University College London

Department of Genetics, Evolution and Environment

The influence of vertebrate species traits
on species' responses to land-use and climate change

Adrienne Etard

Supervision: Dr. Tim Newbold, Dr. Alex Pigot

December 2, 2021

Submitted in part fulfilment of the requirements

for the degree of Doctor of Philosophy in Ecology at University College London

Word count: ∼XXXXXX

CBER

Centre for Biodiversity &
Environment Research

Declaration

The work presented in this thesis is my own. Some parts have been conducted in collaboration with other researchers, and this is indicated where relevant. All else is appropriately referenced.

This document was compiled with LaTeX; the source code and files are available at:

Acknowledgments

Impact statement

Abstract

Key words: vertebrates; traits; land-use change; climate change; functional diversity; vulnerability.

Contents

List of Tables

List of Figures

1 | General Introduction

Thesis outline

The content of Chapter 2 was published in October 2020 in *Global Ecology and Biogeography*, in a research article co-authored by myself, Tim Newbold and Sophie Morrill, who contributed data on XXX (Etard et al., 2020).

2 | Global gaps and biases in terrestrial vertebrate trait data

Keywords

Abstract

Trait data are increasingly 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, I assess whether terrestrial vertebrate trait data are taxonomically, spatially and phylogenetically biased. I compile seven ecological traits and quantify coverage as the proportion of species for which an estimate is available. For a species, I define completeness as the proportion of non-missing values across traits. I assess whether coverage and completeness differ across classes and examine phylogenetic biases in trait data. To investigate spatial biases, I test whether wider-ranging species have more complete trait data than narrow-ranging species. Additionally, I test whether species-rich regions, which are of most concern for conservation, are less well-sampled than species-poor regions. My results show that mammals and birds are well-sampled even in species-rich regions. For reptiles and amphibians (herptiles), only body size presents a high coverage (>80%), as well as habitat related variables (for 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. Hence, trait information is less available on average in some of the most diverse areas and in geographically rarer species, both critical for biodiversity conservation. Gaps in trait data may impede our ability to conduct large scale analyses, while

biases can impact the validity of extrapolations. A short-term solution to the problem is to estimate missing trait data using imputation techniques, while a longer-term and more robust filling of existing gaps requires continued data collection efforts.

2.1 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 and Valladares, 2016; Zamudio et al., 2016), to investigate global patterns of life forms and functions (Díaz et al., 2016), or to assess species' vulnerability to environmental changes (Bohm et al., 2016; Pacifici et al., 2015; Pearson et al., 2014). Because traits influence species' ability to cope with environmental changes (Newbold et al., 2013) and underpin species' contributions to ecosystem processes (Lavorel and 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 297 publications since its release (Activity report, 18/06/2020, `https://www.try-db.org/TryWeb/Home.php`). Such databases hence constitute invaluable research tools and have the potential to greatly advance the field.

Vertebrates are one of the most studied taxa (Titley 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 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 may have to collate trait data from a range of sources (such as in Cooke et al. (2019a,b) or in González-Suárez et al. (2018)), a time-consuming prerequisite which may be a limiting step of the research process. Indeed, collating data from heterogeneously-formatted sources presents many

11

challenges (Schneider et al., 2019), particularly when working across a large number of species. For instance, traits may be measured differently across datasets; units may be inconsistent; and taxonomic resolution and nomenclature may vary.

The lack of a curated, readily available global database for vertebrate ecological traits impedes our ability to conduct cross-taxon comparative studies at global scales. However, efforts to collate data into a single database are limited by the availability of underlying data. Because there exist 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 'Raunkiæran shortfall' by Hortal et al. (2015). Furthermore, incomplete trait data are likely to be biased. Biases in trait data may 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. They notably found that the availability of mammalian trait data was 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 as well as 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 and 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 and Freckleton (2008)). Nevertheless, case deletion may lead to biased parameter estimates and erroneous conclusions when values are not missing at random (González-Suárez et al., 2012). Therefore, it is critical to determine the most appropriate way to deal with data incompleteness. For instance, previous studies using terrestrial vertebrate trait data have employed multiple imputation techniques to fill in the gaps (Cooke et al., 2019a; González-Suárez et al., 2012). Yet, imputation techniques could be sensitive to non-randomness in trait data. Phylogenetic biases (where some clades are under-sampled compared to other clades) could notably impact the performance of several imputation approaches. It is thus vital to characterise the gaps in trait data prior to any analysis. However, there has been no study to date investigating global patterns in the availability of trait data across terrestrial vertebrates.

Here, I aim to assess the global state of trait data in terrestrial vertebrates. I 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 specialisation. I 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). I investigate taxonomic, spatial and phylogenetic biases in trait coverage and completeness.

Given that biodiversity research is globally biased towards birds and mammals (Titley et al., 2017), I hypothesise 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), I hypothesize that species rarity influences completeness, focusing on species geographical range size 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, I test whether species geographical range size explains trait completeness.

It is well established that global research effort is distributed unequally (United Nations Educational Scientific and Cultural and Organization, 2015), with patterns underpinned by various geographical and socioeconomic factors. For instance, countries with higher gross domestic product tend to host a larger number of research institutions (Martin et al., 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 so these data biases are of great concern for biodiversity conservation. Whether species-rich regions are systematically under-sampled for traits compared to species-poor regions is thus important to assess, given the significance of species-rich areas for global conservation. Here, I investigate spatial biases in trait completeness, hypothesizing that species-rich areas are on average less well sampled than species-poor areas.

Finally, I 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.

> **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 thesis, I broaden this definition to include 'ecological' traits (e.g., the 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.

2.2 Methods

I produced class-specific trait datasets that were made available on figshare (DOI: `10.6084/m9.figshare.10075421`). Data compilation and all analyses were conducted with R version 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 2.7; van Rossum, G. (1995)).

2.2.1 Trait data collection

Sources and taxonomic matching

I used freely accessible secondary sources in my compilation (Table 2.1), 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). I developed a procedure aiming at identifying one accepted name for each of the binomial names found across sources. When I could not find an accepted name, I used the original name. Figure 2.1 summarizes the main steps; similar solutions have been used in other large-scale studies (Cooke et al., 2019a).

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 and Szöcs, 2013) R packages. One accepted name was assigned to each synonym. I produced a "Synonym" dataset that I have also made available. I 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, I make two versions of the 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, 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 2, S2.1; Figure S2.1).
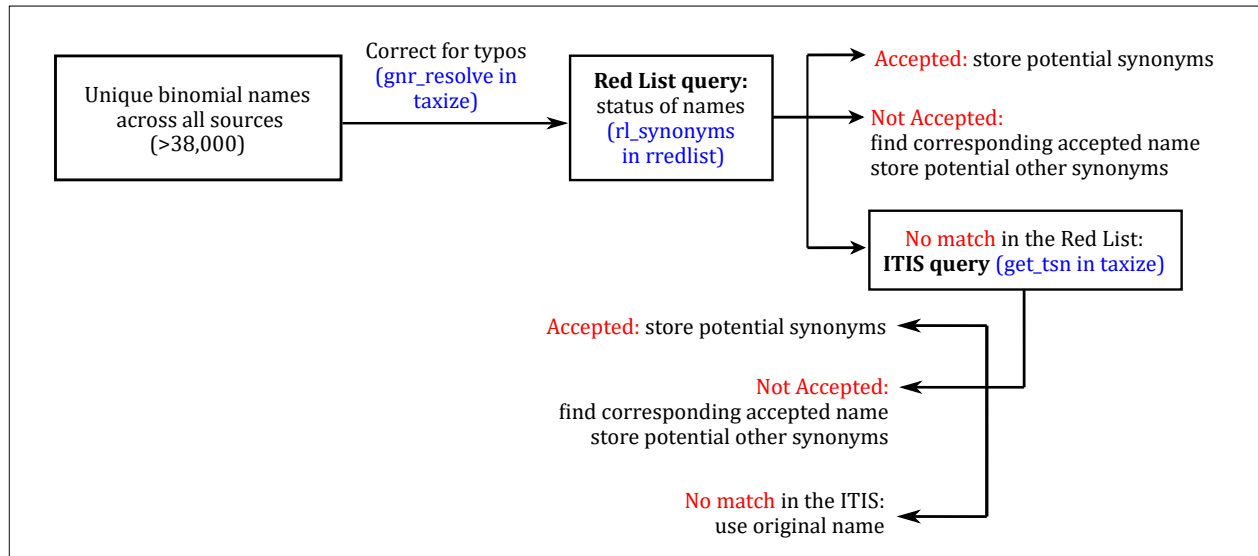
Figure 2.1: Procedure used to identify the accepted names of species. I extracted, where possible, the accepted names of species from either the IUCN Red List or the Integrated Taxonomic Information System (ITIS).

Compilation methods

For continuous traits, I 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 and Revilla, 2012; Siefert et al., 2015), it was not feasible to obtain measures of intraspecific variability from all sources; therefore, estimates were provided

Table 2.1: Data sources for each trait. 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 I derived trophic levels from species diet. [1] `http://datazone.birdlife.org/home`; [2] `https://www.iucnredlist.org/resources/spatial-data-download`; [3] `http://apiv3.iucnredlist.org/api/v3/docs$#$general`.

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[1]		✓				✓				✓	
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) [2]	A, B, M									✓	
IUCN (2020) [3]	All										✓

as a single measure for each species. For some species and some traits, measures were provided separately for females and males. In such cases, I first obtained the mean of these two measures.

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, I 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 the amphibian dataset, Pearson's correlation coefficient between log(Body mass) and log(Body length) was 0.71 (data points shown in the Supporting Information, Figure S2).

- **Longevity**

  I defined longevity as the life span of an individual and maximum longevity as the longest life span reported. I used closely related traits when longevity/maximum longevity was not available or when longevity/maximum longevity had a poorer coverage than a related trait. I selected the age at sexual maturity for amphibians; Pearson's correlation coefficient between log(Age at sexual maturity) and log(Maximum longevity) was 0.55 (Supporting Information, Figure S2). I compiled the generation length for mammals and birds. The correlation between log(Generation length) and log(Longevity) was 0.74 for mammals and 0.70 for birds (Supporting Information, Figure S3). Finally, I 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. I reported measures of central tendencies provided by the sources where applicable; otherwise, I calculated range midpoints (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 2.1). For birds, I 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), I 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**

  I used IUCN habitat data (IUCN, 2020), which describe species habitat preferences and the suitability and importance of each habitat. I 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 I used unweighted sums in this work.

  TO DO *Add relationship between number of habitats and number of strictly natural habitats used.*

- **Use of artificial habitats**

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

Finally, the compiled datasets contain an additional column, 'Note', where I reported species found to be extinct or extinct in the wild (EW). I used species Red List status and information from Meiri (2018) to flag such species. I reported 75 extinct/EW species for mammals, 160 for birds, 34 for amphibians and 53 for reptiles. It is likely that the datasets contain extinct species that I could not flag, because they were not recorded as extinct in the sources I used.

Phylogenies

I used class-specific phylogenetic trees downloaded on 13 April 2020. For mammals, I used 'complete' trees from Faurby et al. (2018, 2020), downloaded from `https://zenodo.org/record/3690867#.Xyc5wyhKhPZ`. For amphibians, birds and squamates, I obtained trees from `https://data.vertlife.org/`. The original sources were as follows: Jetz et al. (2012) for birds; Jetz and 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, I obtained consensus trees using the TreeAnnotator program of the BEAST software (Bouckaert et al., 2019).

Species distributions

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

In addition, for all classes, I 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).

2.2.2 Investigating gaps and biases in trait data

I used trait coverage and completeness to investigate taxonomic, phylogenetic and spatial biases in the trait data. Table 2.2 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 2.2). Also note that I did not filter out species identified as extinct or extinct in the wild, because they represented a small proportion of the datasets (0.48% for amphibians, 1.4% for both birds and mammals, and 0.50% for reptiles) and also because I could not exclude such species systematically, because it is likely that I did not

flag them all.

Table 2.2: Number of species for each analysis. All species represented in the trait datasets were included in (1). All species from the class-specific phylogenetic trees or from the distribution maps that matched with species in the trait datasets were included in (2) and (3).

	(1) Taxonomic biases	(2) Phylogenetic biases	(3) 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

Taxonomic biases

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

Phylogenetic biases

Initially, to assess whether more closely related species were more likely to be similar in trait completeness, I estimated the phylogenetic signal in completeness with Pagel's λ (Pagel, 1999) in each class. I 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). I 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 2.2.

I 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).

Spatial biases

I first investigated whether wider-ranging species were more likely to be better sampled than narrow-ranging species. I tested for a relationship between species range size and trait completeness. I fitted a generalized linear model with a Poisson error distribution (directly using the number of sampled traits, *Ntraits*, rather than the proportion (completeness)). Class was added as a predictor interacting with range size; thus the model was:

20

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

Second, I mapped assemblage-level median completeness. Assemblages were characterized at the pixel level at 50 km^2 resolution. I determined pixel-level composition and richness by stacking species geographical distributions. I then calculated median completeness across species in each pixel. I 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, I 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).

I 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, I fitted such models for herptiles only. I 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 and Piras, 2015; Bivand et al., 2013a; Bivand et al., 2013b)). Given that responses could vary geographically, I 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 Palaearctic. To improve normality, I 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:

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

The value of *W* was estimated using the functions tri2nb and nb2listw of the spdep package (Bivand and Wong, 2018; Bivand et al., 2013b). Fitting the model using all grid cells was computationally intractable; therefore, I randomly sampled cells for this analysis (using 30% of the grid cells in each realm). I selected grid cells where species richness was higher than three to avoid sampling issues. I 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.

2.3 Results

2.3.1 Taxonomic biases in trait information

Trait coverage for mammals and birds was overall high (Figure 2.2(a); 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 2.2(a); 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 2.2(b)). The distribution of trait completeness varied significantly among classes (pairwise Wilcoxon rank sum test: p–value<0.0001 in all cases). Distributions were highly left skewed in mammals and birds (skewness: -2 and -1.6). 84% 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: 0.4), and slightly right skewed for amphibians (skewness: 0.02). 56% of all reptiles and 57% of amphibians fell in the 0-50% completeness range.

2.3.2 Phylogenetic biases in trait completeness

As expected from the distribution of trait completeness in mammals and birds (Figure 2.2), within-family median trait completeness was high across most tips of the phylogenetic trees (Supporting Information Figure S11 and S12; I present the avian and mammalian phylogenies in the Supporting Information because there was little variation in completeness across tips). For birds, λ was 0.71 ($\pm$ 0.0053). For mammals, λ was 0.78 ($\pm$ 0.0035). 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 2.3). In amphibians, groups of families belonging to the order Anura (frogs) showed both the best and worst median completeness (Figure 2.3(a)). 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
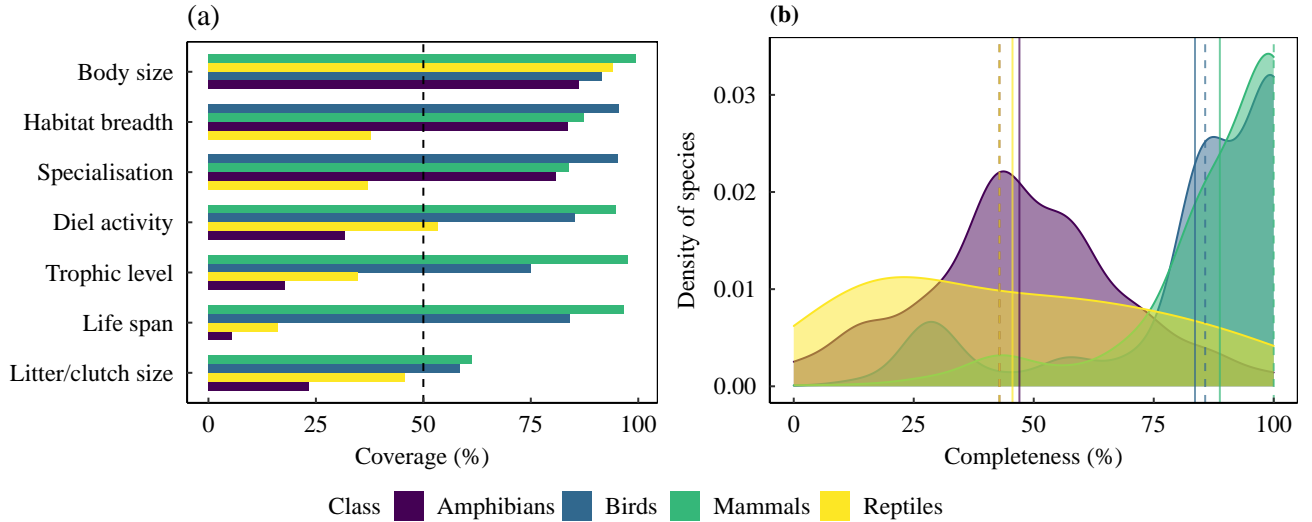
Figure 2.2: Trait coverage and completeness across species. (a) We defined coverage as the proportion of species for which an estimate is available for a given trait. The dashed line represents 50% coverage. **(b)** Trait completeness is the proportion of estimated traits for a species. Here, we show 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 dataset; 9 for birds – 0.077% of species; 7 for mammals – 0.13% 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.

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 0.63 ($\pm$ 0.0039). In reptiles, most snakes were poorly sampled, whereas families in other suborders appeared to be sampled better overall (Figure 2.3(b)). 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 0.69 ($\pm$ 0.0032). The sampling in herptiles was thus also uneven with regard to the phylogeny.

It is important to underline that Figure 2.3 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 the dataset comprised two species of completeness 14% and 86%, respectively. I present similar plots to those in Figure 2.3 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).

23
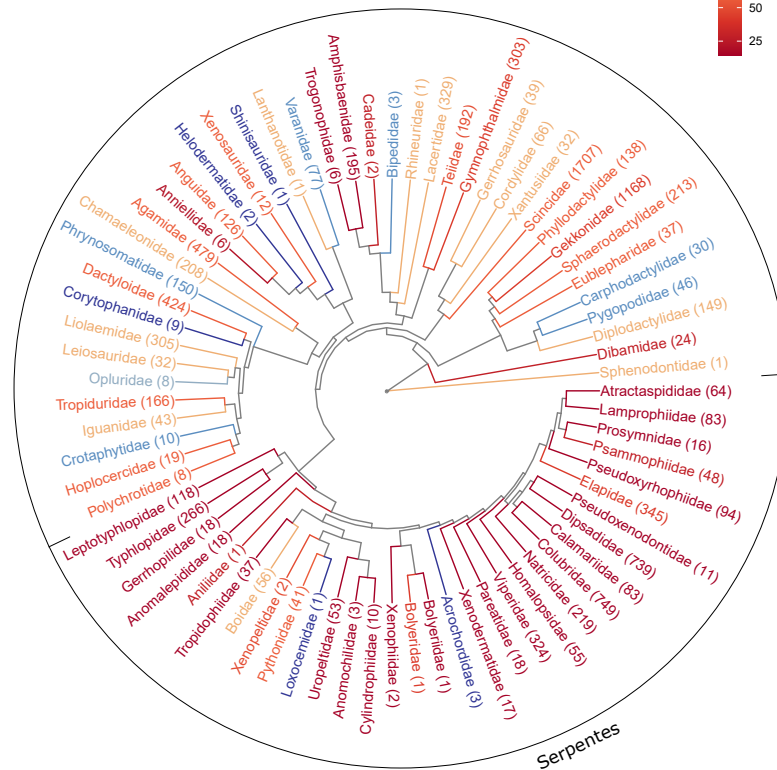
Figure 2.3: Within-family median trait completeness in herptiles. The number next to each family name represents the number of species included in the calculation of the median.

2.3.3 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 2.4; 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).
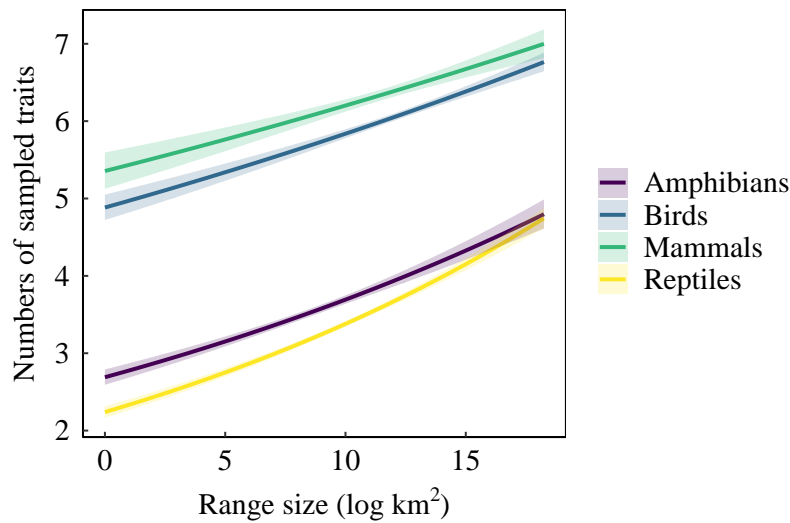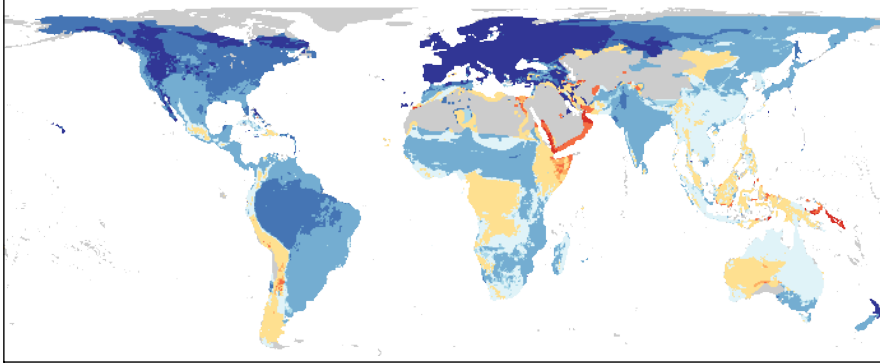

Figure 2.4: Relationship between number of sampled traits and geographical range size. Models were fitted using a Poisson error distribution. Class was added as a predictor interacting with range size. Rates of increase were not significantly different for mammals and birds but differed for reptiles and amphibians, with the steepest rates of increase for reptiles.
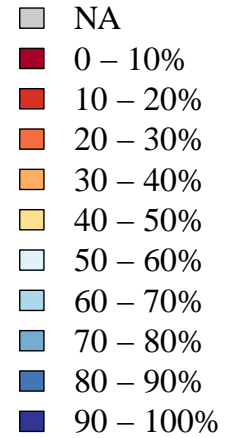
There were marked spatial variations in median trait completeness in herptiles (Figure 2.6). North America and Europe were well sampled for both amphibians and reptiles. Moreover, Southeast Asia and the Congo basin were on average less well sampled. In other regions, contrasting patterns emerged between 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: $\rho = 0.91$, p-value < 0.0001; amphibians: $\rho = 0.92$, p-value < 0.0001). For reptiles, completeness
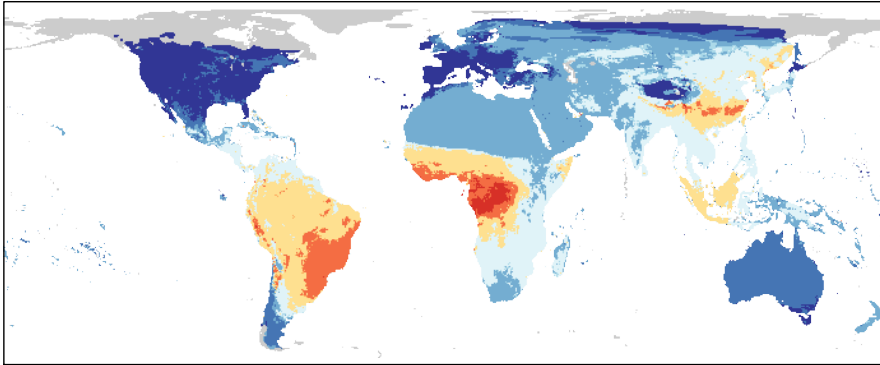
(a) Amphibians

(b) Reptiles

Median completeness:

- ☐ NA
- ■ 0 – 10%
- ■ 10 – 20%
- ■ 20 – 30%
- ■ 30 – 40%
- ■ 40 – 50%
- ■ 50 – 60%
- ■ 60 – 70%
- ■ 70 – 80%
- ■ 80 – 90%
- ■ 90 – 100%

Figure 2.5: Spatial distribution of assemblage-level median trait completeness in herptiles. Similar maps for birds and mammals are shown in the Supporting Information (Figure S7).

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.
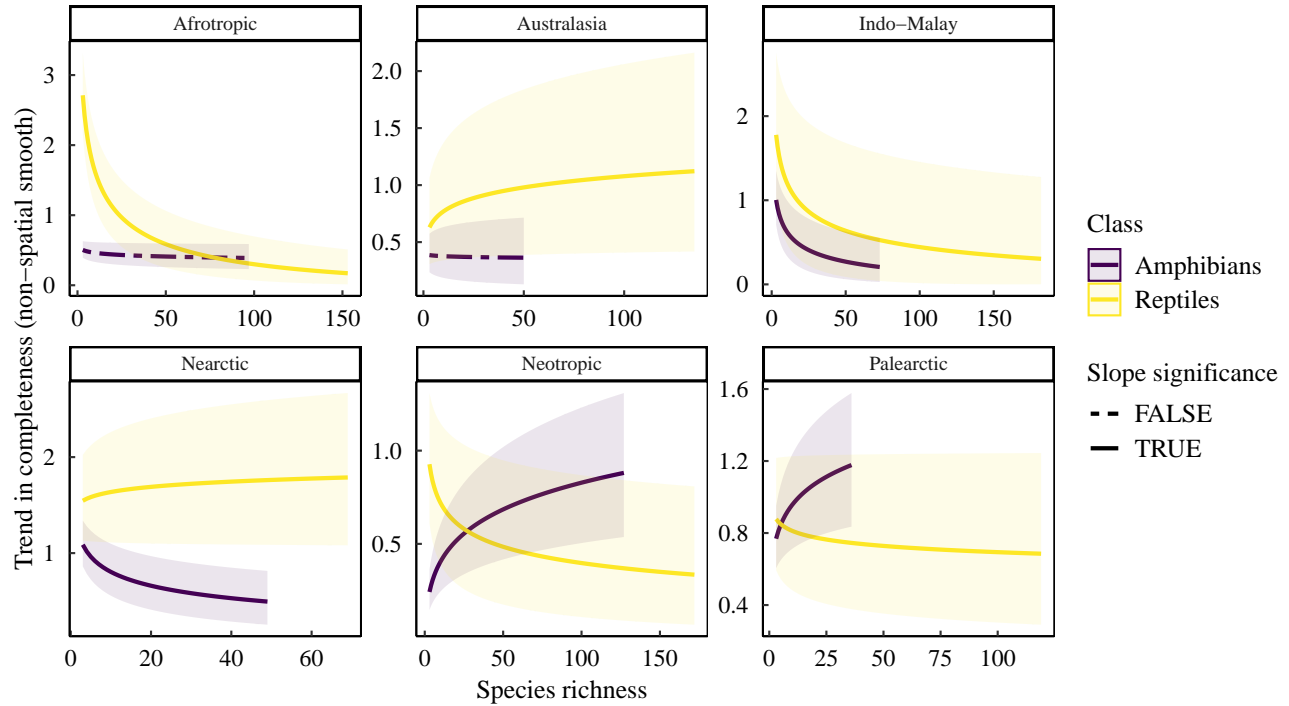
Figure 2.6: Spatial model trends for herptiles. The lines represent in-sample predictions (± SE) for the trend components of the spatial models (trends after accounting for spatial autocorrelation).

2.4 Discussion

The results of this chapter illustrate the taxonomic, spatial and phylogenetic dimensions of the knowledge gaps in trait data, termed the Raunkiæran shortfall by Hortal et al. (2015). To the best of my knowledge, this work constitutes the first comparative assessment of global gaps for terrestrial vertebrate trait data, despite their use in numerous studies. I showed that the trait data present 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 I 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 I highlighted. Nevertheless, biases in the data could have been introduced at later stages, notably with the selection of sources and traits. The global compilation I obtained in this chapter reflects, in part, the

27

interest and focus of the secondary data sources I used. It is possible that the addition of new sources from regional journals or other authorities could diminish spatial biases in the data by increasing coverage for certain areas. Nevertheless, I argue that by focusing on widely used traits, these 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.

I believe that the results presented here 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). I have made two versions of the data compilations available, one in which my 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. I 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.

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 (**Collen2016**; **Purvis2000**; **Ripple2017**) and are more negatively impacted by anthropogenic pressures than wider-ranging species (**Newbold2018a**). 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 (**Barlow2018**). 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 (Pacifici et al., 2015; Ramírez-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 in this chapter, 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 undersampled. 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-latitudinal locations, and argued that these patterns were not underpinned by knowledge gaps in species distributions. Nevertheless, 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 and Freckleton, 2008). For example, retaining complete cases only from our trait datasets would generate trait data disproportionally 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 (**Peugh2004**).

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 et al., 2019a). **Penone2014** used a simulation approach to evaluate the performance of four of these techniques, namely PhyloPars (**Bruggeman2009**), random forest algorithms as implemented in R with missForest (**Stekhoven2016**; **Stekhoven2012**), multivariate imputation by chained equations (MICE; **micepackage**) and k-nearest neighbour (kNN; **Troyanskaya2001**). **Penone2014** 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, **Penone2014** 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 **Penone2014** highlights that there are robust imputation techniques allowing 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 and Freckleton, 2008); indeed, imputation techniques sometimes perform no better than case deletion. More work should be conducted to assess imputation performance in various contexts (see Johnson 2021), and the datasets compiled in this chapter 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. The results of this chapter 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. I believe that both primary research and subsequent efforts to integrate new data and existing databases are required if we are to collectively strive towards the unification of trait databases.

To conclude, this work constitutes, to my knowledge, the first assessment of the global gaps and biases in terrestrial vertebrate trait information. I 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, I believe that primary research, combined with efforts to complete existing datasets, is the only way to fill the current data gaps genuinely and robustly. I hope that the compiled trait dataset and these findings can prove useful for guiding further data collection efforts and for conducting macroecological analyses.

3 | Intense human land uses negatively affect vertebrate functional diversity

4 | Different traits explain sensitivity to land use

5 | Traits & sensitivity to climate change

6 | General Discussion

7 | Conclusion

Bibliography

Baraldi, A. N. and Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*. DOI: `10.1016/j.jsp.2009.10.001`.

Bivand, R. and Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63.18, 1–36.

Bivand, R. and Wong, D. W. S. (2018). Comparing implementations of global and local indicators of spatial association. *TEST*, 27.3, 716–748.

Bivand, R., Hauke, J., and Kossowski, T. (2013a). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, 45.2, 150–179.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V. (2013b). *Applied spatial data analysis with R, Second edition*. Springer, NY.

Bohm, M., Cook, D., Ma, H., Davidson, A. D., Garcïa, A., Tapley, B., Pearce-Kelly, P., and Carr, J. (2016). Hot and bothered: Using trait-based approaches to assess climate change vulnerability in reptiles. *Biological Conservation*. DOI: `10.1016/j.biocon.2016.06.002`.

Bolnick, D. I., Amarasekare, P., Araújo, M. S., Bürger, R., Levine, J. M., Novak, M., Rudolf, V. H., Schreiber, S. J., Urban, M. C., and Vasseur, D. A. (2011). Why intraspecific trait variation matters in community ecology. *Trends in Ecology and Evolution*. DOI: `10.1016/j.tree.2011.01.009`.

Bouckaert, R. et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*. DOI: `10.1371/journal.pcbi.1006650`.

Cardoso, D. et al. (2017). Amazon plant diversity revealed by a taxonomically verified species list. eng. *Proceedings of the National Academy of Sciences of the United States of America*, 114.40, 10695–10700. DOI: `10.1073/pnas.1706756114`.

Chamberlain, S. (2018). *rredlist: 'IUCN' Red List Client*. R package version 0.5.0.

Chamberlain, S. A. and Szöcs, E. (2013). taxize : taxonomic search and retrieval in R [version 2; referees: 3 approved]. *F1000Research*. DOI: `10.12688/f1000research.2-191.v2`.

Collen, B., Ram, M., Zamin, T., and McRae, L. (2008). The Tropical Biodiversity Data Gap: Addressing Disparity in Global Monitoring. *Tropical Conservation Science*, 1.2, 75–88. DOI: `10.1177/194008290800100202`.

Cooke, R. S., Bates, A. E., and Eigenbrod, F. (2019a). Global trade-offs of functional redundancy and functional dispersion for birds and mammals. *Global Ecology and Biogeography*, October 2018, 1–12. DOI: `10.1111/geb.12869`.

Cooke, R. S., Eigenbrod, F., and Bates, A. E. (2019b). Projected losses of global mammal and bird ecological strategies. *Nature Communications*. DOI: `10.1038/s41467-019-10284-z`.

Cooper, N., Bielby, J., Thomas, G. H., and Purvis, A. (2008). Macroecology and extinction risk correlates of frogs. *Global Ecology and Biogeography*. DOI: `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., Schweitzer, J. A., and Palkovacs, E. P. (2018). The ecological importance of intraspecific variation. *Nature Ecology and Evolution*. DOI: `10.1038/s41559-017-0402-5`.

Díaz, S. et al. (2016). The global spectrum of plant form and function. *Nature*. DOI: `10.1038/nature16489`.

Escudero, A. and Valladares, F. (2016). Trait-based plant ecology: moving towards a unifying species coexistence theory: Features of the Special Section. DOI: `10.1007/s00442-016-3578-5`.

ESRI (2011). ArcGIS desktop: Release 10. *Environmental Systems Research Institute*.

Etard, A., Morrill, S., and Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global Ecology and Biogeography*, November 2019, 1–16. DOI: `10.1111/geb.13184`.

Faurby, S., Davis, M., Pedersen, R., Schowanek, S. D., Antonelli1, A., and Svenning, J.-C. (2018). PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology*, 99.11, 2626–2626. DOI: `10.1002/ecy.2443`.

Faurby, S., Pedersen, R., Davis, M., Schowanek, S. D., Jarvie, S., Antonelli, A., and Svenning, J.-C. (2020). PHYLACINE 1.2.1: An update to the Phylogenetic Atlas of Mammal Macroecology. DOI: `10.5281/zenodo.3690867`.

Feldman, A., Sabath, N., Pyron, R. A., Mayrose, I., and Meiri, S. (2016). Body sizes and diversification rates of lizards, snakes, amphisbaenians and the tuatara. *Global Ecology and Biogeography*, 25.2, 187–197. DOI: `10.1111/geb.12398`.

Gainsbury, A. M., Tallowin, O. J., and Meiri, S. (2018). An updated global data set for diet preferences in terrestrial mammals: testing the validity of extrapolation. *Mammal Review*. DOI: `10.1111/mam.12119`.

Galán-Acedo, C., Arroyo-Rodríguez, V., Andresen, E., and Arasa-Gisbert, R. (2019). Ecological traits of the world's primates. *Scientific data*. DOI: `10.1038/s41597-019-0059-9`.

González-Suárez, M. and Revilla, E. (2012). Variability in life-history and ecological traits is a buffer against extinction in mammals. *Ecology Letters*, 16.2, 242–251. DOI: `10.1111/ele.12035`.

González-Suárez, M., Lucas, P. M., and Revilla, E. (2012). Biases in comparative analyses of extinction risk: Mind the gap. *Journal of Animal Ecology*. DOI: `10.1111/j.1365-2656.2012.01999.x`.

González-Suárez, M., Zanchetta Ferreira, F., and Grilo, C. (2018). Spatial and species-level predictions of road mortality risk using trait data. *Global Ecology and Biogeography*, 27.9, 1093–1105. DOI: `10.1111/geb.12769`.

Hortal, J., Bello, F. de, Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., and Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*. DOI: `10.1146/annurev-ecolsys-112414-054400`.

Isaac, N. J., Mallet, J., and Mace, G. M. (2004). Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution*. DOI: `10.1016/j.tree.2004.06.004`.

IUCN (2020). The IUCN Red List of Threatened Species. http://www.iucnredlist.org. Version 2020.1.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., and Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*. DOI: `10.1038/nature11631`.

Jetz, W. and 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*. DOI: `10.1038/s41559-018-0515-5`.

Jones, K. E. et al. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*. DOI: `10.1890/08-1494.1`.

Jones, O. R., Purvis, A., and Quicke, D. L. (2012). Latitudinal gradients in taxonomic overdescription rate affect macroecological inferences using species list data. *Ecography*. DOI: `10.1111/j.1600-0587.2011.06956.x`.

Kattge, J. et al. (2011). TRY - a global database of plant traits. *Global Change Biology*. DOI: `10.1111/j.1365-2486.2011.02451.x`.

Kissling, W. D., Dalby, L., Fløjgaard, C., Lenoir, J., Sandel, B., Sandom, C., Trøjelsgaard, K., and Svenning, J. C. (2014). Establishing macroecological trait datasets: Digitalization, extrapolation, and validation of diet preferences in terrestrial mammals worldwide. *Ecology and Evolution*. DOI: `10.1002/ece3.1136`.

Lavorel, S. and Garnier, E. (2002). *Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the Holy Grail*. DOI: `10.1046/j.1365-2435.2002.00664.x`.

Martin, L. J., Blossey, B., and Ellis, E. (2012). Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Frontiers in Ecology and the Environment*, 10.4, 195–201. DOI: `10.1890/110154`.

May, R. M. (2011). Why worry about how many species and their loss? *PLoS Biology*. DOI: `10.1371/journal.pbio.1001130`.

Meiri, S., Feldman, A., and Kratochvíl, L. (2015). Squamate hatchling size and the evolutionary causes of negative offspring size allometry. *Journal of Evolutionary Biology*, 28.2, 438–446. DOI: `10.1111/jeb.12580`.

Meiri, S. (2018). Traits of lizards of the world: Variation around a successful evolutionary design. *Global Ecology and Biogeography*, 27.10, 1168–1172. DOI: `10.1111/geb.12773`.

Myhrvold, N. P., Baldridge, E., Chan, B., Sivam, D., Freeman, D. L., and Ernest, S. K. M. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology*. DOI: `10.1890/15-0846R.1`.

Nakagawa, S. and Freckleton, R. P. (2008). *Missing inaction: the dangers of ignoring missing data*. DOI: `10.1016/j.tree.2008.06.014`.

Newbold, T., Scharlemann, J. P. W., Butchart, S. H. M., Sekercioğlu, C. H., Alkemade, R., Booth, H., and Purves, D. W. (2013). Ecological traits affect the response of tropical forest bird species to land-use intensity. *Proceedings. Biological sciences / The Royal Society*. DOI: `10.1098/rspb.2012.2131`.

Novosolov, M., Raia, P., and Meiri, S. (2013). The island syndrome in lizards. *Global Ecology and Biogeography*. DOI: `10.1111/j.1466-8238.2012.00791.x`.

Novosolov, M., Rodda, G. H., North, A. C., Butchart, S. H., Tallowin, O. J., Gainsbury, A. M., and Meiri, S. (2017). Population density–range size relationship revisited. *Global Ecology and Biogeography*. DOI: `10.1111/geb.12617`.

Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., and Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*. DOI: `10.1038/sdata.2017.123`.

Pacifici, M. et al. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*. DOI: `10.1038/nclimate2448`.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*. DOI: `10.1038/44766`.

Pearson, R. G., Stanton, J. C., Shoemaker, K. T., Aiello-Lammens, M. E., Ersts, P. J., Horning, N., Fordham, D. A., Raxworthy, C. J., Ryu, H. Y., Mcnees, J., and Akçakaya, H. R. (2014). Life history and spatial traits predict extinction risk due to climate change. *Nature Climate Change*. DOI: `10.1038/nclimate2113`.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.

Roll, U. et al. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. *Nature Ecology and Evolution*. DOI: `10.1038/s41559-017-0332-2`.

Santini, L., Benítez-López, A., Ficetola, G. F., and Huijbregts, M. A. J. (2018). Length–mass allometries in amphibians. *Integrative Zoology*, 13.1, 36–45. DOI: `10.1111/1749-4877.12268`.

Scharf, I., Feldman, A., Novosolov, M., Pincheira-Donoso, D., Das, I., Böhm, M., Uetz, P., Torres-Carvajal, O., Bauer, A., Roll, U., and Meiri, S. (2015). Late bloomers and baby boomers: Ecological drivers of longevity in squamates and the tuatara. *Global Ecology and Biogeography*. DOI: `10.1111/geb.12244`.

Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., Le Provost, G., Manning, P., Ostrowski, A., Penone, C., and Simons, N. K. (2019). *Towards an ecological trait-data standard*. DOI: `10.1111/2041-210X.13288`.

Schwarz, R. and 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*. DOI: `10.1111/jbi.13067`.

Siefert, A. et al. (2015). A global meta-analysis of the relative extent of intraspecific trait variation in plant communities. *Ecology Letters*, 18.12, 1406–1419. DOI: `10.1111/ele.12508`.

Slavenko, A., Tallowin, O. J., Itescu, Y., Raia, P., and Meiri, S. (2016). Late Quaternary reptile extinctions: size matters, insularity dominates. *Global Ecology and Biogeography*. DOI: `10.1111/geb.12491`.

Sodhi, N. S., Bickford, D., Diesmos, A. C., Lee, T. M., Koh, L. P., Brook, B. W., Sekercioglu, C. H., and Bradshaw, C. J. (2008). Measuring the meltdown: Drivers of global amphibian extinction and decline. *PLoS ONE*. DOI: `10.1371/journal.pone.0001636`.

Stark, G., Tamar, K., Itescu, Y., Feldman, A., and Meiri, S. (2018). Cold and isolated ectotherms: drivers of reptilian longevity. *Biological Journal of the Linnean Society*. DOI: `10.1093/biolinnean/bly153/5145102`.

Titley, M. A., Snaddon, J. L., and Turner, E. C. (2017). Scientific research on animal biodiversity is systematically biased towards vertebrates and temperate regions. *PLoS ONE*. DOI: `10.1371/journal.pone.0189577`.

Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., and Pyron, R. A. (2016). Fully-sampled phylogenies of squamates reveal evolutionary patterns in threat status. *Biological Conservation*. DOI: `10.1016/j.biocon.2016.03.039`.

United Nations Educational Scientific and Cultural and Organization (2015). *UNESCO Global Science Report: Towards 2030*. Tech. rep.

van Rossum, G. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica (CWI).

Vidan, E., Roll, U., Bauer, A., Grismer, L., Guo, P., Maza, E., Novosolov, M., Sindaco, R., Wagner, P., Belmaker, J., and Meiri, S. (2017). The Eurasian hot nightlife: Environmental forces associated with nocturnality in lizards. *Global Ecology and Biogeography*. DOI: `10.1111/geb.12643`.

Violle, C., Navas, M. L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., and Garnier, E. (2007). *Let the concept of trait be functional!* DOI: `10.1111/j.0030-1299.2007.15559.x`.

Wilman, H., Belmaker, J., Simpson, J., Rosa, C. de la, Rivadeneira, M. M., and Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*. DOI: `10.1890/13-1917.1`.

Wong, M. K., Guénard, B., and Lewis, O. T. (2018). *Trait-based ecology of terrestrial arthropods*. DOI: `10.1111/brv.12488`.

Zamudio, K. R., Bell, R. C., and Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. *Proceedings of the National Academy of Sciences*, 113.29, 8041–8048. DOI: `10.1073/pnas.1602237113`.

Appendices

Appendix 1: Supporting information for Chapter 1

Appendix 2: Supporting information for Chapter 2

S2.1 Taxonomic corrections

Across the different sources, similar species could appear under different binomial names. This was a problem when matching datasets by species. Moreover, it is possible that within a source, a given species was appearing under two or more different, synonymic names. As such, taxonomic synonymy created duplicated rows for the same species, overall falsely increasing the total number of species and potentially inflating the number of missing trait values. Taxonomic synonymy was hence a major issue. Due to the large number of species across datasets, extensive manual checks could not be applied. The presence of typos in species names had the same effect as synonymy, erroneously duplicating species. We attempted to correct for taxonomy first by correcting for typos, and second by identifying species which were entered under non-accepted names and replacing these with the accepted name. To this end, we developed an automated procedure, complemented with a few manual entries where errors were opportunistically spotted. Such errors in taxonomy were notably spotted when attempting to retrieve trait data for subsets of species, for analyses unrelated to the work conducted here. Taxonomic synonymy was as such checked manually for 91 species (56 birds, 7 mammals and 28 reptiles); in that case, information was extracted from other diverse sources (such as the Reptile Database (`http://www.reptile-database.org/`); Avibase (`https://avibase.bsc-eoc.org/avibase.jsp?lang=EN&pg=home`); AmphibiaWeb (`https://amphibiaweb.org/`); and additional manual checks using the IUCN Red List for mammals). A column in the Synonym dataset mentions where manual checks were applied (in which case the Synonym dataset was manually corrected).

Automated procedure and outputs.

Extracting names from the IUCN Red List and the Integrated Taxonomic Information System (ITIS). The objectives of the automated procedure were to (1) extract species synonymic binomial names from the IUCN Red List or from ITIS (using the rredlist (Chamberlain, 2018) and taxize (Chamberlain and Szöcs,

2013) R packages); and (2) identify the status of each name (accepted or not accepted). We started by generating a list of all names featuring in any of the sources. These 'original' names were corrected for typos (using gnr_resolve function, taxize package). Then, the IUCN Red List was queried and any listed synonyms were stored, as well as the status of each synonym (accepted or not accepted). When species were not found in the IUCN Red List, synonyms were extracted from ITIS. When species were not found in ITIS either, corrected names (original names corrected for typos) were used. Family and order designations were extracted using the same procedure and some entries were retrieved from the Global Biodiversity Information Facility taxonomic backbone when not available in the Red List or in the ITIS (GBIF, `https://www.gbif.org/tools/species-lookup`).

NB: for species entered with the forms *Genus cf.*, *Genus aff.* or *Genus spp.*, the accepted binomial name was left empty.

Output. We generated a 'Synonym' dataset containing records of all names (for 14124, 8743, 6090, and 11678 binomial names for birds, amphibians, mammals and reptiles respectively), their status and their potential synonyms.

Harmonising taxonomy in trait datasets. Taxonomy across datasets was finally homogenised by replacing synonyms with a uniquely identified accepted name. As a consequence, the total number of identified unique species decreased (Figure S2.1 A). The species presenting the highest number of synonyms was the East African mole rat (*Tachyoryctes splendens*), for which we found 12 synonymic names (Figure **??** B).

The automated procedure was not perfect, and taxonomic errors are likely to have persisted in the trait datasets. The Red List and the ITIS were not comprehensive taxonomic sources, and for clades with high degrees of synonymy in names, such as reptiles or amphibians, neither the Red List or the ITIS contained enough information.Taxonomy may be further improved by using class-specific sources in an automated procedure.
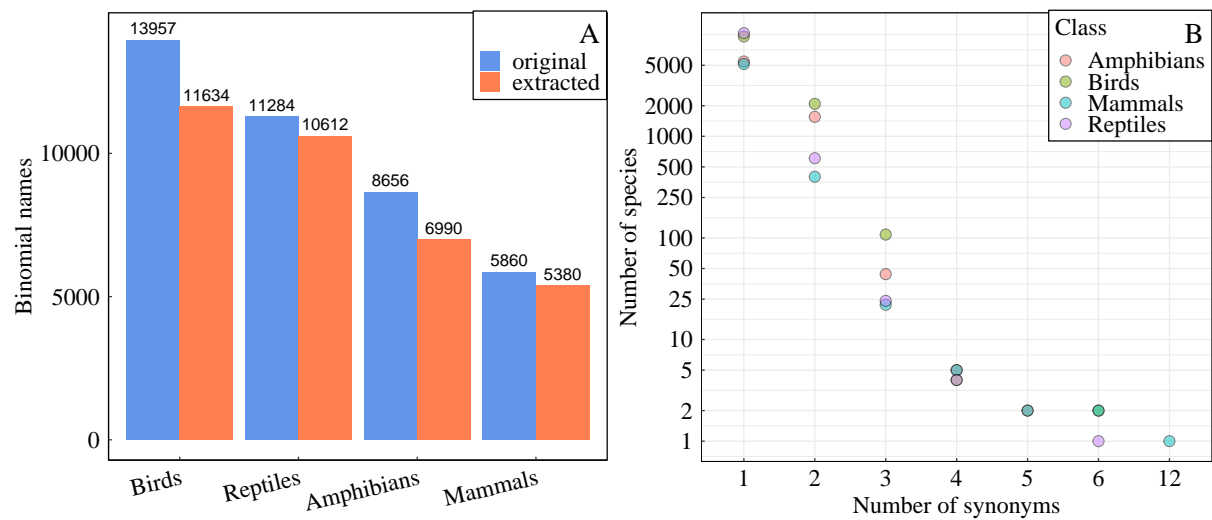
Figure S2.1: Difference in species number before and after taxonomic correction (A) and distribution of number of synonyms across datasets (B). (A) shows the number of species (binomial names) extracted from all sources (blue bars), and the number of uniquely identified accepted names (in red). Replacing non-accepted synonyms by one identified accepted name reduced the number of species in all classes, with the largest reduction for birds. **(B)** shows the distribution of the number of synonymic names. In all four classes, more than 5,000 species were known under one name only. Nevertheless, a large number of species had two identified synonyms (range: 400 species for mammals - 2086 for birds). The most potentially replicated species was the East African mole rat *Tachyoryctes splendens*, for which 12 synonyms were identified.

S2.2 Additional information for trait compilation

Correlations among closely related traits
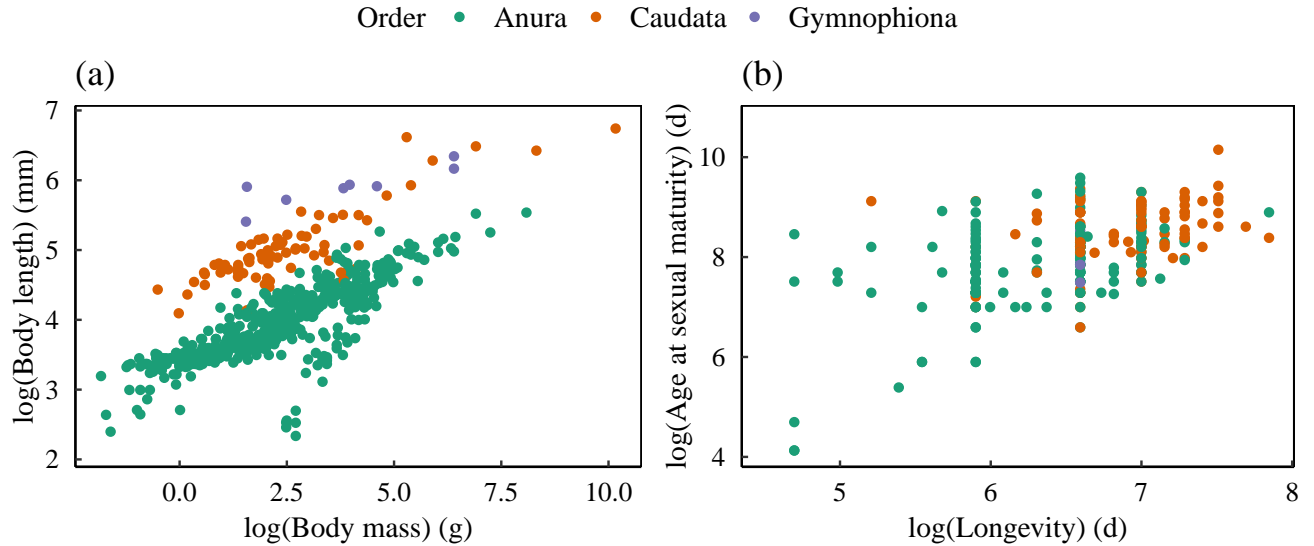
Figure S2.2: (a) Body mass versus body length and (b) Longevity versus age at sexual maturity in amphibians. The Pearson's correlation coefficient was 0.71 in (a) and 0.55 in (b) (order was not included in these coefficients).
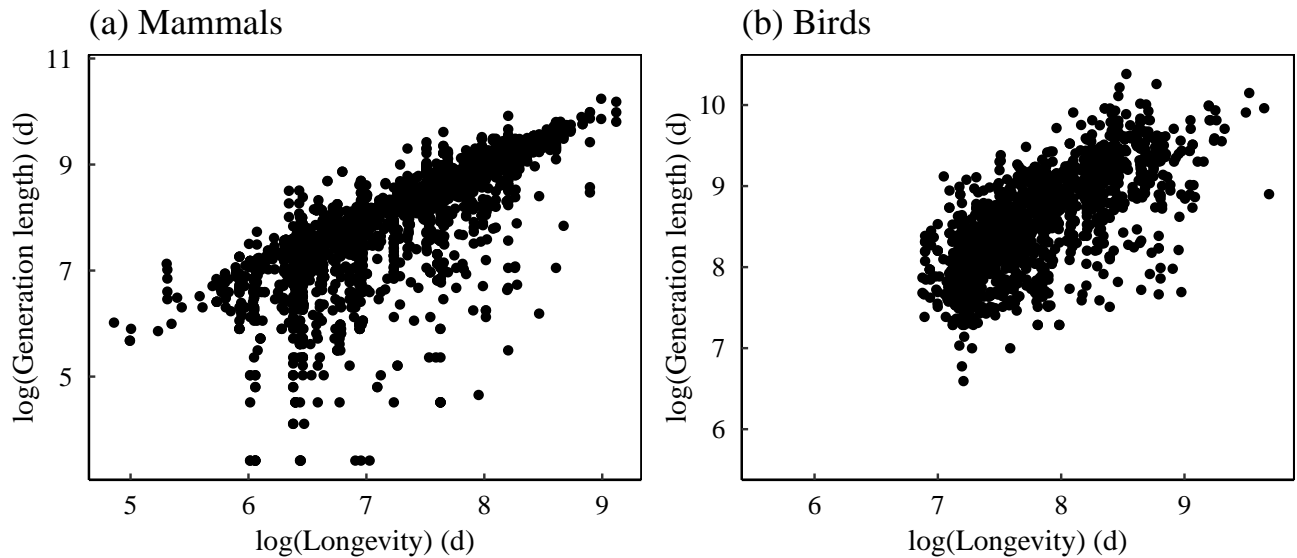
Figure S2.3: Generation length versus longevity data in mammals and birds. The Pearson's correlation coefficient was 0.74 in (a) and 0.70 in (b).

S2.3   Cutting distribution maps by altitudinal limits
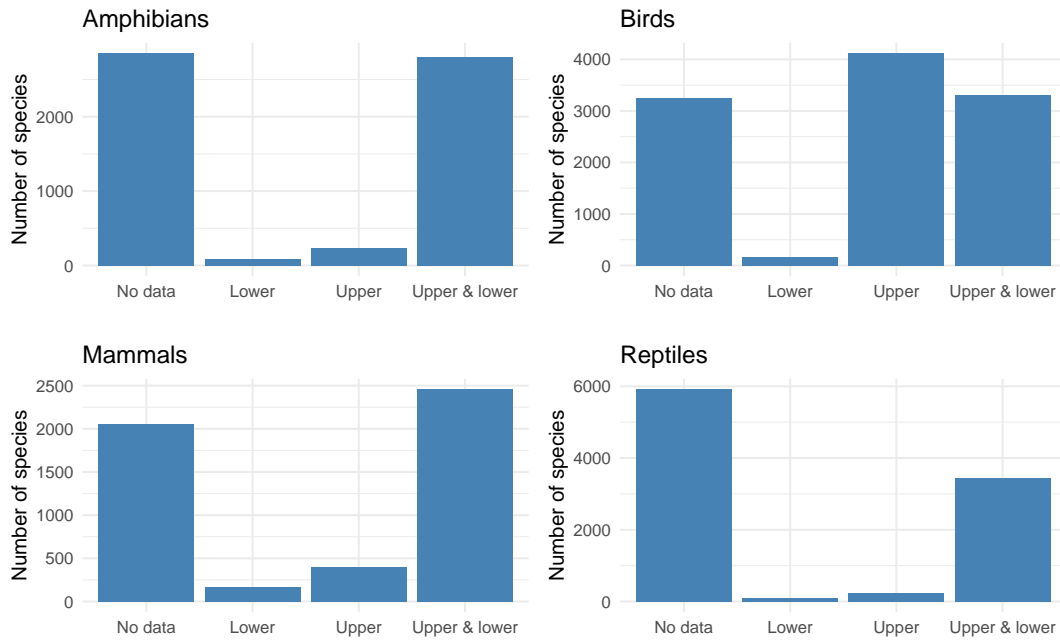
Figure S2.4: Availability of altitudinal limits across species. Upper and lower altitudinal limits were extracted from the IUCN Red List.
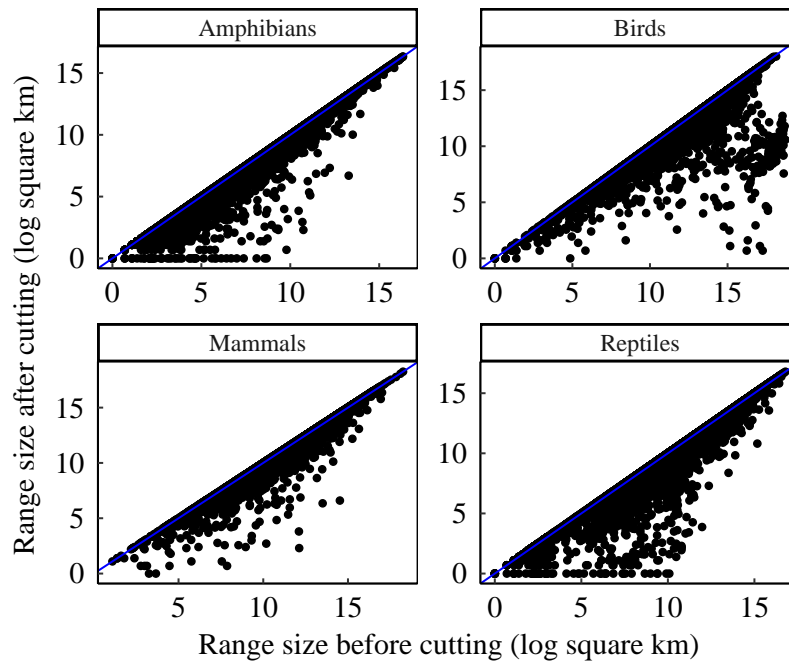
Figure S2.5: Range sizes before VS after cutting by altitudinal limits.

S2.4   Impact of taxonomic corrections of trait coverage
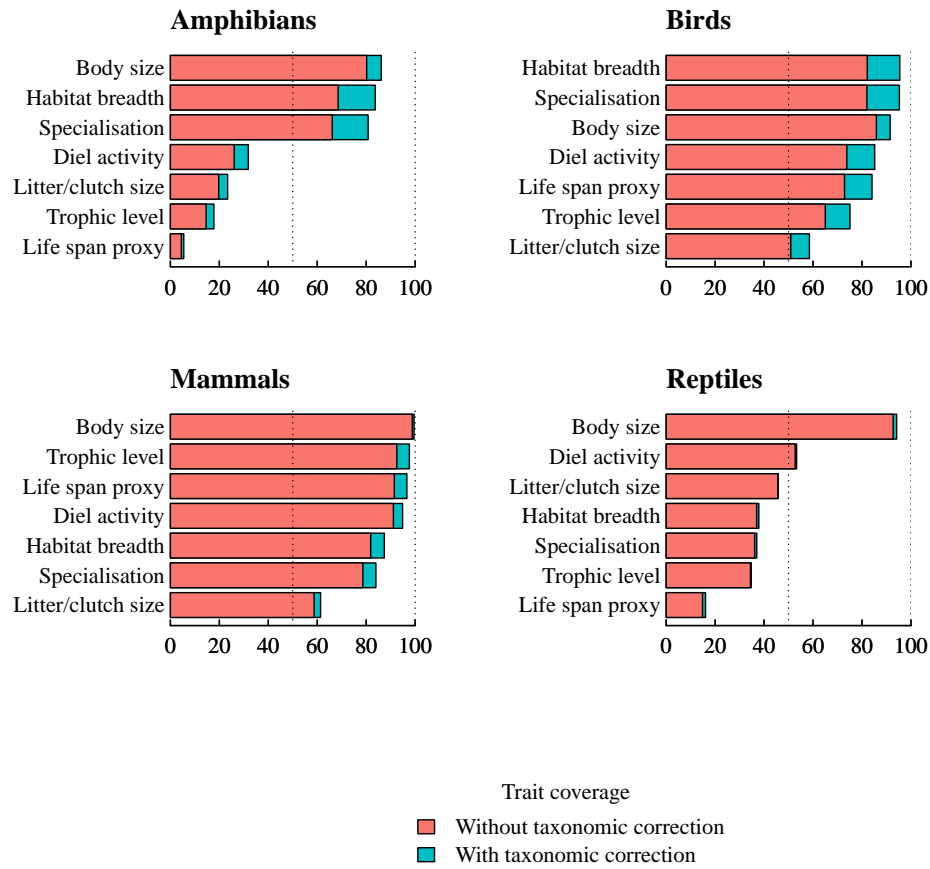
Figure S2.6: Trait coverage when no taxonomic correction is applied before matching sources by binomial names, versus when we apply the described procedure. Identification of synonyms allowed to increase trait coverage in most cases.

S2.5   Assemblage-level median, mean and standard deviation of trait completeness (maps)
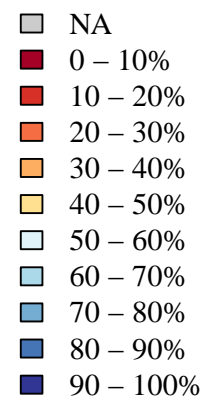
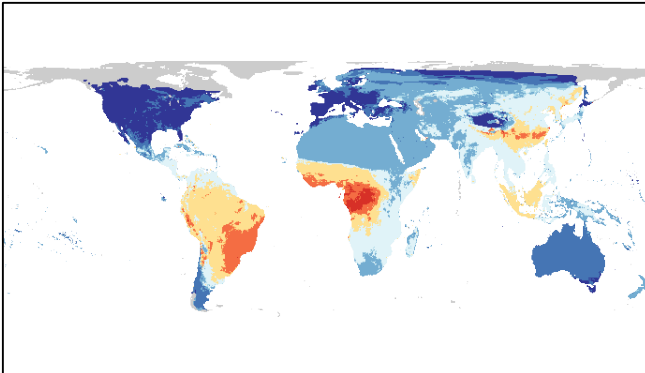**(a) Mammals**
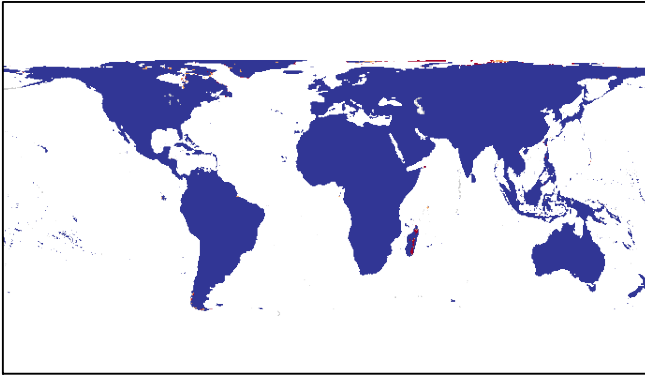
(b) Birds

Median completeness
(birds and mammals):
- NA
- 80 – 90%
- 90 – 91%
- 91 – 92%
- 92 – 93%
- 93 – 94%
- 94 – 95%
- 95 – 96%
- 96 – 97%
- 97 – 98%
- 98 – 99%
- 99 – 100%
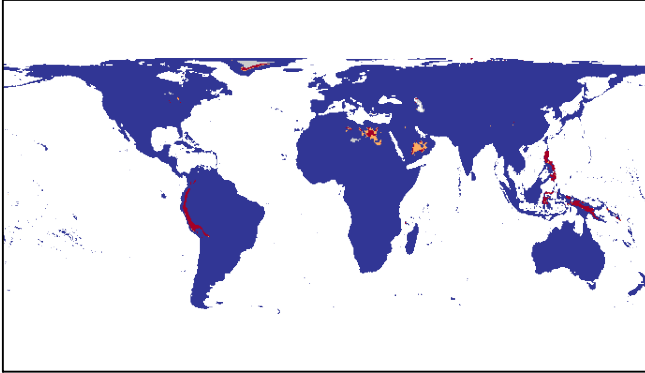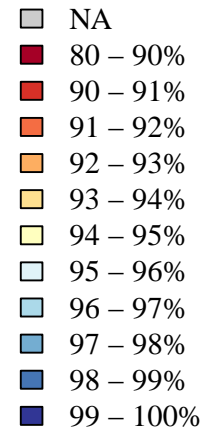
(c) Amphibians

Median completeness
(herptiles):
- NA
- 0 – 10%
- 10 – 20%
- 20 – 30%
- 30 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 70%
- 70 – 80%
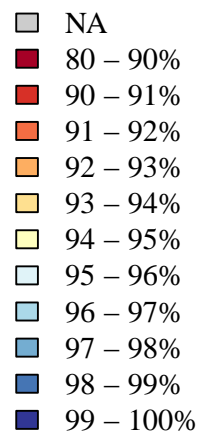- 80 – 90%
- 90 – 100%
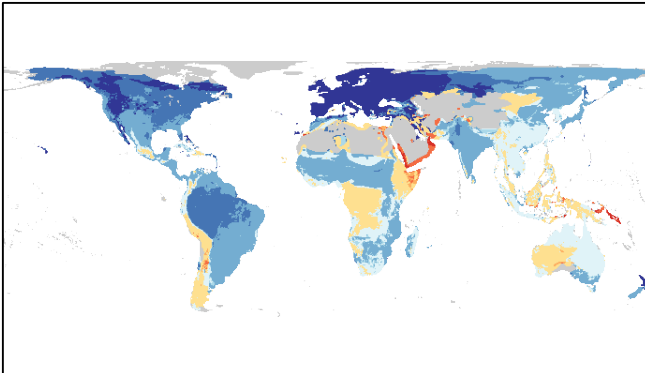
(d) Reptiles

Figure S2.7: Spatial distribution of assemblage-level median trait completeness. Note that the color breaks differ for mammals and birds and for herptiles.
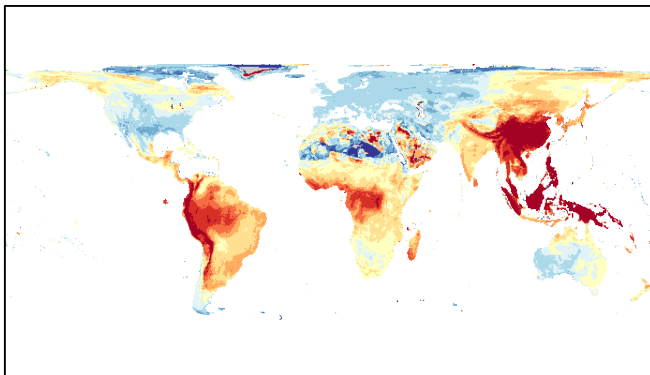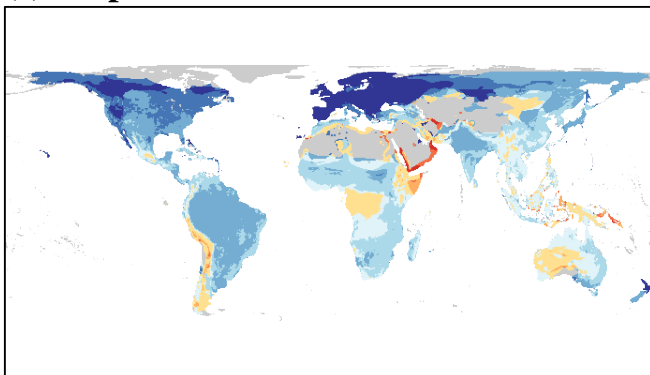
(a) Mammals

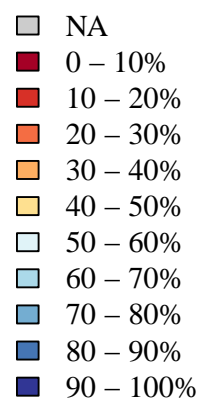**(b) Birds**

Mean completeness
(birds and mammals):

- ☐ NA
- ☐ 80 – 90%
- ☐ 90 – 91%
- ☐ 91 – 92%
- ☐ 92 – 93%
- ☐ 93 – 94%
- ☐ 94 – 95%
- ☐ 95 – 96%
- ☐ 96 – 97%
- ☐ 97 – 98%
- ☐ 98 – 99%
- ☐ 99 – 100%
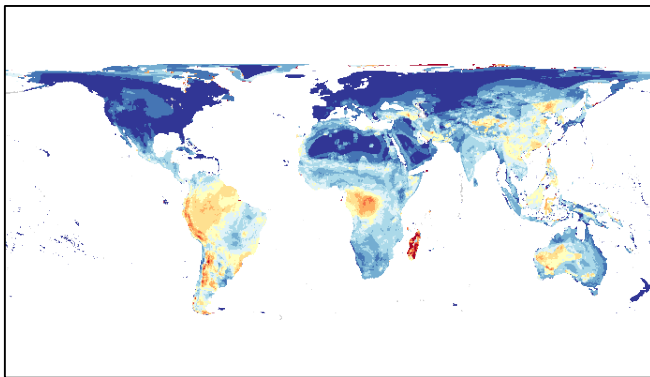
(c) Amphibians

Mean completeness
(herptiles):

- ☐ NA
- ☐ 0 – 10%
- ☐ 10 – 20%
- ☐ 20 – 30%
- ☐ 30 – 40%
- ☐ 40 – 50%
- ☐ 50 – 60%
- ☐ 60 – 70%
- ☐ 70 – 80%
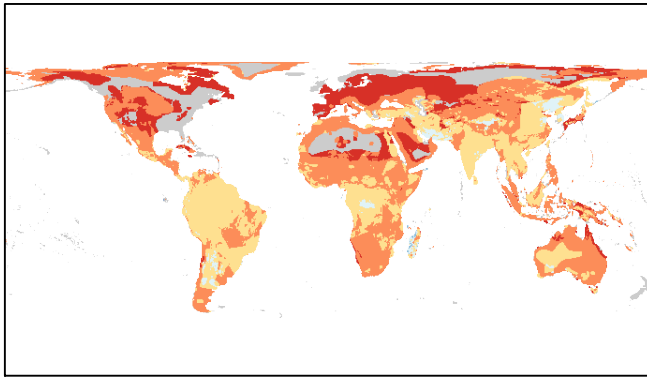- ☐ 80 – 90%
- ☐ 90 – 100%

(d) Reptiles
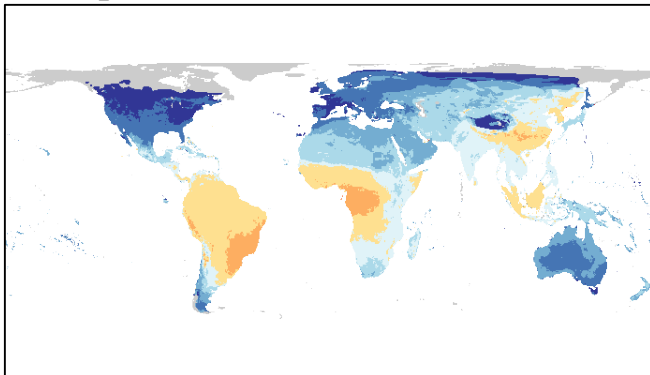
Figure S2.8: Spatial distribution of assemblage-level mean trait completeness. Note that the color breaks differ for mammals and birds and for herptiles.
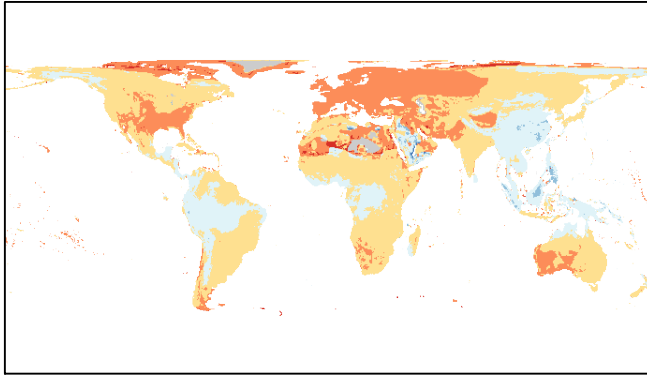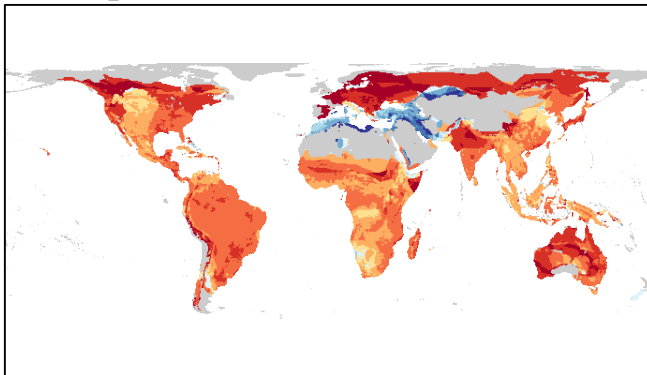
(a) Mammals

(b) Birds

Standard deviation of completeness (birds and mammals):

- ☐ NA
- ■ 0 – 5%
- ■ 5 – 10%
- ■ 10 – 15%
- ☐ 15 – 20%
- ■ 20 – 25%
- ■ 25 – 30%

(c) Amphibians

Standard deviation of completeness (herptiles):

- ☐ NA
- ■ 0– 10%
- ■ 10 – 15%
- ■ 15 – 20%
- ■ 20 – 25%
- ■ 25 – 30%
- ☐ 30 – 35%
- ☐ 35 – 40%
- ■ 40 – 50%
- ■ 50 – 60%
- ■ 60 – 75%

(d) Reptiles
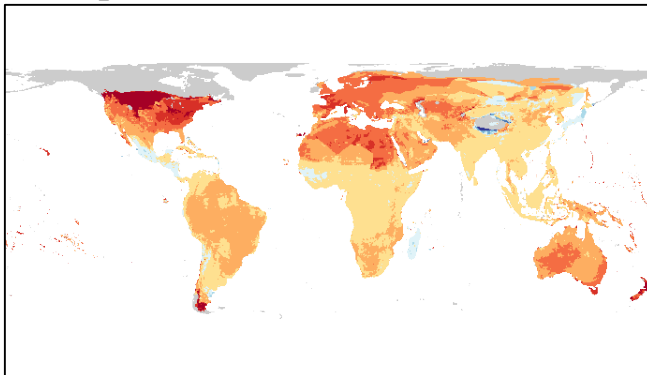
Figure S2.9: Spatial distribution of assemblage-level standard deviation of trait completeness. Note that the color breaks differ for mammals and birds and for herptiles.
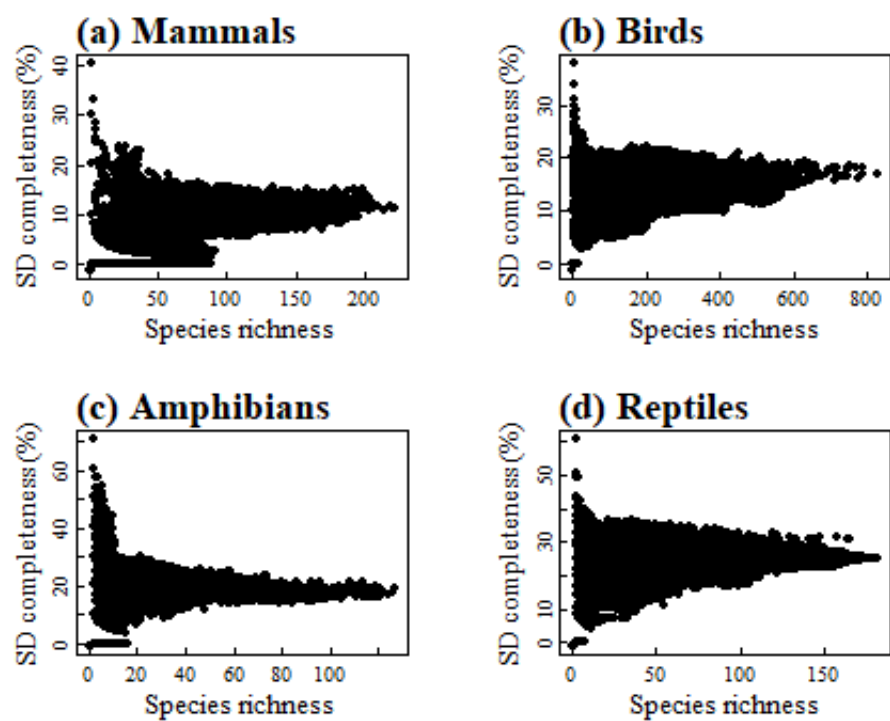
Figure S2.10: Assemblage-level species richness against standard deviation in completeness.

S2.6   Phylogenetic patterns in trait completeness

Figure S2.11: Within-family median trait completeness in mammals.

Figure S2.12: Within-family median trait completeness in birds.

Figure S2.13: Within-family standard deviation in completeness (herptiles).

58
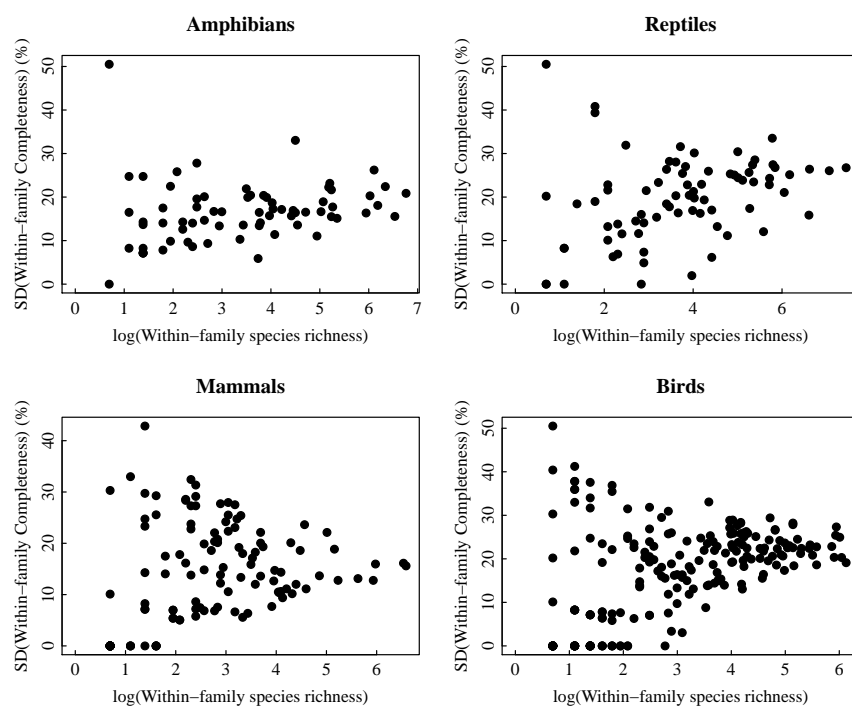
Figure S2.14: Within-family species richness against the within-family standard deviation of completeness.

S2.7 Model coefficients for Range size against Number of sampled traits (Poisson model).

**Table S2.1: Coefficients of the model investigating whether species range size explained the number of sampled traits.** Class was added as an interacting predictor. The reference level for class is Mammals. The model was fitted using Poisson error distribution.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	1.678	0.022	76.809	$< 2e-16$
log Range Size	0.015	0.002	8.086	$6.16e-16$
Class Birds	-0.092	0.028	-3.350	0.000809
Class Amphibians	-0.689	0.029	-24.099	$< 2e-16$
Class Reptiles	-0.872	0.027	-31.856	$< 2e-16$
log Range Size:Class Birds	0.003	0.002	1.415	0.157
log Range Size:Class Amphibians	0.017	0.003	6.427	$1.30e-10$
log Range Size:Class Reptiles	0.026	0.002	11.159	$< 2e-16$

Table S2.2: Coefficients of the model investigating whether species range size explained the number of sampled traits, *using range maps not cut by altitudinal limits*. Class was added as an interacting predictor. The reference level for class is Mammals. The model was fitted using Poisson error distribution.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	1.665	0.023	72.070	$< 2e - 16$
log Range Size	0.015	0.002	8.167	$3.16e - 16$
Class Birds	-0.110	0.029	-3.763	0.0002
Class Amphibians	-0.700	0.030	-23.721	$< 2e - 16$
Class Reptiles	-0.928	0.029	-32.403	$< 2e - 16$
log Range Size:Class Birds	0.004	0.002	1.840	0.066
log Range Size:Class Amphibians	0.018	0.003	6.564	$5.24e - 11$
log Range Size:Class Reptiles	0.031	0.002	12.630	$< 2e - 16$
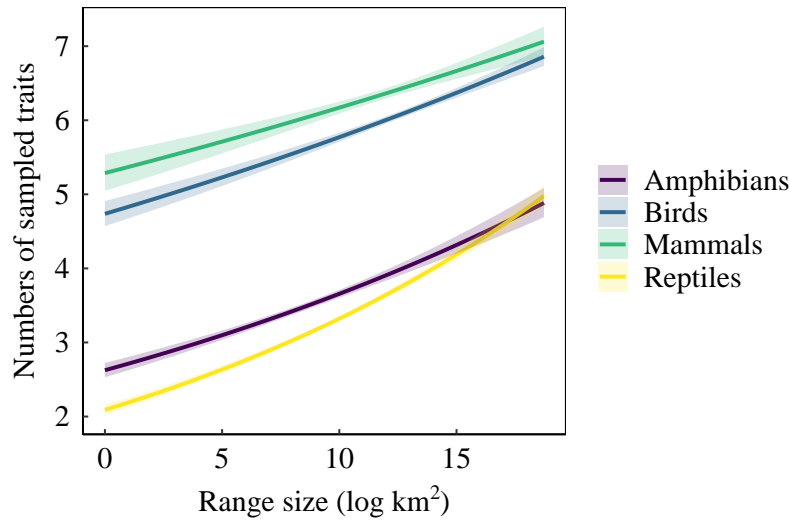
Figure S2.15: Relationship between number of sampled traits and geographical range size *using distribution maps not cut by altitudinal limits*. Models were fitted using a Poisson error distribution. Class was added as a predictor interacting with range size. Rates of increase were not significantly different for mammals and birds, but differed for reptiles and amphibians, with steeper rates of increase for reptiles overall. Cutting range maps by altitudinal limits had little effects on the results (Figure 4 in Main text).

S2.8 Spatial models summaries

Table S2.3: Spatial model summary for amphibians. The spatial model was fitted to explain assemblage-level median completeness with species richness. Biogeographic realm was added as an interacting explanatory variable.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept - Realm: Afrotropic	0.0738	0.0064	11.4908	0
log(Species richness)	-0.0025	0.0017	-1.4261	0.1538
Realm: Australasia	-0.0109	0.0095	-1.1453	0.2521
Realm: Indo-Malay	0.0455	0.0119	3.8294	0.0001
Realm: Nearctic	0.0441	0.0082	5.3905	0.000000
Realm: Neotropic	-0.0377	0.0083	-4.5538	0.00001
Realm: Palearctic	0.0047	0.0067	0.6992	0.4844
log(Species richness):Australasia	0.0018	0.0038	0.4789	0.6320
log(Species richness):Indo-Malay	-0.0147	0.0039	-3.7294	0.0002
log(Species richness):Nearctic	-0.0097	0.0030	-3.2003	0.0014
log(Species richness):Neotropic	0.0144	0.0026	5.6454	0.000000
log(Species richness):Palearctic	0.0109	0.0029	3.7358	0.0002

Table S2.4: Spatial model summary for reptiles. The spatial model was fitted to explain assemblage-level median completeness with species richness. Biogeographic realm was added as an interacting explanatory variable.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept - Realm: Afrotropic	0.2001	0.0144	13.9349	0
log(Species richness)	-0.0316	0.0031	-10.0547	0
Realm: Australasia	-0.1284	0.0189	-6.7851	0
Realm: Indo-Malay	-0.0453	0.0263	-1.7215	0.0852
Realm: Neartic	-0.0788	0.0140	-5.6366	0.000000
Realm: Neotropic	-0.0932	0.0145	-6.4425	0
Realm: Palearctic	-0.1030	0.0131	-7.8787	0
log(Species richness):Australasia	0.0386	0.0046	8.4019	0
log(Species richness):Indo-Malay	0.0124	0.0061	2.0397	0.0414
log(Species richness):Neartic	0.0346	0.0038	9.1601	0
log(Species richness):Neotropic	0.0220	0.0034	6.4231	0
log(Species richness):Palearctic	0.0286	0.0033	8.6153	0

S2.9 Trait coverage and taxonomic matching

Here, we briefly explore the robustness of our work to taxonomic uncertainty by comparing trait coverage obtained with our procedure for taxonomic matching against trait coverage obtained when extracting synonyms from class-specific sources, which could contain more information, notably for herptiles. We aligned taxonomy again using the rangeBuilder R package, which allows the extraction of accepted names from class-specific sources. Overall, our results are robust to the use of different taxonomic backbones; the main conclusions are likely to be unaffected by taxonomic uncertainty.
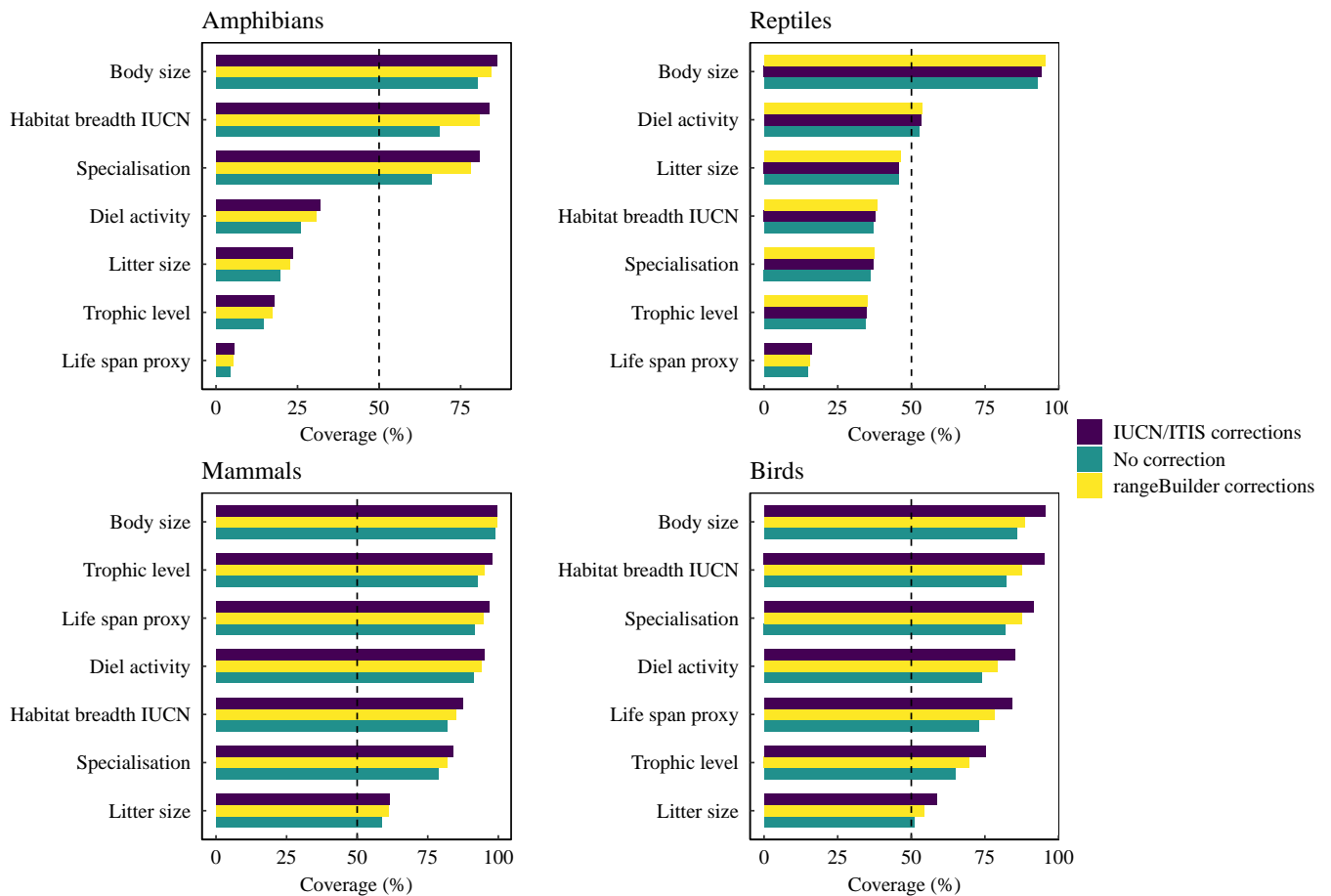
Figure S2.16: Comparison of trait coverage among datasets corrected for taxonomy using the described procedure (purple bars), datasets corrected using the rangeBuilder package (extraction of synonyms from class-specific sources, yellow bars) and datasets where no taxonomic correction was applied when matching sources (green bars).

Appendix 3: Supporting information for Chapter 3

Appendix 4: Supporting information for Chapter 4