas for global conservation. Here, we investigate spatial biases in trait completeness, hypothesizing that species-rich areas are on average less well sampled than species-poor areas.

Finally, we investigate phylogenetic biases in the trait data. We assess whether particular clades have received more attention than others by looking for patterns in the distribution of trait completeness across the terminal branches of phylogenetic trees in each class.

2 | METHODS

We produced class-specific trait datasets that we have made available on figshare (<https://doi.org/10.6084/m9.figshare.10075421>).

Data compilation and all analyses were conducted with R v.3.5.1 (R Core Team, 2018). Distribution maps were processed using both R and the ArcPy package available in ArcGIS v.10.6 (ESRI, 2011) (implemented in Python v.2.7; van Rossum, 1995).

2.1 | Trait data collection

2.1.1 | Sources and taxonomic matching

We used freely accessible secondary sources in our compilation (Table 1), selected for their broad taxonomic coverage and/or for their frequent use in macroecological studies. Across sources, similar species could appear under synonymic names. This was a potential problem for matching sources by binomial names. Indeed, synonymy can artefactually decrease trait coverage, when trait information is not available across all synonyms. Notably, difficulties arise when species have been divided into several subspecies or when different subspecies are clumped together. Systematic manual checks could not be applied considering the scale of the data collection (there were >39,000 unique binomial names across sources). We developed a procedure aiming at identifying one accepted name for each of the binomial names found across sources. When we could not find an accepted name, we used the original name. Figure 1 summarizes the main steps; similar solutions have been used in other large-scale studies (Cooke, Bates, et al., 2019).

Briefly, the procedure consisted of extracting synonyms from the IUCN (IUCN, 2020) or from the Integrated Taxonomic Information System (ITIS; <https://www.itis.gov/>), using the rredlist (Chamberlain, 2018) and taxize (Chamberlain & Szöcs, 2013) R packages. One accepted name was assigned to each synonym. We produced a "Synonym" dataset that we have also made available. We then normalized taxonomy across sources by replacing binomial names with their identified accepted name where applicable.

Given that different taxonomic backbones could be used to correct for taxonomy, we make two versions of our trait compilations available (corrected and not corrected for taxonomy), meaning that users are free to apply their own corrections; for example, taxonomy could be aligned to that of class-specific sources, such as The Reptile Database, the American Museum of Natural History's Amphibian Species of the World, the Mammal Diversity Database or the International Ornithological Congress World Bird List. Datasets corrected for taxonomy contain 11,634 species of birds,

TABLE 1 Data sources for each trait

Sources	Taxa	Traits									
		BM	BL	L/ML	MA	GL	LCS	TL	DA	RS	H
Oliveira et al. (2017)	Amphibians	(✓)	✓	(✓)	✓		✓	~✓	✓		
Cooper et al. (2008)							✓				
Sodhi et al. (2008)			✓								
Wilman et al. (2014)	Birds	✓						~✓	✓		
BirdLife ^a		✓				✓					✓
Jones et al. (2009)	Mammals	✓	(✓)	(✓)	(✓)		✓		✓		
Kissling et al. (2014)								✓			
Gainsbury et al. (2018)								✓			
Wilman et al. (2014)		✓								✓	
Pacifici et al. (2015)		✓				✓					
Scharf et al. (2015)	Reptiles	✓		✓	(✓)		✓	✓	✓		
Vidan et al. (2017)										✓	
Stark et al. (2018)		✓		✓			✓		✓		
Schwarz and Meiri (2017)							✓				
Novosolov et al. (2017)		✓						✓			
Novosolov et al. (2013)							✓				
Slavenko et al. (2016)		✓									
Feldman et al. (2016)		✓									
Meiri (2018)					✓		✓	✓	✓		
Meiri et al. (2015)								✓	✓		
Roll et al. (2017)											✓
Myhrvold et al. (2015)	B, M, R	✓	✓	✓	(✓)		✓				
IUCN (2020) ^b	A, B, M										✓
IUCN (2020) ^c	All										✓

Abbreviations: A = amphibians; B = birds; BL = body length; BM = body mass; DA = diel activity time; GL = generation length; H = habitat data; LCS = litter or clutch size; L/ML = longevity or maximum longevity; M = mammals; MA = age at sexual maturity; R = reptiles; RS = range size; TL = trophic level. Note. Data sources may contain more traits than shown here. Tick marks in parentheses indicate that the trait was present in the data source but that another closely related trait with a better coverage was used instead. The tilde character (~) before a tick mark indicates that we derived trophic levels from species diet.

^a<http://datazone.birdlife.org/home>

^b<https://www.iucnredlist.org/resources/spatial-data-download>

^c<http://apiv3.iucnredlist.org/api/v3/docs#general>

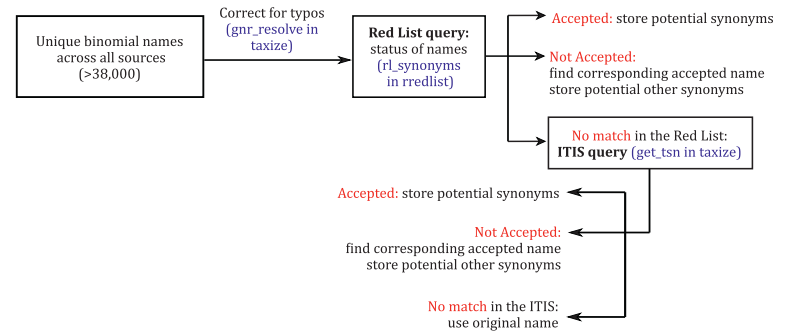
5,381 mammals, 10,612 reptiles and 6,990 amphibians. Where no taxonomic correction was applied when matching sources, the compiled datasets contain 13,501 birds, 5,791 mammals, 11,012 reptiles and 8,583 amphibians. For more information, see the Supporting Information (Appendix S1; Figure S1).

2.1.2 | Compilation methods

For continuous traits, we took the median value within species when multiple estimates were available from different sources, after

removal of any repeated values, which were assumed to represent estimates duplicated across secondary compilations and derived from the same underlying primary sources. Although intraspecific variation is increasingly being recognized to have important effects on ecological systems (Bolnick et al., 2011; Des Roches et al., 2018; González-Suárez & Revilla, 2013; Siefert et al., 2015), it was not feasible to obtain measures of intraspecific variability from all sources; therefore, estimates were provided as a single measure for each species. For some species and some traits, measures were provided separately for females and males. In such cases, we first obtained the mean of these two measures.

FIGURE 1 Procedure used to identify the accepted names of species. We extracted, where possible, the accepted names of species from either the Red List or the Integrated Taxonomic Information System (ITIS).



Across sources, there were multiple traits related to each of body size and life span. For instance, body mass and/or body length information could be provided. Different proxies were also available for life span, such as the age at sexual maturity or generation length. In such cases, we focused on the trait presenting the highest coverage.

Body size

Adult body mass estimates were compiled for mammals, birds and reptiles. Body length information was compiled for amphibians, because the coverage for body length was higher than that for body mass. Body mass and body length are known to scale allometrically, although the allometric relationship differs across amphibian clades (Santini et al., 2018). In our amphibian dataset, Pearson's correlation coefficient between $\log(\text{Body mass})$ and $\log(\text{Body length})$ was .71 (data points shown in Supporting Information Figure S2).

Longevity

We defined longevity as the life span of an individual and maximum longevity as the longest life span reported. We used closely related traits when longevity/maximum longevity was not available or when longevity/maximum longevity had a poorer coverage than a related trait. We selected the age at sexual maturity for amphibians; Pearson's correlation coefficient between $\log(\text{Age at sexual maturity})$ and $\log(\text{Maximum longevity})$ was .55 (Supporting Information Figure S2). We compiled the generation length for mammals and birds. The correlation between $\log(\text{Generation length})$ and $\log(\text{Longevity})$ was .74 for mammals and .70 for birds (Supporting Information Figure S3). Finally, we used maximum longevity directly for reptiles.

Litter or clutch size

The number of offspring (litter size) or eggs (clutch size) was compiled directly from the sources and treated as equivalent across classes. We reported measures of central tendencies provided by the sources where applicable; otherwise, we calculated range mid-points (mean of smallest and largest reported litter/clutch sizes).

Trophic level

In all classes, species were described as omnivores, carnivores or herbivores. For reptiles and mammals, this information was compiled

directly from the sources. For amphibians and birds, trophic levels were not provided. For these two classes, we inferred trophic levels from dietary information (Table 1). For birds, we used the primary diet (based on food items recorded as composing $\geq 50\%$ of the diet of a species). Diet for amphibians was described without respect to the percentage use of food items; simply as a binary record of whether or not food items were used. In both cases, species recorded to only consume plant-based resources (seeds, nectar, fruit or other plant material) were classified as herbivores. Species consuming only animal resources (invertebrates or vertebrates) were classified as carnivores. Species consuming a mixture of plant and animal resources were classified as omnivores.

Activity time

Species were described as being either nocturnal or non-nocturnal. Despite a higher resolution of activity time information in some of the sources (e.g., species being described as cathemeral, crepuscular or diurnal), we adopted the classification of the source with the lowest resolution (EltonTraits: Wilman et al. (2014), for birds), in order to have consistent information across classes. As such, all species defined as diurnal, cathemeral or crepuscular were classified as non-nocturnal, as opposed to species classified as strictly nocturnal.

Habitat breadth

We used IUCN habitat data (IUCN, 2020), which describe species habitat preferences and the suitability and importance of each habitat. We defined habitat breadth as the number of habitats a species was known to use, using level 2 of the IUCN Habitat Classification Scheme for description of habitat types (divided into: Forest, Savanna, Shrubland, Grassland, Wetland, Rocky areas, Caves and subterranean, Desert, Marine, Marine intertidal or coastal/supratidal, Artificial, Introduced vegetation, and Other/Unknown.) Note that the sum could be weighted to account for habitat suitability and importance, but we used unweighted sums in this work.

Use of artificial habitats

For a species, we recorded whether any artificial habitat was reported to be suitable in the IUCN habitat data.

Finally, our compiled datasets contain an additional column, "Note", where we reported species found to be extinct or extinct in the wild (EW). We used species Red List status and information

from Meiri (2018) to flag such species. We reported 75 extinct/EW species for mammals, 160 for birds, 34 for amphibians and 53 for reptiles. It is likely that our datasets contain extinct species that we could not flag, because they were not recorded as extinct in the sources we used.

2.1.3 | Phylogenies

We used class-specific phylogenetic trees downloaded on 13 April 2020. For mammals, we used “complete” trees from Faurby et al., (2018, 2020), downloaded from <https://zenodo.org/record/3690867#.Xyc5wyhKhPZ>. For amphibians, birds and squamates, we obtained trees from <https://data.vertlife.org/>. The original sources were as follows: Jetz et al. (2012) for birds; Jetz & Pyron (2018) for amphibians; and Tonini et al. (2016) for squamates. For each class, a distribution of 1,000 trees was available. For plotting purposes, we obtained consensus trees using the TreeAnnotator program of the BEAST software (Bouckaert et al., 2014).

2.1.4 | Species distributions

We obtained extent-of-occurrence distribution maps for reptiles from Roll et al. (2017), available at: <https://datadryad.org/stash/dataset/doi:10.5061/dryad.83s7k> (downloaded 13 April 2020). For mammals and amphibians, species distribution maps were obtained from the IUCN Red List (IUCN (2020), downloaded 13 April 2020); for birds, they were obtained from BirdLife International (<http://datazone.birdlife.org/species/requestdis>, downloaded 17 April 2020).

For amphibians, mammals and birds, we selected areas of extant or probably extant presence only. Additionally, we selected areas where species were resident or present during the breeding season, and we excluded areas occupied during the non-breeding season or where species were considered vagrant.

In addition, for all classes, we excluded occupied areas that fell outside the known elevational limits of species, where such data were available. Lower and upper elevational limits were retrieved from the IUCN Red List (queried using the `rredlist` package) and were available for approximately half of the species (Supporting Information Figure S4). Decreases in range sizes were observed after cutting distribution maps by the known elevational limits (Supporting Information Figure S5).

2.2 | Investigating gaps and biases in trait data

We used trait coverage and completeness to investigate taxonomic, phylogenetic and spatial biases in the trait data. Table 2 summarizes the sample sizes (number of species) in each of the following analyses. Note that species for which completeness was 0% were included in all analyses (for more details, see Figure 2). Also note that we did

TABLE 2 Number of species for each analysis

	(a) Taxonomic biases	(b) Phylogenetic biases	(c) Spatial biases
Amphibians	6,990	6,170	5,650
Birds	11,634	8,315	10,802
Mammals	5,381	5,171	5,046
Reptiles	10,612	9,404	9,382

All species represented in the trait datasets were included in (a). All species from the class-specific phylogenetic trees or from the distribution maps that matched with species in the trait datasets were included in (b) and (c).

not filter out species identified as extinct or extinct in the wild, because they represented a small proportion of the datasets (.48% for amphibians, 1.4% for both birds and mammals, and .50% for reptiles) and also because we could not exclude such species systematically, because it is likely that we did not flag them all.

2.2.1 | Taxonomic biases

We tested whether completeness varied across taxonomic class using pairwise Wilcoxon rank sum tests. We tested for the extent and performance of our taxonomic corrections by looking at trait coverage when taxonomic corrections are applied and when no correction is applied (Supporting Information Figure S6).

2.2.2 | Phylogenetic biases

Initially, to assess whether more closely related species were more likely to be similar in trait completeness, we estimated the phylogenetic signal in completeness with Pagel's λ (Pagel, 1999) in each class. We used a bootstrapping approach, calculating λ for each of 50 trees randomly sampled in each class (using the `phylosig` function of the `phytools` R package; Revell, 2012). We then estimated the mean and 95% confidence intervals (95% CIs) of λ . Sample sizes for computing λ (number of species represented in both the phylogenies and trait datasets) are shown in Table 2.

We then plotted within-family median completeness in phylogenetic trees built at the family level, using the consensus trees. Within-family median completeness was calculated using taxonomic information in the trait datasets (sample sizes shown in Table 2).

2.2.3 | Spatial biases

We first investigated whether wider-ranging species were more likely to be better sampled than narrow-ranging species. We tested for a relationship between species range size and trait completeness. We fitted a generalized linear model with a Poisson error distribution [directly using the number of sampled traits, “ N_{traits} ”, rather than

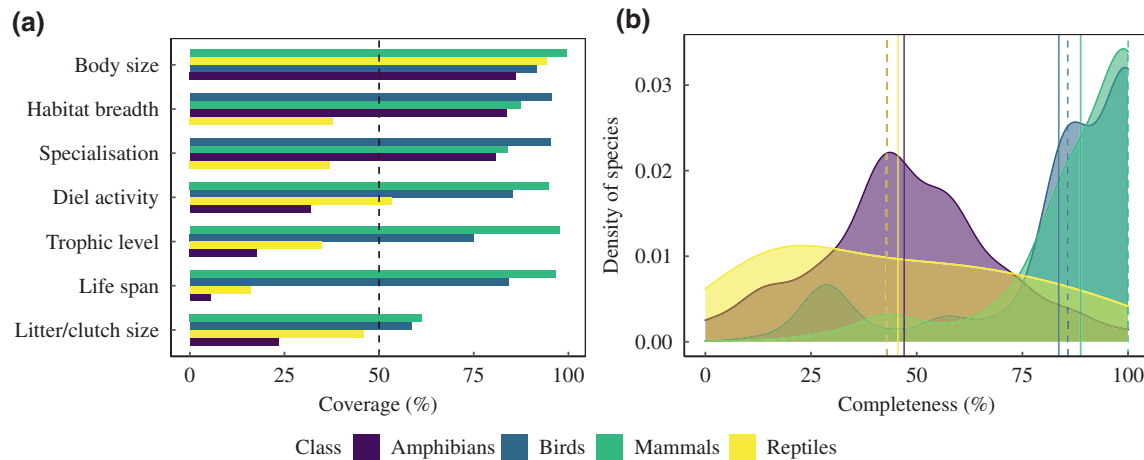


FIGURE 2 Trait coverage and completeness across species. (a) We defined coverage as the proportion of species for which an estimate is available for a given trait. The dashed line represents 50% coverage. (b) Trait completeness is the proportion of estimated traits for a species. Here, we show the distribution of completeness. Continuous lines represent the mean trait completeness for each class, whereas dashed lines represent the median trait completeness. Note that there were species with 0% completeness (230 species for amphibians, 3.3% of amphibian species in our trait data; nine for birds, .077% of species; seven for mammals, .13% of species; and 161 for reptiles, 1.5% of species). Species with 0% completeness were retained in the datasets when there was information for traits we did not select in the analyses but no known value for the traits we did select. For instance, the body mass of the amphibian species *Rhinella centralis* was known, but other trait values (including body length) were missing, meaning that *Rhinella centralis* had 0% completeness for the set of traits we considered.

the proportion (completeness)]. Class was added as a predictor interacting with range size; thus the model was:

$$N_{\text{traits}} \sim \log(\text{Range size}) \times \text{Class}.$$

Second, we mapped assemblage-level median completeness. Assemblages were characterized at the pixel level at 50 km² resolution. We determined pixel-level composition and richness by stacking species geographical distributions. We then calculated median completeness across species in each pixel. We show the resulting maps for herptiles in the main text, and for mammals and birds in Supporting Information Figure S7 (median completeness was very high across most pixels for mammals and birds). In addition, we provide maps of assemblage-level mean completeness and standard deviation for all classes in the Supporting Information (Figures S8 and S9 show maps; Figure S10 shows standard deviation against species richness).

We then tested for a spatial correlation between species richness and median completeness. Given that median completeness was very high across most pixels for mammals and birds, we fitted such models for herptiles only. We fitted spatial autoregressive lag models to explain assemblage-level median completeness as a function of species richness [using the function `lagsarlm` of the `spatialreg` package (Bivand, Pebesma, & Gómez-Rubio, 2013; Bivand & Piras, 2015; Bivand, Hauke, & Kossowski, 2013)]. Given that responses could vary geographically, we included the biogeographical realm as an interacting factor [using the World Wide Fund for Nature (WWF) ecoregion shapefile to characterise realms, obtained from <https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world>]; the considered realms were Afrotropics, Australasia, Indo-Malayan, Nearctic, Neotropics and Palearctic. To improve

normality, we arc-sin square-root transformed completeness values and log-transformed species richness. The `lagsarlm` function allows for a consideration of spatial autocorrelation in the dependent variable by estimating the autoregressive lag coefficient, ρ , associated with an n -by- n matrix of spatial weights, W . The final model was:

$$\arcsin(\sqrt{\text{Completeness}}) \sim \log(\text{Species richness}) \times \text{realm} + \rho \cdot W \cdot \arcsin(\sqrt{\text{Completeness}}).$$

The value of W was estimated using the functions `tri2nb` and `nb2listw` of the `spdep` package (Bivand, Pebesma, et al., 2013; Bivand & Wong, 2018). Fitting the model using all grid cells was computationally intractable; therefore, we randomly sampled cells for this analysis (using 30% of the grid cells in each realm). We selected grid cells where species richness was higher than three to avoid sampling issues. We fitted separate models for amphibians and reptiles, because when adding class as an interacting predictor, the same cells (with the same coordinates) might be sampled for multiple classes, whereas the `tri2nb` function does not tolerate duplicated coordinates.

3 | Results

3.1 | Taxonomic biases in trait information

Trait coverage for mammals and birds was high overall (Figure 2a; mean and median coverage across traits: 89% and 95% for mammals; 84% and 85% for birds). In both cases, litter/clutch size was the trait with the poorest coverage (61% for mammals and 59% for birds). Coverage exceeded 80% for all other traits (except trophic level for birds, at 75% coverage).

Conversely, trait coverage was more variable for herptiles, and poorer overall (Figure 2a; mean and median trait coverage: 47% and 32% for amphibians; 46% and 38% for reptiles). Coverage exceeded 80% only for body size in both reptiles and amphibians and for habitat-related traits in amphibians only. In all other cases, coverage was <55%, with very little information available for longevity-related traits.

Trait completeness (proportion of non-missing trait values for a species) reflected similar biases (Figure 2b). The distribution of trait completeness varied significantly among classes (pairwise Wilcoxon rank sum test: $p < .0001$ in all cases). Distributions were highly left skewed in mammals and birds (skewness: -2 and -1.6). Eighty-four per cent of all mammalian species and 80% of avian species fell in the 80%–100% completeness range. Moreover, the completeness distribution was moderately right skewed for reptiles (skewness: $.4$) and slightly right skewed for amphibians (skewness: $.02$). Fifty-six per cent of all reptiles and 57% of amphibians fell in the 0%–50% completeness range.

3.2 | Phylogenetic biases in trait completeness

As expected from the distribution of trait completeness in mammals and birds (Figure 2), within-family median trait completeness was high across most tips of the phylogenetic trees (Supporting Information Figure S11 and S12; we present the avian and mammalian phylogenies in the Supporting Information because there was little variation in completeness across tips). For birds, λ was $.71$ ($\pm .0053$). For mammals, λ was $.78$ ($\pm .0035$). This indicated that, despite completeness generally being high across tips, the sampling was not evenly distributed across the phylogeny.

In herptiles, clusters of families with similar median trait completeness appeared (Figure 3). In amphibians, groups of families belonging to the order Anura (frogs) showed both the best and worst median completeness (Figure 3a). The best-sampled families included the tailed frogs of the family Ascaphidae (two species) and species of the family Leiopelmatidae (four species endemic to New Zealand). The family Ceratobatrachidae (containing c. 90 species occurring in Southeast Asia and in some Pacific islands), the family Ranidae (true frogs, 450 species considered here) and the family Rhacophoridae (shrub frogs, 382 species considered here) figured among the worst-sampled families. For amphibians, λ was $.63$ ($\pm .0039$). In reptiles, most snakes were poorly sampled, whereas families in other suborders appeared to be sampled better overall (Figure 3b). Within snakes, the pythons, boas, the three species of the family Acrochordidae and the python-like species of the family Loxocemidae were better sampled than other snake families. In reptiles, λ was $.69$ ($\pm .0032$). The sampling in herptiles was thus also uneven with regard to the phylogeny.

It is important to underline that Figure 3 shows within-family median completeness, masking the considerable variation in species richness across families, hence masking potential important variation in completeness across species within families. For example, in

the amphibian family Allophrynidae (three recognized species), the within-family median completeness was 50%; but our dataset comprised two species of completeness 14% and 86%, respectively. We present similar plots to those in Figure 3 showing the within-family standard deviation in completeness in the Supporting Information (Figure S13). Within-family standard deviation tended to increase with within-family species richness (Supporting Information Figure S14).

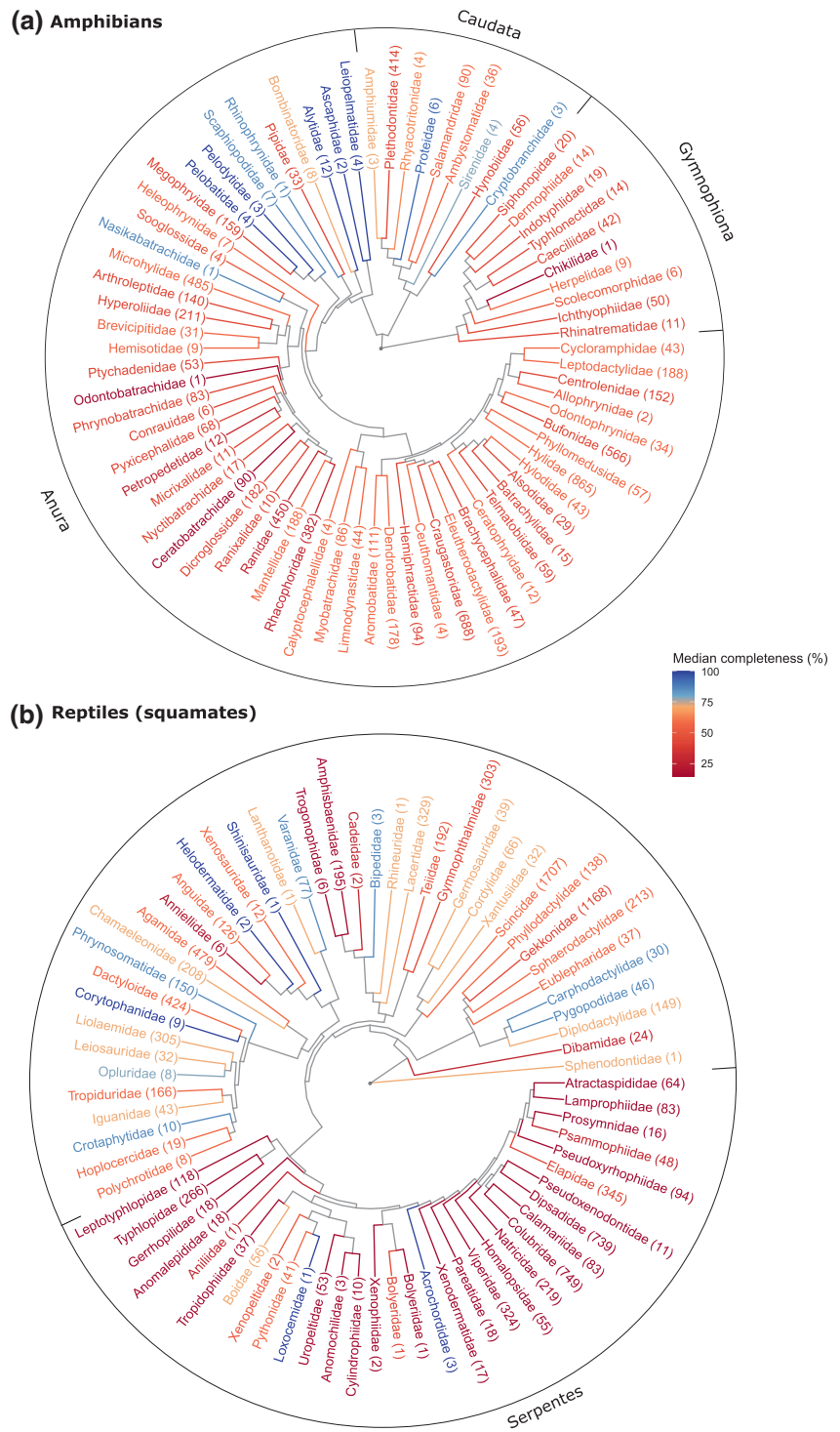
3.3 | Spatial biases in trait completeness

Range size was significantly correlated with the number of sampled traits. Larger range sizes were associated with a higher number of sampled traits (i.e., with higher completeness; Figure 4; Supporting Information Table S1). Similar results were obtained when using distribution maps not cut by elevational limits (Supporting Information Table S2; Figure S15). The rate of increase was steepest for reptiles, then for amphibians, then for birds and mammals (slope estimates for birds and mammals were not significantly different from each other; Supporting Information Table S1).

There were marked spatial variations in median trait completeness in herptiles (Figure 5). North America and Europe were well sampled for both amphibians and reptiles. Moreover, Southeast Asia and the Congo basin were on average less well sampled. In other regions, contrasting patterns emerged between amphibians and reptiles. For instance, median completeness was poorer for amphibians than for reptiles in Australia, but opposite patterns were observed in South America. As in the phylogenetic analyses, assemblage-level median completeness could mask potential important variation in completeness within species of a given assemblage. Assemblage-level mean and standard deviation maps are shown in the Supporting Information (Figures S8 and S9). There was a trend for increasing standard deviation with increasing species richness, with a larger spread in standard deviation at lower species richness (Supporting Information Figure S10).

Spatial models showed that species richness explained median trait completeness in herptiles in most realms (Figure 6; Supporting Information Tables S3 and S4); including spatial lags improved the models (reptiles: $\rho = .91$, $p < .0001$; amphibians: $\rho = .92$, $p < .0001$). For reptiles, completeness was negatively correlated with species richness in the most species-rich realms (Afrotropics, Indo-Malayan and Neotropics) and in the Palaearctic; the relationship was steepest in the Afrotropics and shallowest in the Palaearctic. In the Australasian and Nearctic realms, completeness tended to increase with species richness. For amphibians, negative relationships were observed in the Indo-Malay and Nearctic realms, whereas positive trends were observed in the Neotropics and the Palaearctic. The opposite trends between reptiles and amphibians observed in the Australasian and Neotropical realms reflected patterns observed on the maps. The Indo-Malayan was the only realm where median completeness tended to decrease with species richness for both reptiles and amphibians.

FIGURE 3 Within-family median trait completeness in herptiles. The number next to each family name represents the number of species included in the calculation of the median.



4 | Discussion

Our work illustrates the taxonomic, spatial and phylogenetic dimensions of the knowledge gaps in trait data, termed the Raunkiaeran shortfall by Hortal et al. (2015). To the best of our knowledge, this study constitutes the first comparative assessment of global gaps for terrestrial vertebrate trait data, despite their use in numerous studies. We showed that the trait data presents important taxonomic, spatial and phylogenetic biases, with contrasts in the availability of

trait information between, on the one hand, herptiles and, on the other hand, birds and mammals.

Birds and mammals are globally well sampled for the set of traits we considered, even in the most species-rich assemblages. Moreover, the availability of trait information for herptiles is lower overall and phylogenetically and geographically biased. Several factors could interplay to shape these patterns. For instance, species that are more easily detectable (for example, wider ranging) and more charismatic are likely to be better sampled. Diverse socio-economic

predictors could also contribute to geographical biases in trait data sampling; global biases in primary data collection are likely to be one of the most important contributors to the patterns we observe. Nevertheless, biases in the data could have been introduced at later stages, notably with the selection of sources and traits. Our global compilation reflects, in part, the interest and focus of the secondary data sources we used. It is possible that the addition of new sources from regional journals or other authorities could diminish

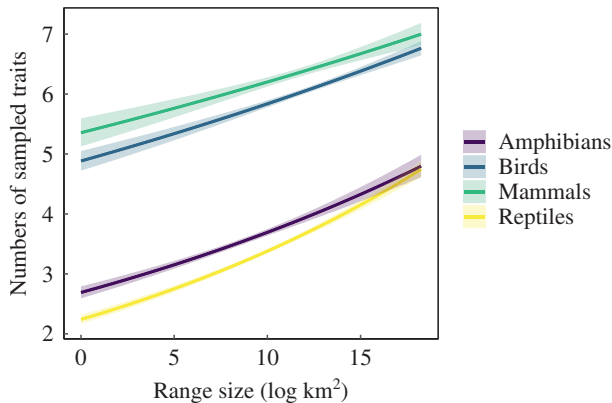


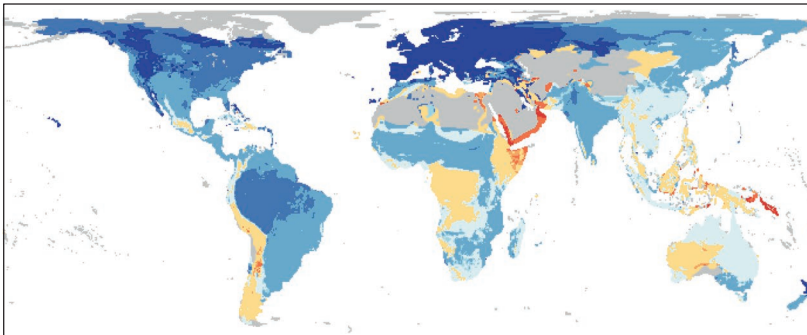
FIGURE 4 Relationship between number of sampled traits and geographical range size. Models were fitted using a Poisson error distribution. Class was added as a predictor interacting with range size. Rates of increase were not significantly different for mammals and birds but differed for reptiles and amphibians, with the steepest rates of increase for reptiles.

spatial biases in the data by increasing coverage for certain areas. Nevertheless, we argue that by focusing on widely used traits, our results are likely to reflect the “true” availability of the data in primary sources and that the shortfalls for other, less used traits would be more pronounced.

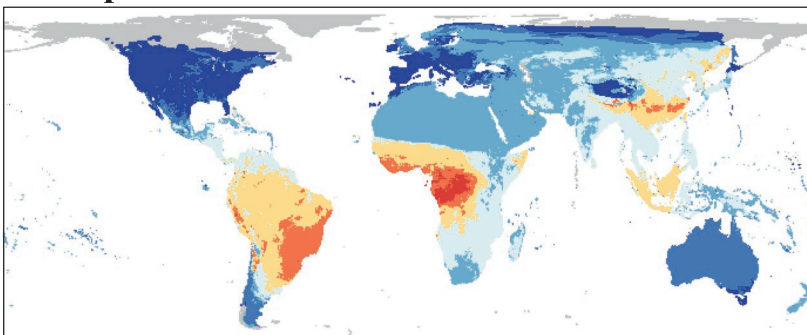
We believe that our results are robust to taxonomic uncertainty, although taxonomic matching might potentially be improved further using class-specific sources, such as the Reptile Database or AmphibiaWeb, for identification of synonyms (but see Supporting Information Appendix S9, Figure S16). We have made two versions of our data compilations available, one in which our own corrections were applied and one using the original binomial names of the sources, meaning that users are free to use their own taxonomic backbones and identify synonyms within the compilations.

We believe that taxonomic matching is a recurring issue when working across thousands of species. Taxonomic synonymy artefactually inflates the numbers of identified species, potentially lowering trait coverage (whereas clumping subspecies together can have the opposite effect). Tackling this problem is difficult (Isaac et al., 2004; Jones et al., 2012), notably because there is no global curated database recording the status of species names, and also because of the nature of taxonomy and the debates around the species concept (May, 2011). Nevertheless, taxonomic uncertainty can have important consequences. For instance, Cardoso et al. (2017) showed that inaccuracies and errors in species checklists contributed to the overestimation of plant diversity in the Amazon.

(a) Amphibians



(b) Reptiles



Median completeness:

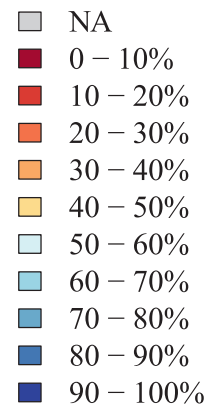


FIGURE 5 Spatial distribution of assemblage-level median trait completeness in herptiles. Similar maps for birds and mammals are shown in the Supporting Information (Figure S7).

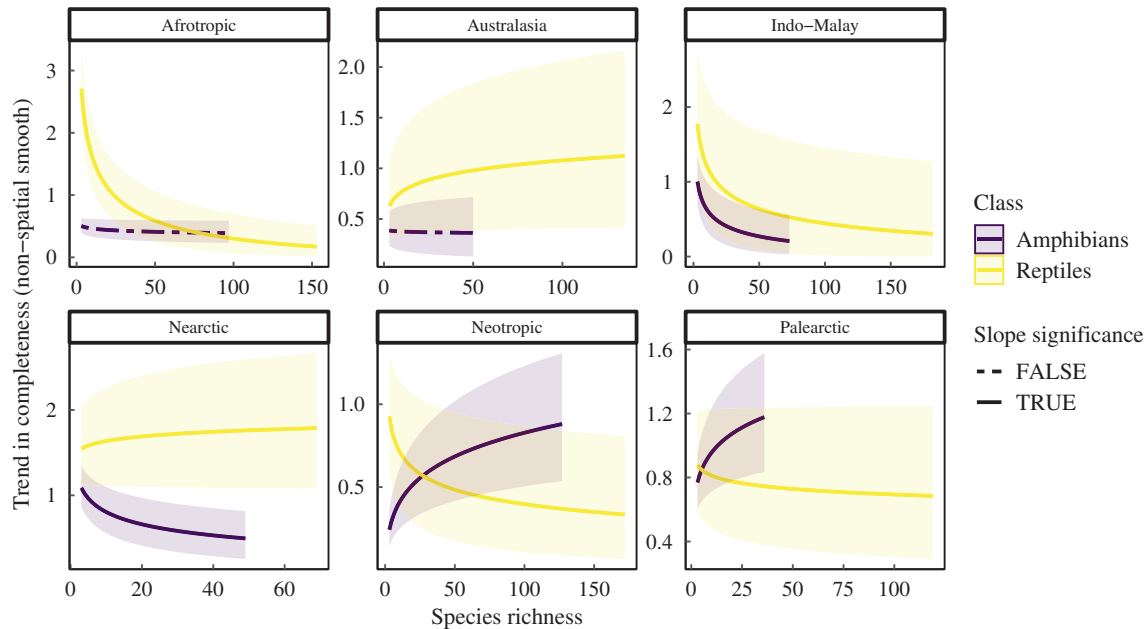


FIGURE 6 Spatial model trends for herptiles. The lines represent in-sample predictions (\pm SE) for the trend components of the spatial models (trends after accounting for spatial autocorrelation).

Biases in trait data have important implications for conservation planning. Past studies have shown that narrow-ranged species, for which fewer trait data are available on average, have higher extinction risks (Collen et al., 2016; Purvis et al., 2000; Ripple et al., 2017) and are more negatively impacted by anthropogenic pressures (Newbold et al., 2018) than wider-ranging species. Trait information is also less available for herptiles in tropical regions such as the Congo basin, Southeast Asia and South America, which are some of the most diverse areas of crucial importance for worldwide conservation (Barlow et al., 2018). Consequently, trait information is on average less available where potentially more crucial to conservation planning. Indeed, trait information can be incorporated into vulnerability assessments and, as such, can help to prioritize conservation efforts. Species traits have been found to mediate species responses to environmental changes across diverse taxonomic groups, and thus can inform on the sensitivity of species to anthropogenic pressures (Flynn et al., 2009; Newbold et al., 2013; Nowakowski et al., 2017). Traits are now commonly used to estimate species vulnerability or extinction risks (Pacifci et al., 2015; Ramirez-Bautista et al., 2020). As opposed to trend-based approaches, which rely on historical population trends (changes in abundance or shifts in distributions) to predict species' vulnerability and extinction risks, trait-based approaches rely on species' intrinsic sensitivity to particular threats. The appeal of trait-based approaches to extinction risk estimation is that, by providing mechanistic insights, they diminish the amount of population information needed. If the responses of species to a threat consistently relate to certain traits, it is possible to generalize patterns across species for which population data are less available (Verberk et al., 2013). Integrating traits into vulnerability assessments is hence of particular interest when field monitoring of species population sizes or distributions is difficult to achieve, but

biases in the data could mean that such information is lacking for some of the most vulnerable species.

Traits that influence species responses to environmental changes have been termed "response traits" (or "response-mediating traits"; Luck et al., 2012), as opposed to "effect traits" that underpin ecosystem functioning (Lavorel & Garnier, 2002). For instance, relative brain size and longevity have been characterized as response traits in birds (Newbold et al., 2013; Sayol et al., 2020), whereas dietary characteristics (e.g., trophic levels or guilds) are both response and effect traits. Hortal et al. (2015) highlighted that, for plants, both response and effect traits have been investigated, whereas for vertebrates the research has been more focused on understanding species responses. This could be because the way vertebrate traits interact to shape some ecosystem processes has not yet been characterized well.

Ecosystem processes sustained by animals might be harder to quantify and might be influenced by a combination of traits. The traits compiled in this work are likely to have a role in diverse processes. Nevertheless, there was one important omission, in that we did not compile species diet, potentially the most straightforward trait to link with diverse processes, such as grazing, pollination, scavenging and seed dispersal. From a practical perspective, we chose traits that had been estimated at least for some of the species in each class, and that were readily available. Diet was excluded because although estimates were available for amphibians, birds and mammals, there was no readily available database for reptilian diet. Movement or dispersal abilities were also excluded because information was not readily available for any class. Although we expect that species diet and dispersal abilities would present similar sampling biases to the ones presented in this work, the addition of such traits to the compilation would represent a valuable contribution

and would notably facilitate studies looking at the functional roles of reptiles.

For practical reasons, we did not consider intraspecific trait variation. Intraspecific variation has been shown to have important effects on ecological systems, and a growing body of literature encourages trait-based research to include intraspecific variability (Guralnick et al., 2016). There have been several calls to produce open-access, global trait datasets (Weiss & Ray, 2019), including a representation of intraspecific trait variation (Kissling et al., 2018). Notably, Schneider et al. (2019) designed a framework to store and share inter- and intraspecific trait data, accompanied by an R package to standardize the data in a proposed format. Such a proposition could constitute an important step towards the unification of individual datasets into a single, comprehensive database for ecological trait data.

The current spatial and taxonomic gaps in trait data might limit our ability to scale studies up, whereas biases in the data can affect the validity of extrapolations to groups or areas that are under-sampled. More generally, biases and gaps in biodiversity data can have important implications for ecological studies. Data gaps can hinder our ability to draw conclusions on observed macroecological patterns. For example, Chaudhary et al. (2016) proposed that marine species richness follows a bimodal distribution, peaking at mid-latitude locations, and argued that these patterns were not underpinned by knowledge gaps in species distributions. Moreover, Menegotto and Rangel (2018) attributed the tropical dip in marine species richness to a lack of species distribution data, explained by lower sampling efforts in tropical areas ("Wallacean" shortfall; Hortal et al., 2015). Biases and gaps in trait data could also affect studies in closely related fields, such as functional ecology [for instance, past studies have shown that functional diversity indices are sensitive to missing data (Májeková et al., 2016; Pakeman, 2014)] or community assembly (Perronne et al., 2017).

Ecologists should, therefore, take particular care when designing trait-based studies, because both data quality and data gaps are likely to influence the results and the generality of the conclusions. There exist diverse methods to deal with missing trait values, should data missingness be problematic. Complete removal of missing values ("case deletion") is commonly used but presents several issues, because it reduces sample size and statistical power and introduces potential bias in data subsamples (Nakagawa & Freckleton, 2008). For example, retaining complete cases only from our trait datasets would generate trait data disproportionately representative of mammals and birds, which would be problematic for conducting cross-taxon analysis on terrestrial vertebrates. As such, it is recommended that case deletion be applied only when data are missing completely at random, which is rarely the case (Peugh & Enders, 2004).

Alternatives to case deletion consist of filling in the gaps. In recent years, the development of imputation techniques has provided robust methods to handle missing data. Such imputation techniques have been used to complete trait datasets in recent studies (Cooke, Bates, et al., 2019). Penone et al. (2014) used a

simulation approach to evaluate the performance of four of these techniques, namely PhyloPars (Bruggeman et al., 2009), random forest algorithms as implemented in R with missForest (Stekhoven, 2016; Stekhoven & Bühlmann, 2012), multivariate imputation by chained equations (MICE; Van Buuren & Groothuis-Oudshoorn, 2011) and k-nearest neighbour (kNN; Troyanskaya et al., 2001). Penone et al. (2014) introduced missing values (10%–80%) in a complete trait dataset of carnivorans and measured imputation performance in different scenarios. Given that phylogenetic non-randomness in missing trait values can impact imputation accuracy, Penone et al. (2014) removed values in three different ways (completely at random; with a phylogenetic bias; and with a body mass bias). Out of the four techniques, missForest and PhyloPars performed best when species phylogenetic position was included as a predictor of missing trait values. Such imputations appeared to be robust even when trait coverage was as low as 40%, which might be relevant for many reptilian and amphibian traits. The performance was not significantly affected by phylogenetic non-randomness of the data. Hence, missForest and PhyloPars appear to be well suited when traits are phylogenetically conserved, because they allow species phylogenetic position to be included as a predictor of missing trait values. The study by Penone et al. (2014) highlights that there are robust imputation techniques allowing us to deal with incomplete trait data where biases might otherwise be problematic. Nevertheless, it is important to highlight that some imputation techniques, such as single or mean imputation, can be problematic because they do not allow an estimation of uncertainty and suffer from a lack of accuracy (Nakagawa & Freckleton, 2008); indeed, imputation techniques sometimes perform no better than case deletion. We believe that more work should be conducted to assess imputation performance in various contexts, and our compiled datasets might provide an opportunity for such studies.

Although robust imputation techniques can be useful for filling gaps in trait datasets, they are no substitute for continued data collection efforts. Our results show that data are particularly lacking in herptiles, notably in the Afrotropics, the Neotropics and the Indo-Malayan realms. For these areas, incorporating regional databases into existing datasets could contribute to the reduction of global gaps. We believe that both primary research and subsequent efforts to integrate new data and existing databases are required if we are to strive towards the unification of trait databases.

To conclude, this work constitutes, to our knowledge, the first assessment of the global gaps and biases in terrestrial vertebrate trait information. We show that herptiles are undersampled compared with mammals and birds, with important spatial and phylogenetic variability in the availability of trait information. Imputation techniques are one possible solution to these problems. Nevertheless, we believe that primary research, combined with efforts to complete existing datasets, is the only way to fill the current data gaps genuinely and robustly. We hope that the compiled trait dataset and our findings can prove useful for guiding further data collection efforts and for conducting macroecological analyses.

ACKNOWLEDGMENTS

We thank Stuart Butchart for contributing data and for his feedback on an earlier version of the manuscript. Many thanks to Alex Pigot and Richard Pearson for their valuable inputs, which contributed to improve this manuscript. This work was supported by a Royal Society University Research Fellowship to T.N. and a Royal Society Research Grant to T.N., which supports the PhD studentship of A.E.

DATA AVAILABILITY STATEMENT

The data used in this work were collected from freely accessible sources. Each of these sources is cited in the manuscript. The data compiled in this work were made available on figshare (<https://doi.org/10.6084/m9.figshare.10075421>).

ORCID

Adrienne Etard  <https://orcid.org/0000-0002-1700-2972>
Tim Newbold  <https://orcid.org/0000-0001-7361-0051>

REFERENCES

- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5–37. <https://doi.org/10.1016/j.jsp.2009.10.001>
- Barlow, J., França, F., Gardner, T. A., Hicks, C. C., Lennox, G. D., Berenguer, E., ... Graham, N. A. J. (2018). The future of hyperdiverse tropical ecosystems. *Nature, 559*(7715), 517–526. <https://doi.org/10.1038/s41586-018-0301-1>
- Bivand, R., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis, 45*, 150–179. <https://doi.org/10.1111/gean.12008>
- Bivand, R., & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software, 63*(18). <https://doi.org/10.18637/jss.v063.i18>
- Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R* (2nd ed.). Switzerland AG: Springer Nature. <https://doi.org/10.1007/978-1-4614-7618-4>
- Bivand, R. S., & Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *Test, 27*, 716–748. <https://doi.org/10.1007/s11749-018-0599-x>
- Bohm, M., Cook, D., Ma, H., Davidson, A. D., García, A., Tapley, B., Pearce-Kelly, P., & Carr, J. (2016). Hot and bothered: Using trait-based approaches to assess climate change vulnerability in reptiles. *Biological Conservation, 204*, 32–41. <https://doi.org/10.1016/j.biocon.2016.06.002>
- Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., ... Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology and Evolution, 26*, 183–192. <https://doi.org/10.1016/j.tree.2011.01.009>
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 10*, e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bruggeman, J., Heringa, J., & Brandt, B. W. (2009). PhyloPars: Estimation of missing parameter values using phylogeny. *Nucleic Acids Research, 37*(suppl_2), W179–W184. <https://doi.org/10.1093/nar/gkp370>
- Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., ... Forzza, R. C. (2017). Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences USA, 114*, 10695–10700. <https://doi.org/10.1073/pnas.1706756114>
- Chamberlain, S. (2018). rredlist: "IUCN" Red List Client. Retrieved from <https://cran.r-project.org/package=rredlist>
- Chamberlain, S. A., & Szöcs, E. (2013). taxize: Taxonomic search and retrieval in R [version 2; referees: 3 approved]. *F1000Research*. Retrieved from <https://doi.org/10.12688/f1000research.2-191.v2>
- Chaudhary, C., Saeedi, H., & Costello, M. J. (2016). Bimodality of latitudinal gradients in marine species richness. *Trends in Ecology and Evolution, 31*, 670–676. <https://doi.org/10.1016/j.tree.2016.06.001>
- Collen, B., Dulvy, N. K., Gaston, K. J., Gärdenfors, U., Keith, D. A., Punt, A. E., ... Akçakaya, H. R. (2016). Clarifying misconceptions of extinction risk assessment with the IUCN Red List. *Biology Letters, 12*, 20150843. <https://doi.org/10.1098/rsbl.2015.0843>
- Collen, B., Ram, M., Zamin, T., & McRae, L. (2008). The tropical biodiversity data gap: Addressing disparity in global monitoring. *Tropical Conservation Science, 1*, 75–88. <https://doi.org/10.1177/194008290800100202>
- Cooke, R. S. C., Bates, A. E., & Eigenbrod, F. (2019). Global trade-offs of functional redundancy and functional dispersion for birds and mammals. *Global Ecology and Biogeography, 28*, 484–495. <https://doi.org/10.1111/geb.12869>
- Cooke, R. S. C., Eigenbrod, F., & Bates, A. E. (2019). Projected losses of global mammal and bird ecological strategies. *Nature Communications, 10*, 2279. <https://doi.org/10.1038/s41467-019-10284-z>
- Cooper, N., Bielby, J., Thomas, G. H., & Purvis, A. (2008). Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography, 17*, 211–221. <https://doi.org/10.1111/j.1466-8238.2007.00355.x>
- Des Roches, S., Post, D. M., Turley, N. E., Bailey, J. K., Hendry, A. P., Kinnison, M. T., ... Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. *Nature Ecology and Evolution, 2*, 57–64. <https://doi.org/10.1038/s41559-017-0402-5>
- Díaz, S., Kattge, J., Cornelissen, J. H. C., Wright, I. J., Lavorel, S., Dray, S., ... Gorné, L. D. (2016). The global spectrum of plant form and function. *Nature, 529*(7585), 167–171. <https://doi.org/10.1038/nature16489>
- Escudero, A., & Valladares, F. (2016). Trait-based plant ecology: Moving towards a unifying species coexistence theory: Features of the special section. *Oecologia, 180*, 919–922. <https://doi.org/10.1007/s00442-016-3578-5>
- ESRI. (2011). *ArcGIS desktop: Release 10*. Environmental Systems Research Institute.
- Faurby, S., Davis, M., Pedersen, R. Ø., Schowaneck, S. D., Antonelli, A., & Svenning, J.-C. (2018). PHYLACINE 1.2: The phylogenetic atlas of mammal macroecology. *Ecology, 99*, 2626. <https://doi.org/10.1002/ecy.2443>
- Faurby, S., Pedersen, R. Ø., Davis, M., Schowaneck, S. D., Jarvie, S., Antonelli, A., & Svenning, J.-C. (2020). MegaPast2Future/PHYLACINE_1.2: PHYLACINE version 1.2.1. Zenodo. <https://doi.org/10.5281/zenodo.3690867>
- Feldman, A., Sabath, N., Pyron, R. A., Mayrose, I., & Meiri, S. (2016). Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. *Global Ecology and Biogeography, 25*, 187–197. <https://doi.org/10.1111/geb.12398>
- Flynn, D. F. B., Gogol-Prokurat, M., Nogeire, T., Molinari, N., Richers, B. T., Lin, B. B., ... DeClerck, F. (2009). Loss of functional diversity under land use intensification across multiple taxa. *Ecology Letters, 12*, 22–33. <https://doi.org/10.1111/j.1461-0248.2008.01255.x>
- Gainsbury, A. M., Tallowin, O. J. S., & Meiri, S. (2018). An updated global data set for diet preferences in terrestrial mammals: Testing the validity of extrapolation. *Mammal Review, 48*, 160–167. <https://doi.org/10.1111/mam.12119>
- Galán-Acedo, C., Arroyo-Rodríguez, V., Andresen, E., & Arasa-Gisbert, R. (2019). Ecological traits of the world's primates. *Scientific Data, 6*, 55. <https://doi.org/10.1038/s41597-019-0059-9>

- González-Suárez, M., Lucas, P. M., & Revilla, E. (2012). Biases in comparative analyses of extinction risk: Mind the gap. *Journal of Animal Ecology*, 81, 1211–1222. <https://doi.org/10.1111/j.1365-2656.2012.01999.x>
- González-Suárez, M., & Revilla, E. (2013). Variability in life-history and ecological traits is a buffer against extinction in mammals. *Ecology Letters*, 16, 242–251. <https://doi.org/10.1111/ele.12035>
- González-Suárez, M., Zanchetta Ferreira, F., & Grilo, C. (2018). Spatial and species-level predictions of road mortality risk using trait data. *Global Ecology and Biogeography*, 27, 1093–1105. <https://doi.org/10.1111/geb.12769>
- Guralnick, R. P., Zermoglio, P. F., Wieczorek, J., LaFrance, R., Bloom, D., & Russell, L. (2016). The importance of digitized biocollections as a source of trait data and a new VertNet resource. *Database*, 2016, baw158. <https://doi.org/10.1093/database/baw158>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Isaac, N. J. B., Mallet, J., & Mace, G. M. (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution*, 19, 464–469. <https://doi.org/10.1016/j.tree.2004.06.004>
- IUCN. (2020). The IUCN Red List of Threatened Species. Version 2020-2. Retrieved from <https://www.iucnredlist.org>
- Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, 491(7424), 444–448. <https://doi.org/10.1038/nature11631>
- Jetz, W., & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology and Evolution*, 2, 850–858. <https://doi.org/10.1038/s41559-018-0515-5>
- Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., ... Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, 90, 2648. <https://doi.org/10.1890/08-1494.1>
- Jones, O. R., Purvis, A., & Quicke, D. L. J. (2012). Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. *Ecography*, 35, 333–340. <https://doi.org/10.1111/j.1600-0587.2011.06956.x>
- Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönsch, G., ... Wirth, C. (2011). TRY – a global database of plant traits. *Global Change Biology*, 17, 2905–2935. <https://doi.org/10.1111/j.1365-2486.2011.02451.x>
- Kissling, W. D., Dalby, L., Fløjgaard, C., Lenoir, J., Sandel, B., Sandom, C., ... Svenning, J.-C. (2014). Establishing macroecological trait datasets: Digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecology and Evolution*, 4, 2913–2930. <https://doi.org/10.1002/ece3.1136>
- Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., ... Guralnick, R. P. (2018). Towards global data products of essential biodiversity variables on species traits. *Nature Ecology and Evolution*, 2, 1531–1540. <https://doi.org/10.1038/s41559-018-0667-3>
- Lavorel, S., & Garnier, E. (2002). Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail. *Functional Ecology*, 16, 545–556. <https://doi.org/10.1046/j.1365-2435.2002.00664.x>
- Luck, G. W., Lavorel, S., McIntyre, S., & Lumb, K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *Journal of Animal Ecology*, 81, 1065–1076. <https://doi.org/10.1111/j.1365-2656.2012.01974.x>
- Májeková, M., Paal, T., Plowman, N. S., Bryndová, M., Kasari, L., Norberg, A., ... De Bello, F. (2016). Evaluating functional diversity: Missing trait data and the importance of species abundance structure and data transformation. *PLoS ONE*, 11, e0149270. <https://doi.org/10.1371/journal.pone.0149270>
- Martin, L. J., Blossey, B., & Ellis, E. (2012). Mapping where ecologists work: Biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10, 195–201. <https://doi.org/10.1890/110154>
- May, R. M. (2011). Why worry about how many species and their loss? *PLoS Biology*, 9, e1001130. <https://doi.org/10.1371/journal.pbio.1001130>
- Meiri, S. (2018). Traits of lizards of the world: Variation around a successful evolutionary design. *Global Ecology and Biogeography*, 27, 1168–1172. <https://doi.org/10.1111/geb.12773>
- Meiri, S., Feldman, A., & Kratochvil, L. (2015). Squamate hatchling size and the evolutionary causes of negative offspring size allometry. *Journal of Evolutionary Biology*, 28, 438–446. <https://doi.org/10.1111/jeb.12580>
- Menegotto, A., & Rangel, T. F. (2018). Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nature Communications*, 9, 4713. <https://doi.org/10.1038/s41467-018-07217-7>
- Myhrvold, N. P., Baldrige, E., Chan, B., Sivam, D., Freeman, D. L., & Ernest, S. K. M. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology*, 96, 3109–000. <https://doi.org/10.1890/15-0846R.1>
- Nakagawa, S., & Freckleton, R. P. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology and Evolution*, 23, 592–596. <https://doi.org/10.1016/j.tree.2008.06.014>
- Newbold, T., Hudson, L. N., Contu, S., Hill, S. L. L., Beck, J., Liu, Y., ... Purvis, A. (2018). Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide. *PLoS Biology*, 16, e2006841. <https://doi.org/10.1371/journal.pbio.2006841>
- Newbold, T., Scharlemann, J. P. W., Butchart, S. H. M., Sekerciöglu, C. H., Alkemade, R., Booth, H., & Purves, D. W. (2013). Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings of the Royal Society B: Biological Sciences*, 280(1750), 20122131. <https://doi.org/10.1098/rspb.2012.2131>
- Novosolov, M., Raia, P., & Meiri, S. (2013). The island syndrome in lizards. *Global Ecology and Biogeography*, 22, 184–191. <https://doi.org/10.1111/j.1466-8238.2012.00791.x>
- Novosolov, M., Rodda, G. H., North, A. C., Butchart, S. H. M., Tallowin, O. J. S., Gainsbury, A. M., & Meiri, S. (2017). Population density–range size relationship revisited. *Global Ecology and Biogeography*, 26, 1088–1097. <https://doi.org/10.1111/geb.12617>
- Nowakowski, A. J., Thompson, M. E., Donnelly, M. A., & Todd, B. D. (2017). Amphibian sensitivity to habitat modification is associated with population trends and species traits. *Global Ecology and Biogeography*, 26, 700–712. <https://doi.org/10.1111/geb.12571>
- Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphibiO, a global database for amphibian ecological traits. *Scientific Data*, 4, 170123. <https://doi.org/10.1038/sdata.2017.123>
- Pacifici, M., Foden, W. B., Visconti, P., Watson, J. E. M., Butchart, S. H. M., Kovacs, K. M., ... Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, 5, 215–224. <https://doi.org/10.1038/nclimate2448>
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, 401(6756), 877–884. <https://doi.org/10.1038/44766>
- Pakeman, R. J. (2014). Functional trait metrics are sensitive to the completeness of the species' trait data? *Methods in Ecology and Evolution*, 5, 9–15. <https://doi.org/10.1111/2041-210X.12136>
- Pearson, R. G., Stanton, J. C., Shoemaker, K. T., Aiello-Lammens, M. E., Ersts, P. J., ... Akçakaya, H. R. (2014). Life history and spatial traits predict extinction risk due to climate change. *Nature Climate Change*, 4, 217–221. <https://doi.org/10.1038/nclimate2113>
- Penone, C., Davidson, A. D., Shoemaker, K. T., Di Marco, M., Rondinini, C., Brooks, T. M., ... Costa, G. C. (2014). Imputation of missing

- data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution*, 5, 961–970. <https://doi.org/10.1111/2041-210X.12232>.
- Perronne, R., Munoz, F., Borgy, B., Reboud, X., & Gaba, S. (2017). How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. *Perspectives in Plant Ecology, Evolution and Systematics*, 25, 29–44. <https://doi.org/10.1016/j.ppees.2017.01.004>
- Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525–556. <https://doi.org/10.3102/00346543074004525>.
- Purvis, A., Gittleman, J. L., Cowlshaw, G., & Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*, 267(1456), 1947–1952.
- R Core Team. (2018). *R software: Version 3.5.1*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://doi.org/10.1007/978-3-540-74686-7>
- Ramírez-Bautista, A., Thorne, J. H., Schwartz, M. W., & Williams, J. N. (2020). Trait-based climate vulnerability of native rodents in southwestern Mexico. *Ecology and Evolution*, 10, 5864–5876. <https://doi.org/10.1002/ece3.6323>
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <http://dx.doi.org/10.1111/j.2041-210x.2011.00169.x>
- Ripple, W. J., Wolf, C., Newsome, T. M., Hoffmann, M., Wirsing, A. J., & McCauley, D. J. (2017). Extinction risk is most acute for the world's largest and smallest vertebrates. *Proceedings of the National Academy of Sciences USA*, 114, 10678–10683. <https://doi.org/10.1073/pnas.1702078114>
- Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., ... Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution*, 1, 1677–1682. <https://doi.org/10.1038/s41559-017-0332-2>
- Santini, L., Benítez-López, A., Ficetola, G. F., & Huijbregts, M. A. J. (2018). Length-mass allometries in amphibians. *Integrative Zoology*, 13, 36–45. <https://doi.org/10.1111/1749-4877.12268>.
- Sayol, F., Sol, D., & Pigot, A. L. (2020). Brain size and life history interact to predict urban tolerance in birds. *Frontiers in Ecology and Evolution*, 8, 58. <https://doi.org/10.3389/fevo.2020.00058>.
- Scharf, I., Feldman, A., Novosolov, M., Pincheira-Donoso, D., Das, I., Böhm, M., ... Meiri, S. (2015). Late bloomers and baby boomers: Ecological drivers of longevity in squamates and the tuatara. *Global Ecology and Biogeography*, 24, 396–405. <https://doi.org/10.1111/geb.12244>.
- Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., ... Simons, N. K. (2019). Towards an ecological trait-data standard. *Methods in Ecology and Evolution*, 10, 2006–2019. <https://doi.org/10.1111/2041-210X.13288>.
- Schwarz, R., & Meiri, S. (2017). The fast-slow life-history continuum in insular lizards: a comparison between species with invariant and variable clutch sizes. *Journal of Biogeography*, 44, 2808–2815. <https://doi.org/10.1111/jbi.13067>.
- Siefert, A., Violle, C., Chalmandrier, L., Albert, C. H., Taudiere, A., Fajardo, A., ... Wardle, D. A. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology Letters*, 18, 1406–1419. <https://doi.org/10.1111/ele.12508>.
- Slavenko, A., Tallwin, O. J. S., Itescu, Y., Raia, P., & Meiri, S. (2016). Late Quaternary reptile extinctions: Size matters, insularity dominates. *Global Ecology and Biogeography*, 25, 1308–1320. <https://doi.org/10.1111/geb.12491>.
- Sodhi, N. S., Bickford, D., Diesmos, A. C., Lee, T. M., Koh, L. P., Brook, B. W., ... Bradshaw, C. J. A. (2008). Measuring the meltdown: Drivers of global amphibian extinction and decline. *PLoS ONE*, 3, e1636. <https://doi.org/10.1371/journal.pone.0001636>.
- Stark, G., Tamar, K., Itescu, Y., Feldman, A., & Meiri, S. (2018). Cold and isolated ectotherms: Drivers of reptilian longevity. *Biological Journal of the Linnean Society*, 125, 730–740. <https://doi.org/10.1093/biolinean/bly153>
- Stekhoven, D. J. (2016). Nonparametric missing value imputation using random forest. *R package version 1.4*. Retrieved from <https://doi.org/10.1093/bioinformatics/btr597>
- Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—Non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28, 112–118. <https://doi.org/10.1093/bioinformatics/btr597>
- Titley, M. A., Snaddon, J. L., & Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLoS ONE*, 12, e0189577. <https://doi.org/10.1371/journal.pone.0189577>.
- Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., & Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, 204, 23–31. <https://doi.org/10.1016/j.biocon.2016.03.039>
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., ... Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17, 520–525. <https://doi.org/10.1093/bioinformatics/17.6.520>
- United Nations Educational, Scientific and Cultural Organization (UNESCO). (2015). *UNESCO global science report: Towards 2030*. UNESCO Publishing.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Multivariate imputation by chained equations. *Journal of Statistical Software*, 45, 1–67.
- van Rossum, G. (1995). *Python tutorial* (Technical Report CS-R9526). Centrum voor Wiskunde en Informatica (CWI).
- Verberk, W. C. E. P., van Noordwijk, C. G. E., & Hildrew, A. G. (2013). Delivering on a promise: Integrating species traits to transform descriptive community ecology into a predictive science. *Freshwater Science*, 32, 531–547. <https://doi.org/10.1899/12-092.1>.
- Vidan, E., Roll, U., Bauer, A., Grismer, L., Guo, P., Maza, E., ... Meiri, S. (2017). The Eurasian hot nightlife: Environmental forces associated with nocturnality in lizards. *Global Ecology and Biogeography*, 26, 1316–1325. <https://doi.org/10.1111/geb.12643>.
- Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., & Garnier, E. (2007). Let the concept of trait be functional! *Oikos*, 116, 882–892. <https://doi.org/10.1111/j.0030-1299.2007.15559.x>.
- Weiss, K. C. B., & Ray, C. A. (2019). Unifying functional trait approaches to understand the assemblage of ecological communities: Synthesizing taxonomic divides. *Ecography*, 42, 2012–2020. <https://doi.org/10.1111/ecog.04387>.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*, 95, 2027. <https://doi.org/10.1890/13-1917.1>.
- Wong, M. K. L., Guénard, B., & Lewis, O. T. (2018). Trait-based ecology of terrestrial arthropods. *Biological Reviews*, 94, 999–1022. <https://doi.org/10.1111/brv.12488>.
- Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences USA*, 113, 8041–8048. <https://doi.org/10.1073/pnas.1602237113>

BIOSKETCHES

Adrienne Etard is interested in assessing the impacts of global land-use and climate change on the trait structure and the functioning of vertebrate communities. **Sophie Morrill** is a Master of Research graduate who is combining her knowledge of the natural world with her interest in science communication to conduct developmental research for natural history television programmes. **Tim Newbold's** research seeks to understand how human activities, in particular through land-use and climate change, are reshaping terrestrial ecological communities worldwide.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

How to cite this article: Etard A, Morrill S, Newbold T. Global gaps in trait data for terrestrial vertebrates. *Global Ecol. Biogeogr.* 2020;00:1–16. <https://doi.org/10.1111/geb.13184>