



University College London  
Department of Genetics, Evolution and Environment

# **Traits influence responses to land-use and climate change in terrestrial vertebrates**

**Adrienne Etard**

Supervision: Dr. Tim Newbold, Dr. Alex Pigot

April 6, 2022

Submitted in part fulfilment of the requirements  
for the degree of Doctor of Philosophy in Ecology at University College London

Word count: ~XXXXXX

# **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 L<sup>A</sup>T<sub>E</sub>X; the source code and files are available at: <https://github.com/AdrienneEtard>

# Acknowledgments

# **Abstract**

# **Impact statement**

# **Data and code access**

## **Authorship and contributions**

# Contents

<b>List of Tables</b>	<b>10</b>
<b>List of Figures</b>	<b>11</b>
<b>1 General introduction</b>	<b>15</b>
<b>2 Global gaps and biases in trait data for terrestrial vertebrates</b>	<b>17</b>
2.1 Introduction . . . . .	18
2.2 Methods . . . . .	21
2.2.1 Trait data collection . . . . .	21
2.2.2 Investigating gaps and biases in trait data . . . . .	27
2.3 Results . . . . .	29
2.3.1 Taxonomic biases in trait information . . . . .	29
2.3.2 Phylogenetic biases in trait completeness . . . . .	30
2.3.3 Spatial biases in trait completeness . . . . .	33
2.4 Discussion . . . . .	35
<b>3 Intensive human land uses negatively affect vertebrate functional diversity</b>	<b>40</b>
3.1 Introduction . . . . .	40
3.2 Methods . . . . .	42
3.2.1 Vertebrate assemblages . . . . .	42
3.2.2 Functional traits and diversity indices . . . . .	44
3.2.3 Effects of land use and land-use intensity on FRic and FDis (Hypothesis 1) . . . . .	45
3.2.4 Investigating functional under-dispersion (Hypothesis 2) . . . . .	47
3.2.5 Functional loss and functional gain (Hypothesis 3) . . . . .	47
3.3 Results . . . . .	48
3.3.1 Effects of land use on FRic and FDis . . . . .	48

3.3.2	Changes in the probability of occurrence of functional under-dispersion . . . . .	51
3.3.3	Functional loss and gain . . . . .	51
3.4	Discussion . . . . .	52
<b>4</b>	<b>Geographical range area, habitat breadth and specialisation on natural habitats predict land-use responses and climate-change sensitivity more consistently than life-history and dietary traits in terrestrial vertebrates</b>	<b>58</b>
4.1	Introduction . . . . .	58
4.2	Methods . . . . .	58
4.3	Results . . . . .	58
4.4	Discussion . . . . .	58
<b>5</b>	<b>Energetic constraints and trophic levels explain species persistence in disturbed land uses</b>	<b>59</b>
5.1	Introduction . . . . .	60
5.2	Methods . . . . .	60
5.2.1	Vertebrate assemblage composition . . . . .	60
5.2.2	Energy availability by land-use type and land-use intensity . . . . .	61
5.2.3	Resting Metabolic Rates (RMR) & imputations of missing RMR values . . . . .	61
5.2.4	Trophic level and body mass information . . . . .	62
5.2.5	Effects of land use, land-use intensity and trophic level on assemblage-level total RMR (prediction 1; Figure 2(a)) . . . . .	63
5.2.6	Effects of land use, land-use intensity, trophic level and residual RMR on species occurrence probability (Prediction 2; Figure 2(b)) . . . . .	64
5.3	Results . . . . .	65
5.4	Discussion . . . . .	65
<b>6</b>	<b>General discussion</b>	<b>66</b>
<b>7</b>	<b>Conclusion</b>	<b>67</b>
<b>Bibliography</b>		<b>68</b>
<b>Appendices</b>		<b>80</b>
<b>Appendix 1: Supporting information for Chapter 1</b>		<b>81</b>

<b>Appendix 2: Supporting information for Chapter 2</b>	<b>81</b>
S2.1 Taxonomic corrections . . . . .	82
S2.2 Additional information for trait compilation . . . . .	85
S2.3 Cutting distribution maps by altitudinal limits . . . . .	86
S2.4 Impact of taxonomic corrections on trait coverage . . . . .	87
S2.5 Assemblage-level median, mean and standard deviation of trait completeness (maps) . . . . .	88
S2.6 Phylogenetic patterns in trait completeness . . . . .	93
S2.7 Model coefficients for Range size against Number of sampled traits (Poisson model). . . . .	98
S2.8 Spatial models summaries . . . . .	100
S2.9 Trait coverage and taxonomic matching . . . . .	101
<b>Appendix 3: Supporting information for Chapter 3</b>	<b>101</b>
S3.1 Land-use categories in PREDICTS and sample sizes . . . . .	102
S3.2 Trait data & imputation of missing trait values . . . . .	105
S3.2.1 Choice of imputation technique . . . . .	107
S3.2.2 Phylogenetic signal in traits . . . . .	107
S3.2.3 Implementation of missForest imputations . . . . .	108
S3.3 Degree of multicollinearity among traits . . . . .	109
S3.4 Imputation performance . . . . .	110
S3.5 Functional loss and functional gain . . . . .	113
S3.6 Diagnostic plots . . . . .	116
S3.7 Model robustness . . . . .	125
S3.8 Model robustness – time since land-use conversion . . . . .	132
<b>Appendix 4: Supporting information for Chapter 4</b>	<b>133</b>
<b>Appendix 5: Supporting information for Chapter 5</b>	<b>135</b>
<b>Bibliography for the Appendices</b>	<b>137</b>

# List of Tables

<b>Chapter 2</b>	<b>17</b>
2.1    Data sources for each trait . . . . .	23
2.2    Number of species for each analysis . . . . .	27
<b>Appendix 2</b>	<b>81</b>
S2.1    Coefficients of the model investigating whether species range size explained the number of sampled traits . . . . .	98
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 . . . . .	98
S2.3    Spatial model summary for amphibians . . . . .	100
S2.4    Spatial model summary for reptiles . . . . .	100
<b>Appendix 3</b>	<b>101</b>
S3.1    Land-use categories in the PREDICTS database . . . . .	102
S3.2    Phylogenetic signal in continuous and categorical traits . . . . .	108
S3.3    Variance inflation factors across considered (imputed) traits . . . . .	109
S3.4    Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain across all species . . . . .	114
S3.5    Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain within each class . . . . .	115
S3.6    Summary of the model explaining FRic by land use and time since land-use conversion, fitted on the subset of data for which there are information on time since land-use conversion	132
S3.7    Summary of the model explaining FDis by land use and time since land-use conversion, fitted on the subset of data for which I have information on time since land-use conversion	134

# List of Figures

<b>Chapter 2</b>	<b>17</b>
2.1    Procedure used to identify the accepted names of species . . . . .	22
2.2    Number of natural and artificial habitats used by a species against number of strictly natural habitats used by a species . . . . .	26
2.3    Trait coverage and completeness across species . . . . .	30
2.4    Within-family median trait completeness in herptiles . . . . .	32
2.5    Relationship between number of sampled traits and geographical range size . . . . .	33
2.6    Spatial distribution of assemblage-level median trait completeness in herptiles . . . . .	34
2.7    Spatial model trends for herptiles . . . . .	35
<b>Chapter 3</b>	<b>40</b>
3.1    Overview of the study design and functional metrics. . . . .	43
3.2    Effects of land use, land-use intensity and region on FRic (a) and on FDis (b) across all PREDICTS vertebrates. . . . .	49
3.3    Effects of land use, land-use intensity, region and taxonomic class on FRic (a) and FDis (b). .	50
3.4    Effects of land use, land-use intensity and region on the probability of occurrence of functional under-dispersion. . . . .	52
3.5    Effects of land use and land-use intensity on (a) functional loss and (b) functional gain across pairs of sites in temperate areas and tropical areas. . . . .	53
3.6    Effects of land use, land-use intensity, region and taxonomic class on functional loss and functional gain across pairs of sites. . . . .	54
<b>Chapter 4</b>	<b>58</b>

<b>Chapter 5</b>	<b>59</b>
------------------	-----------

<b>Appendix 2</b>	<b>81</b>
S2.1 Difference in species number before and after taxonomic correction (A) and distribution of number of synonyms across datasets (B) . . . . .	84
S2.2 (a) Body mass versus body length and (b) longevity versus age at sexual maturity in amphibians . . . . .	85
S2.3 Generation length versus longevity data in birds . . . . .	85
S2.4 Availability of altitudinal limits across species . . . . .	86
S2.5 Range sizes before versus after cutting by altitudinal limits . . . . .	86
S2.6 Trait coverage when no taxonomic correction is applied before matching sources by binomial names, versus when we applied the described procedure . . . . .	87
S2.7 Spatial distribution of assemblage-level median trait completeness in each class . . . . .	89
S2.8 Spatial distribution of assemblage-level mean trait completeness in each class . . . . .	90
S2.9 Spatial distribution of assemblage-level standard deviation of trait completeness in each class	91
S2.10 Assemblage-level species richness against standard deviation in completeness in each class	92
S2.11 Within-family median trait completeness in mammals . . . . .	94
S2.12 Within-family median trait completeness in birds . . . . .	95
S2.13 Within-family standard deviation in completeness . . . . .	96
S2.14 Within-family species richness against the within-family standard deviation of completeness	97
S2.15 Relationship between number of sampled traits and geographical range size using distribution maps not cut by altitudinal limits . . . . .	99
S2.16 Comparison of trait coverage among datasets corrected for taxonomy in different ways . .	101

<b>Appendix 3</b>	<b>101</b>
-------------------	------------

S3.1 Number of sites in each land use and land-use intensity for which FRic and FDis were calculated, across all species . . . . .	103
S3.2 Number of sites in each land use, land-use intensity and for which FRic and FDis were calculated, within each class . . . . .	104
S3.3 Relationship between habitat breadth and geographical range size across species in each class	105
S3.4 Trait coverage for the vertebrate species sampled in the PREDICTS database . . . . .	106

S3.5	Distribution of trait completeness across the vertebrate species sampled in the PREDICTS database . . . . .	106
S3.6	Distribution of continuous traits considered in the calculation of the functional diversity metrics . . . . .	110
S3.7	Frequency distribution for categorical traits considered in the calculation of the functional diversity metrics . . . . .	111
S3.8	Out-of-bag imputation errors for the continuous traits (a) and categorical traits (b) . . . . .	112
S3.9	Diagnostic plots for Model 1a . . . . .	116
S3.10	Diagnostic plots for Model 1b . . . . .	117
S3.11	Diagnostic plots for Model 2a . . . . .	118
S3.12	Diagnostic plots for Model 2b . . . . .	119
S3.13	Diagnostic plots for Model 3 . . . . .	120
S3.14	Diagnostic plots for Model 4a . . . . .	121
S3.15	Diagnostic plots for Model 4b . . . . .	122
S3.16	Diagnostic plots for Model 5a . . . . .	123
S3.17	Diagnostic plots for Model 5b . . . . .	124
S3.18	Effects of land use, land-use intensity and region on FRic (a) and FDis (b) across vertebrates, for the subset of species with complete trait data . . . . .	125
S3.19	Effects of land use, land-use intensity and region on FRic (a) and FDis (b), for the subset of species with complete trait data, with geographical range size as an additional trait . . . . .	126
S3.20	Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when calculating FRic and FDis with each set of imputed traits . . . . .	127
S3.21	Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when resampling primary vegetation sites twenty independent times . . . . .	128
S3.22	Effects of land use, land-use intensity and taxonomic class on FRic (a) and FDis (c), for the subset of species with complete trait data . . . . .	129
S3.23	Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis obtained with the imputed trait data (black points) or with the complete data subsets (red points) . . . . .	130
S3.24	Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis, obtained when calculating FRic and FDis with each set of imputed traits . . . . .	131

S3.25 Effects of land use on FRic for the model that includes time since land-use conversion (blue points) versus the model that does not take time since land-use conversion into account (red points) . . . . .	133
---	-----

# 1 | General introduction

## Thesis outline

Chapter 2: this chapter was conducted in collaboration with In this chapter, I collected trait data for terrestrial vertebrates, targeting seven traits that are commonly used in the field (e.g., body mass, trophic level, etc). I assessed the availability of the trait data across vertebrate classes and I investigated whether the data present taxonomic, phylogenetic and spatial biases. This (the content of this chapter has been published: <https://onlinelibrary.wiley.com/doi/full/10.1111/geb.13184>).

Chapter 2: Here, I combined the trait data collected in the previous chapter with the PREDICTS database (a database containing records of species occurrence in different land uses), in order to investigate the effects of land-use disturbance on local functional diversity indices (namely functional richness and functional dispersion) with a “space-for-time” approach. The content of this chapter has been accepted for publication.

Chapter 3: In this chapter, I investigated which traits shape species responses to land-use and climate change. I used the trait data collected in the first chapter. For the land-use aspects, I used the PREDICTS database to investigate whether traits influence species persistence in disturbed land uses. For the climate change aspects, I quantified sensitivity to climate change across vertebrates from the properties of species climatic niche space, and I then investigated whether traits influenced sensitivity to climate change. This chapter is in progress.

Chapter 4: In this chapter, I collected resting metabolic rates for vertebrate species. I combined these metabolic rates with the PREDICTS database to investigate the impacts of land-use disturbance on assemblage-level energetic requirements. I also investigated whether metabolic rates influenced species responses to land-use disturbance, asking whether resting metabolic rates can be considered as response traits. The content of this chapter is going to be submitted as a research article in the coming days.

**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 trait data for terrestrial vertebrates

### Keywords

Terrestrial vertebrates; traits; coverage; completeness; taxonomic biases; spatial biases; phylogenetic biases.

### 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

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 ‘Raunkiær 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 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), 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 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. I 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 may be 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**

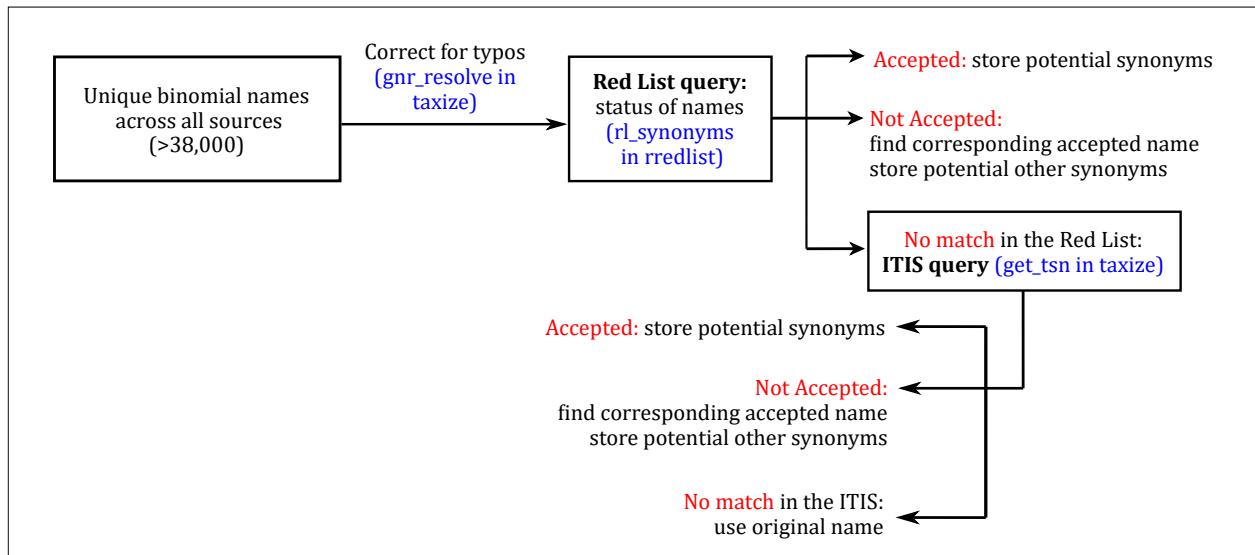
#### **Sources and taxonomic matching**

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 synonymous 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 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). Figure reproduced from Etard et al. (2020).

## 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. <sup>1</sup> <http://datazone.birdlife.org/home>; <sup>2</sup> <https://www.iucnredlist.org/resources/spatial-data-download>; <sup>3</sup> <http://apiv3.iucnredlist.org/api/v3/docs##general>. Table reproduced from Etard et al. (2020).

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 <sup>1</sup>		✓				✓					✓
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) <sup>2</sup>	A, B, M										✓
IUCN (2020) <sup>3</sup>	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 Appendix 2, S2.2, Figure S2.2).

- **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 (Appendix 2, S2.2, Figure S2.2). 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 (data points shown in Appendix 2, S2.2, Figure S2.3). 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, I 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.

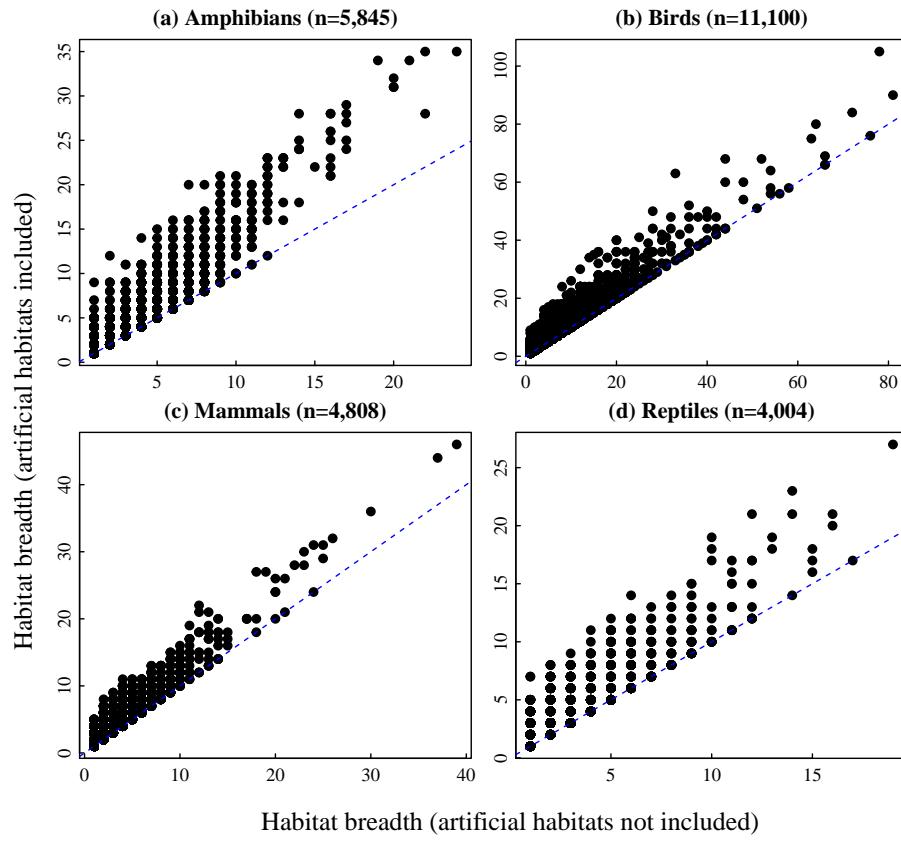
- **Habitat breadth**

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 total number of habitats, determined by including those that qualify as artificial, correlates positively with the number of natural habitats used (Figure 2.2).

- **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.



**Figure 2.2: Number of natural and artificial habitats used by a species against number of strictly natural habitats used by a species.** Pearson's correlation coefficients show a high positive correlation between these two metrics of habitat breadth in all terrestrial vertebrate classes: 0.92 for amphibians (a), 0.95 for birds (b), 0.94 for mammals (c), and 0.90 for reptiles (d).

## 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, Appendix 2, S2.3, Figure S2.4). Decreases in range sizes were observed after cutting distribution maps by the known elevational limits (Appendix 2, S2.3, Figure S2.5).

## 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 2.3). 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). *Table reproduced from Etard et al. (2020).*

	(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 (Appendix 2, S2.4, Figure S2.6).

## 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  $\lambda$  (Pagel, 1999) in each class. I used a bootstrapping approach, calculating  $\lambda$  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  $\lambda$ . Sample sizes for computing  $\lambda$  (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, ' $N_{traits}$ ', rather than the proportion (completeness)). Class was added as a predictor interacting with range size; thus the model was:

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

Second, I mapped assemblage-level median completeness. Assemblages were characterized at the pixel level at 50 km<sup>2</sup> 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 Appendix 2 (S2.5, Figure S2.7; 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 Appendix 2 (S2.5; Figures S2.8 and S2.9 show corresponding maps; Figure S2.10 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 et al., 2013a; Bivand and Piras, 2015; 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,  $\rho$ , associated with an n-by-n matrix of spatial weights,  $W$ . The final model was:

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

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

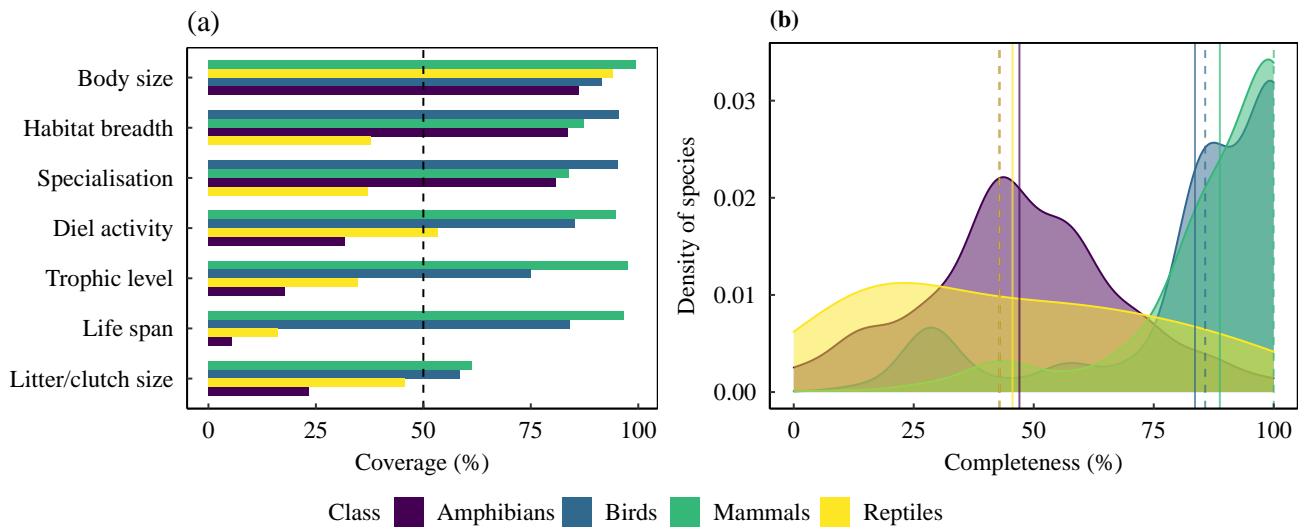
### 2.3.1 Taxonomic biases in trait information

Trait coverage for mammals and birds was overall high (Figure 2.3(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 2.3(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 2.3(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.



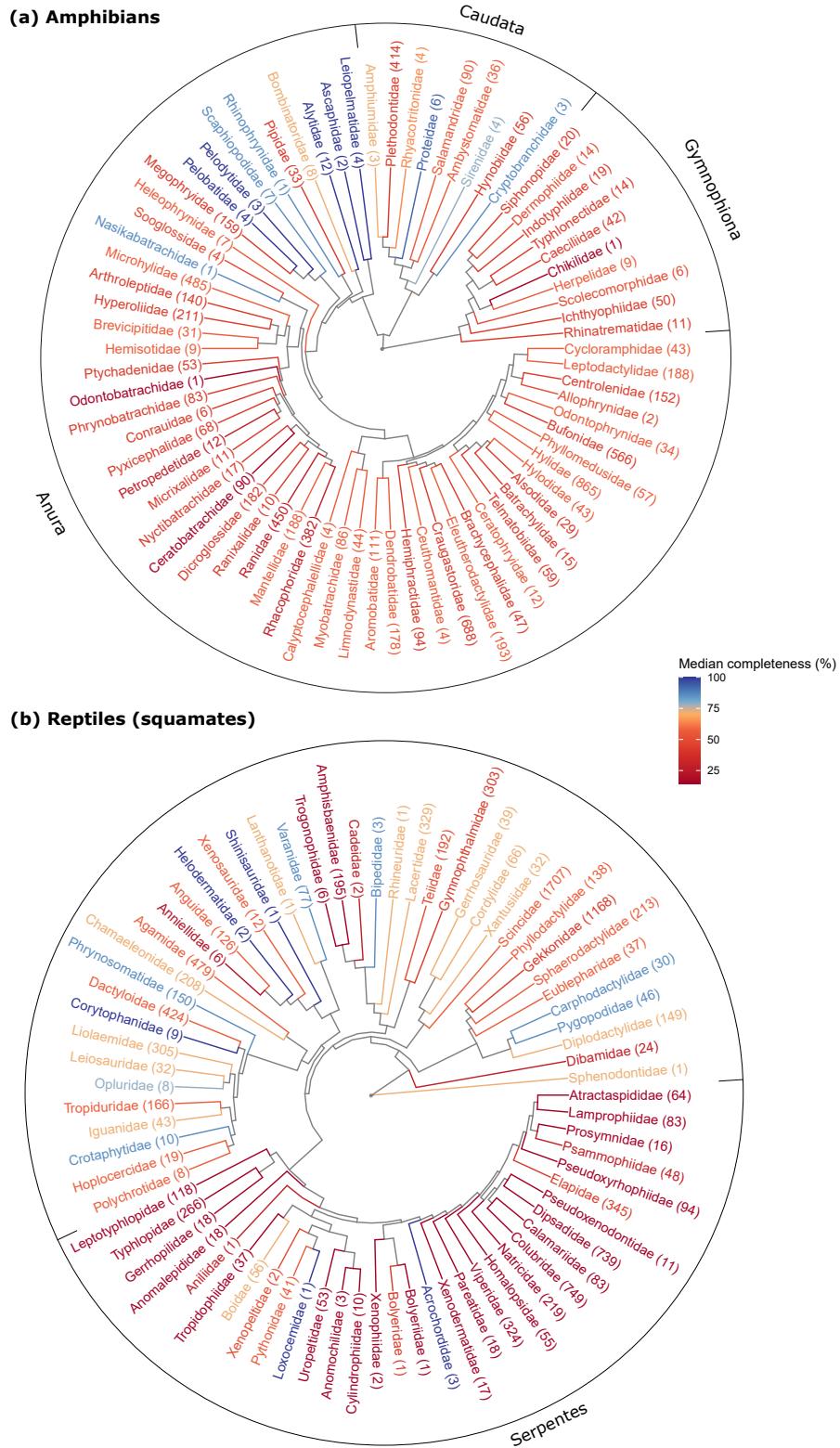
**Figure 2.3: Trait coverage and completeness across species.** (a) I defined coverage as the proportion of species for which an estimate is available for a given trait. The dashed line represents 50% coverage. (b) Trait completeness is the proportion of estimated traits for a species. Here, I show the 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 the 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 I did not select in the analyses, but no known value for the traits I 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 I considered. *Figure reproduced from Etard et al. (2020).*

### 2.3.2 Phylogenetic biases in trait completeness

As expected from the distribution of trait completeness in mammals and birds (Figure 2.3), within-family median trait completeness was high across most tips of the phylogenetic trees (Appendix 2, Figures S2.11 and S2.12; I present the avian and mammalian phylogenies in the Appendix because there was little variation in completeness across tips). For birds,  $\lambda$  was 0.71 ( $\pm 0.0053$ ). For mammals,  $\lambda$  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 2.4). In amphibians, groups of families belonging to the order Anura (frogs) showed both the best and worst median completeness (Figure 2.4(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 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,  $\lambda$  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 2.4(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,  $\lambda$  was  $0.69 (\pm 0.0032)$ . The sampling in herptiles was thus also uneven with regard to the phylogeny.

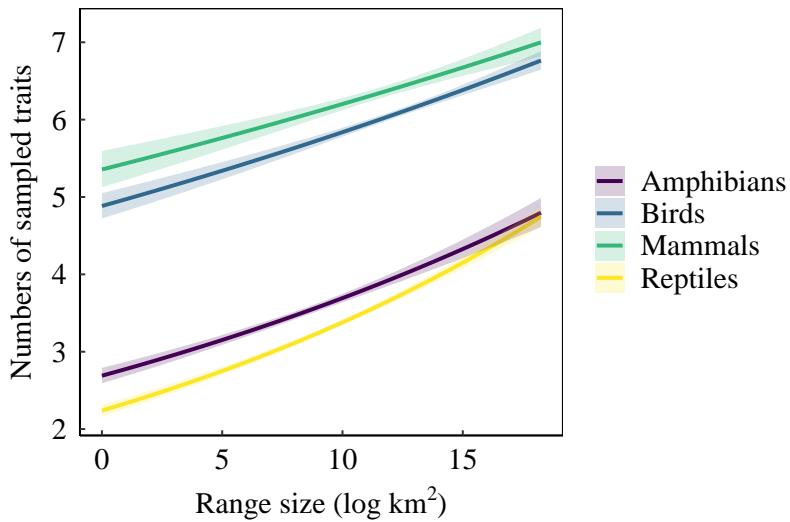
It is important to underline that Figure 2.4 shows within-family median completeness, masking the considerable variation in species richness across families, hence masking potential important variation in completeness across species within families. For example, in the amphibian family Allophrynididae (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 2.4 showing the within-family standard deviation in completeness in Appendix 2 (Figure S2.13). Within-family standard deviation tended to increase with within-family species richness (Appendix 2, Figure S2.14).



**Figure 2.4: 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. *Figure reproduced from Etard et al. (2020).*

### 2.3.3 Spatial biases in trait completeness

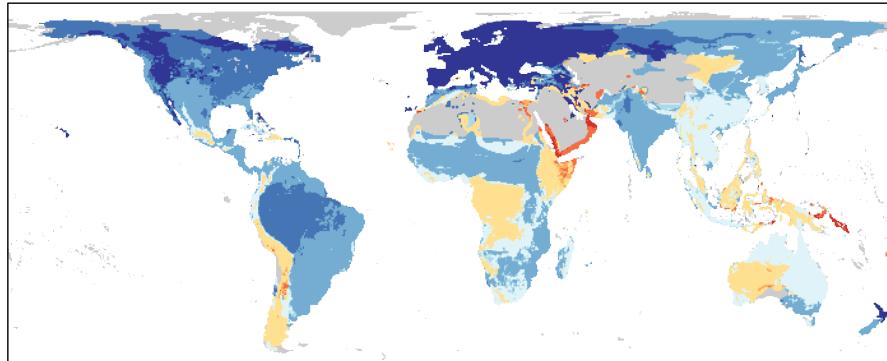
Range size was significantly correlated with the number of sampled traits. Larger range sizes were associated with a higher number of sampled traits (i.e., with higher completeness; Figure 2.5; Appendix 2, S2.7, Table S2.1). Similar results were obtained when using distribution maps not cut by elevational limits (Appendix 2, S2.7, Table S2.2; Figure S2.15). 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; Appendix 2, Table S2.1).



**Figure 2.5: 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 in number of sampled traits with range size were not significantly different for mammals and birds but differed for reptiles and amphibians, with the steepest rates of increase for reptiles. *Figure reproduced from Etard et al. (2020).*

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 Appendix 2 (S2.5, Figures S2.8 and S2.9). There was a trend for increasing standard deviation with increasing species richness, with a larger spread in standard deviation at lower species richness (Appendix 2, section S2.5, Figure S2.10).

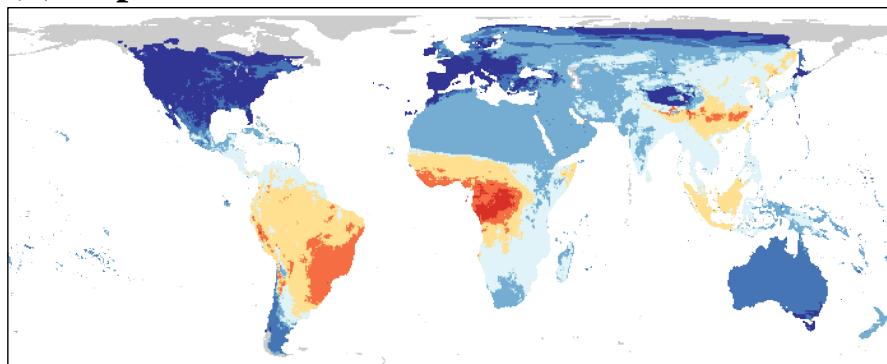
### (a) Amphibians



Median completeness:

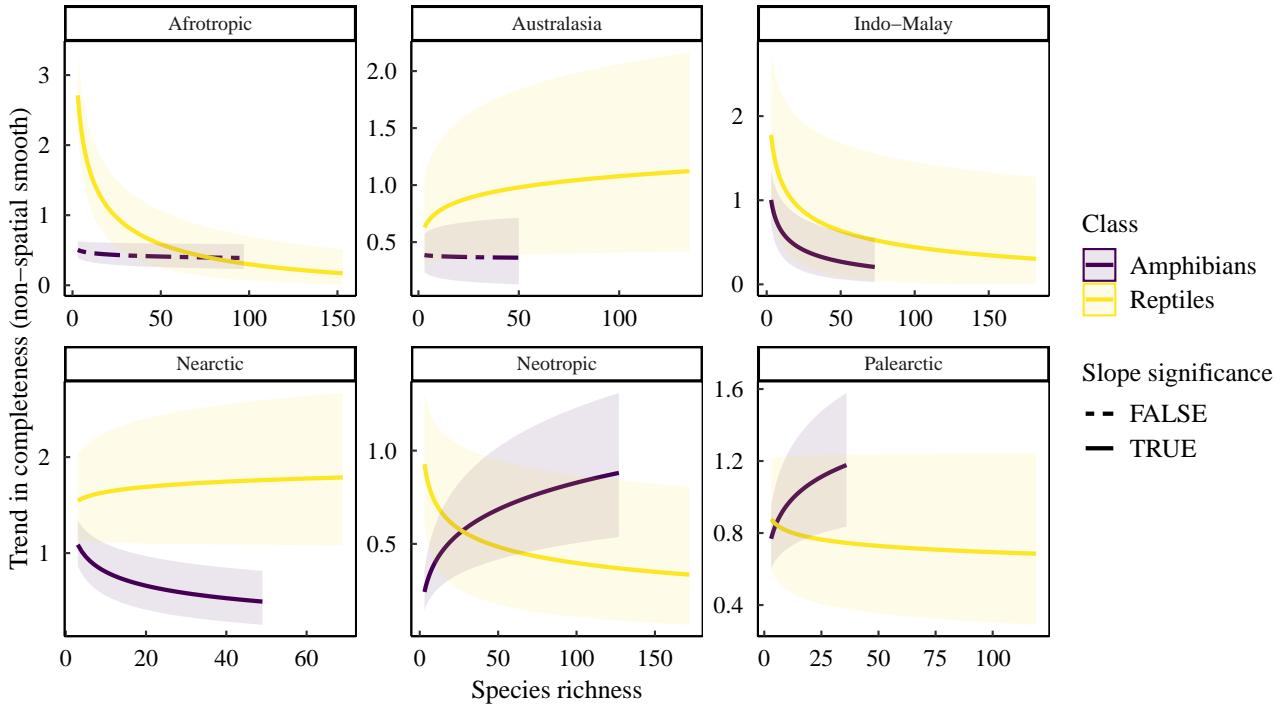
NA
0 – 10%
10 – 20%
20 – 30%
30 – 40%
40 – 50%
50 – 60%
60 – 70%
70 – 80%
80 – 90%
90 – 100%

### (b) Reptiles



**Figure 2.6: Spatial distribution of assemblage-level median trait completeness in herptiles.** Similar maps for birds and mammals are shown in Appendix 2 (S2.5, Figure S2.7). *Figure reproduced from Etard et al. (2020).*

Spatial models showed that species richness explained median trait completeness in herptiles in most realms (Figure 2.7; Appendix 2, S2.8, Tables S2.3 and S2.4); including spatial lags improved the models (amphibians:  $\rho = 0.92$ , p-value < 0.0001; reptiles:  $\rho = 0.91$ , p-value < 0.0001). 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 2.7: Spatial model trends for herptiles.** The lines represent in-sample predictions ( $\pm$  standard error) for the trend components of the spatial models (trends after accounting for spatial autocorrelation). *Figure reproduced from Etard et al. (2020).*

## 2.4 Discussion

The results of this Chapter illustrate the taxonomic, spatial and phylogenetic dimensions of the knowledge gaps in trait data, termed the Raunkiær 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

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 Appendix 2, S2.9, Figure S2.16). 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 (but see Freeman and Pennell (2021): the relative underdescription of species in tropical areas compared to temperate areas –‘taxonomic debt’, also referred to as ‘latitudinal taxonomic gradient’ by the authors– may lead to the underestimation of species richness at low latitudes).

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 than wider-ranging species (Newbold et al., 2018). 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 (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 and 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 I 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, I 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 I 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, I 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 and 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 the 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 and 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 et al., 2019a). 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 and Bühlmann, 2012), multivariate imputation by chained equations (MICE; Buuren and Oudshoorn (2007)) 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 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 | Intensive human land uses negatively affect vertebrate functional diversity**

### **Keywords**

Land use; land-use intensity; terrestrial vertebrates; functional diversity; traits.

### **Abstract**

Land-use change is the leading driver of global biodiversity loss, thus characterising its impacts on the functional structure of ecological communities is an urgent challenge. Using a database describing vertebrate assemblages in different land uses, I assess how the type and intensity of land use affect the functional diversity of vertebrates globally. I find that human land uses alter local functional structure by driving declines in functional diversity, with the strongest effects in the most disturbed land uses (intensely used urban sites, cropland and pastures), and among amphibians and birds. Both tropical and temperate areas experience important functional losses, which are only partially offset by functional gains. Tropical assemblages are more likely to show decreases in functional diversity that exceed those expected from species loss alone. These results indicate that land-use change non-randomly reshapes the functional structure of vertebrate assemblages, raising concerns about the continuation of ecological processes sustained by vertebrates.

### **3.1 Introduction**

Anthropogenic activities are profoundly transforming global biodiversity. Although multiple pressures act in combination, land-use change currently poses the greatest threat to biodiversity (Maxwell et al., 2016; Newbold et al., 2015). However, not all species respond similarly to land-use change. Traits have been found to explain species' sensitivity to land-use change in diverse groups (Newbold et al., 2013; Nowakowski et

al., 2017; Quesnelle et al., 2014; Todd et al., 2017). Previous work has also shown that land-use change leads to non-random modification of assemblage trait composition (or functional diversity) (Chapman et al., 2018; Colin et al., 2018; Flynn et al., 2009; La Sorte et al., 2018; Newbold et al., 2013; Tinoco et al., 2018). Since it is widely acknowledged that biodiversity, and in particular trait diversity, may promote ecosystem functioning and stability, modification to the trait composition of assemblages could have far-reaching and adverse impacts on ecological processes (Hooper et al., 2012; Magioli et al., 2021; Oliver et al., 2015; Tilman and Downing, 1994).

Terrestrial vertebrates support many processes, ranging from pollination (Ratto et al., 2018), to seed dispersal to the regulation of lower trophic levels (Barber et al., 2010; Letnic et al., 2012; Salo et al., 2010; Zhang et al., 2018). However, we lack a global understanding of how the functional diversity of entire vertebrate assemblages responds to changes in land use. Most previous studies have been conducted at regional or local scales (Davison et al., 2021), but these may not be representative of global patterns. Indeed, recent global syntheses have highlighted how biodiversity responses can differ substantially between regions and across latitudes, with higher sensitivity reported for the tropics (Matuoka et al., 2020; Millard et al., 2021; Newbold et al., 2020). Another key issue is the taxonomic coverage of past work. Few studies investigating effects of land use on functional diversity have considered several vertebrate classes together, and comparative studies remain rare. Thus, how land-use change affects the functional diversity of local vertebrate assemblages at global scales, and the potential geographical and taxonomic variation in the effects, still largely remains to be explored.

Here, I aim to assess how human land use and land-use intensity affect the functional diversity of vertebrate assemblages, across and within taxonomic classes. Building on recent work (Matuoka et al., 2020; Millard et al., 2021; Newbold et al., 2020), I investigate differences in response between tropical and temperate regions. I use multiple response metrics to quantify functional diversity. First, functional richness measures the breadth and variety of trait combinations represented in an assemblage (Legras et al., 2018). Second, functional dispersion quantifies how similar species in a given assemblage are in terms of their traits (Laliberte and Legendre, 2010). These metrics can mask important alterations of assemblage composition if functional losses are compensated for by functional gains. To address this, I consider pairwise measures between assemblages, to explore levels of functional loss and functional gain across land uses (Figure 3.1).

To this end, I combine (1) the trait data across terrestrial vertebrates collected in Chapter 2 (also published in Etard et al. (2020)), with (2) global records of species occurrence in eight land-use types of differing intensity of use (the PREDICTS database: Hudson et al. (2014, 2017), Figure 3.1, Table S3.1). The PREDICTS database is currently the most comprehensive database of sampled species occurrence, and for most records

also abundance, across multiple land uses of different land-use intensity. Using the PREDICTS database allows us to contrast biodiversity metrics among intact land uses (primary-vegetation sites, considered to be the undisturbed reference condition), and all other human land-use types. Specifically, I test the following hypotheses, both across and within taxonomic classes:

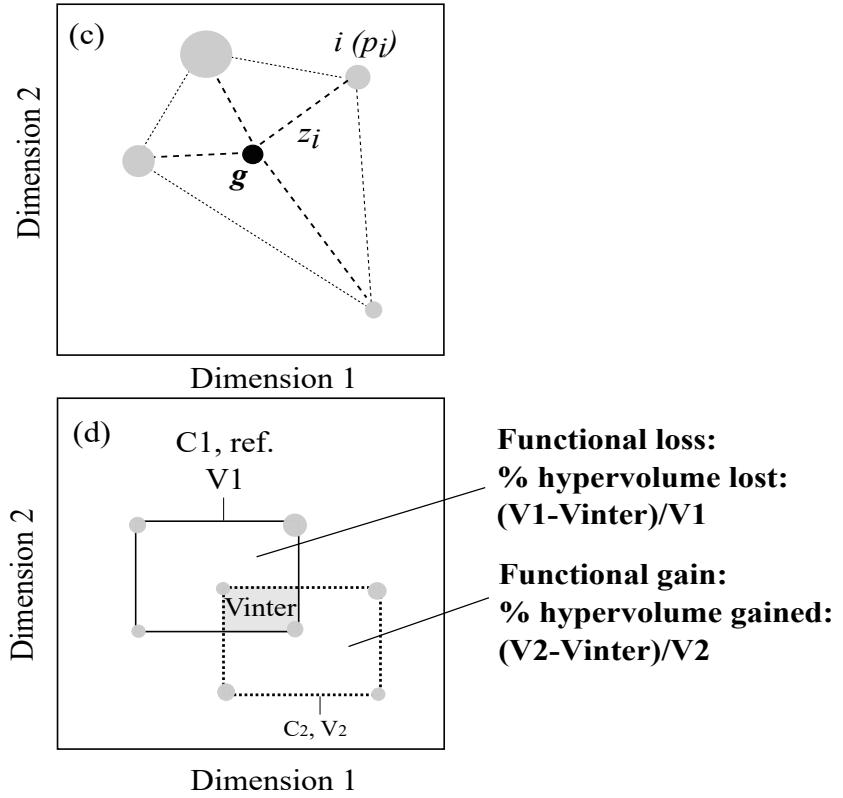
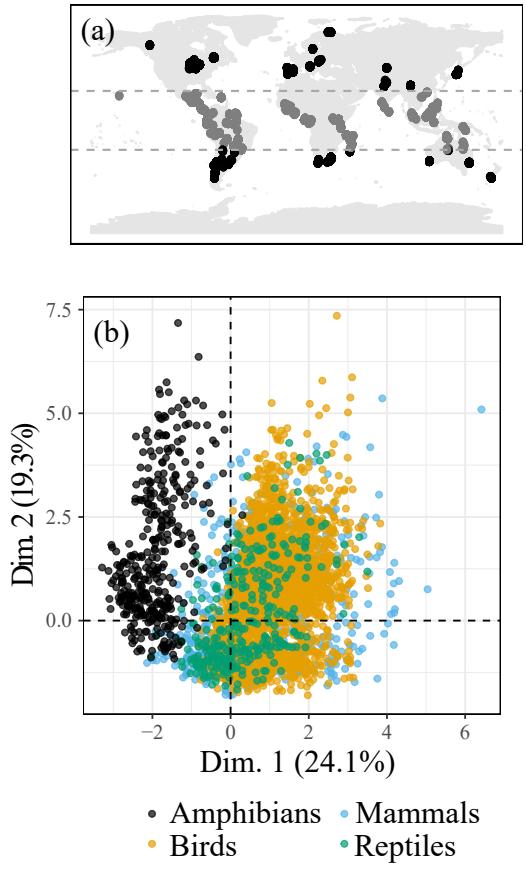
1. I expect decreases in functional diversity in human land uses compared to primary vegetation, caused by contractions of occupied trait space. I expect such effects to be more pronounced where land is used more intensively by humans. This hypothesis builds upon evidence that species with certain traits are more sensitive to land-use disturbance (Newbold et al., 2013), meaning that disturbed land uses will retain only disturbance-tolerant species, more functionally similar to one another. Given the reported higher sensitivity of tropical assemblages to land-use disturbance, I predict that such effects are stronger in the tropics.
2. I hypothesise that decreases in functional diversity in disturbed land uses exceed decreases expected by chance, given local species loss. Thus, I expect disturbed land uses to promote functional under-dispersion. Functional under-dispersion occurs when species within an assemblage are more similar, in term of their traits, than expected by chance (Cadotte and Tucker, 2017; Wong et al., 2018) – or, in other words, when functional dispersion is lower than expected given local species richness. I predict that under-dispersion is more likely to occur in the highly disturbed sites, in both temperate and tropical areas. This hypothesis is based on the idea that species are being removed non-randomly from sensitive areas of the trait space, and increasingly so with higher disturbance level.
3. Finally, I expect decreases in functional diversity in human land uses to be driven by high functional loss, whereby species are being removed from previously occupied areas of the trait space; I expect no functional gain. This hypothesis is based on the idea that the functional trait space in undisturbed land uses represents all of the possible regional trait combinations and that species with functional attributes rendering them unable to persist in altered conditions will be filtered out (Cornwell et al., 2006).

## 3.2 Methods

### 3.2.1 Vertebrate assemblages

I used vertebrate occurrence data from the PREDICTS database (Hudson et al., 2014, 2017), a collection of studies that recorded species occurrence across multiple land uses and land-use intensities. In PREDICTS,

- Tropical sites (3914)
- Temperate sites (2373)



**Figure 3.1: Overview of the study design and functional metrics.** I used occurrence data for vertebrate species from the PREDICTS database ((Hudson et al., 2014, 2017); 180 studies; 431,170 records; 4,339 species; 6,758 sampled sites). (a) shows the spatial distribution of sites I consider. I combine occurrence data with trait data compiled in Chapter 2 to calculate functional metrics. (b) is a representation of the trait data in two dimensions, plotted across PREDICTS vertebrates. Traits that contributed most to dimension 1 were lifespan (29%) and litter/clutch size (22%), while traits that contributed most to dimension 2 were habitat breadth (47%) and use of artificial habitats (35%). (c) and (d) present the conceptual framework for the calculation of the functional diversity metrics: local measures (c) and pairwise metrics (d). (c) Given a trait space, functional richness is calculated as the hypervolume occupied by the minimum convex hull encompassing all species Villéger et al., 2008. Functional dispersion is calculated as the mean distance of the species to the centroid,  $g$  Laliberte and Legendre, 2010. (d) I compute functional loss as the proportion of hypervolume lost from the reference assemblage, and I define functional gain as the proportion of hypervolume of the disturbed assemblage that was gained (proportion of novel trait space in the disturbed assemblage). *Figure reproduced from Etard et al. (2021).*

each study contains several sites, which may be clustered into spatial blocks. Assemblage and land-use data are available at the site level: one site is characterised by a unique land use of given land-use intensity and provides occurrence data for a set of sampled taxa (and the same set of taxa is sought at all other sites within a study). Sites located between  $23.5^{\circ}\text{N}$  and  $23.5^{\circ}\text{S}$  of latitude were considered tropical, and otherwise temperate (Figure 3.1).

Land uses in PREDICTS were assigned to the following categories, based on the descriptions of the

habitat given by the original collectors of the data: primary vegetation (considered to be the undisturbed reference); secondary vegetation; plantation forest; pasture; cropland; urban (considered human, or disturbed; Table S3.1; Hudson et al., 2014, 2017). Secondary vegetation is further divided into three categories: mature, intermediate and young, depending on the stage of recovery of the vegetation. Land-use intensity is reported as minimal, light or intense, according to criteria that depended on the land-use type in question (e.g., crop diversity, degree of mechanisation and chemical inputs in cropland, or bushmeat harvesting and selective logging in primary vegetation; Hudson et al., 2014). I excluded sites for which the land use could not be characterised or for which the stage of recovery of secondary vegetation was unclear. As the PREDICTS database is a collection of independent studies, the design of this study was not balanced: the sample size varied across land uses (Figures S3.1, S3.2), and across taxonomic groups (3103 species of birds; 531 mammals; 379 amphibians; 326 reptiles).

### 3.2.2 Functional traits and diversity indices

Trait choice is a critical step when calculating functional diversity metrics, which are highly sensitive to trait selection (Mouillot et al., 2021). However, trait selection trades off with data availability. Here, a constraint was to use similar traits across the different classes. I thus used the seven traits compiled in Chapter 2 across terrestrial vertebrates. Most of these traits were available for at least 50% of the species in each class (except trophic level in amphibians and lifespan in herptiles; Figure S3.4). In addition, I chose these traits as they were ecologically relevant, thus I broadened the biological definition of traits (i.e., a characteristic measurable at the level of an individual) to include measures of habitat breadth and habitat specialisation (still theoretically measurable at the level of an individual). The final set constituted seven traits that influence species responses to environmental change: body mass, trophic level, lifespan, litter/clutch size, diel activity, habitat breadth and use of artificial habitats. These traits related to life-history, habitat specialisation and use of geographical space (e.g., habitat breadth is a significant predictor of geographical range size in all classes; Figure S3.3). Here, I did not consider estimations of dispersal abilities or home range size as these were available for a small fraction of the species (<3%, Alex Smith and M. Green, 2005; Paradis et al., 1998; Sutherland et al., 2000; Whitmee and Orme, 2013), neither did I include geographical range size which is measured across many individuals, and hence cannot be considered a trait. As in Chapter 2, I did not consider intraspecific trait variation, thus assuming no effect of the environment on trait values.

Trait coverage was variable among classes and traits, with important gaps for reptiles and amphibians (Figures S3.4, S3.5; Chapter 2; Etard et al. (2020)). I imputed missing trait values using random forest algorithms (missForest package: Stekhoven and Bühlmann (2012), Stekhoven (2016)), including traits, tax-

onomic order and phylogenetic eigenvectors as predictors (Debastiani et al., 2021; Penone et al., 2014). To further assess the sensitivity of the results to imputation (see next section), I imputed missing trait values eight times, thereby obtaining eight sets of imputed traits. I randomly selected one imputed trait set for the calculation of functional metrics. Imputations of missing trait values & imputation performance are detailed in Appendix S3.2 and Appendix S3.4 (and see Figures S3.6-S3.8). Post-imputation, continuous traits were log<sub>10</sub>-transformed (except habitat breadth which was square-rooted) and z-scored (standardised to unit variance and zero mean). In addition, I also assessed whether the results were robust to imputation error using a subset of the PREDICTS data considering only species for which I had complete trait information (see next section).

Correlation among traits can be a safeguard against high sensitivity of functional metrics to trait omission, notably where omitted traits correlate strongly with traits that are already included in the calculation (Mouillot et al., 2021). Nevertheless, high multicollinearity among traits has been reported as potentially problematic for the calculation of functional diversity (Cadotte et al., 2011). Thus, I verified that the degree of multicollinearity among traits was not problematically high (with a threshold of 5 for variance inflation factors, Table S3.3). Furthermore, I tested the sensitivity of the results to trait omission, by investigating whether adding geographical range size in the calculation of functional metrics was likely to affect the results.

### 3.2.3 Effects of land use and land-use intensity on FRic and FDis (Hypothesis 1)

For each assemblage, I measured functional richness using ‘FRic’ (Villéger et al., 2008), and functional dispersion using ‘FDis’ (Laliberte and Legendre (2010); Figure 3.1), from the FD package (Laliberte and Legendre (2010); Laliberté et al. (2015)). I assessed the effects of land use, land-use intensity, and region (temperate versus tropical) on FRic and FDis across and within taxonomic classes using linear mixed-effects models (lme4 package, Bates et al. (2015)). Land use and land-use intensity were not ranked in the models. A random intercept of study identity accounted for variation in experimental design across studies, while a random intercept representing spatial blocks of sampled sites, nested within study, accounted for spatial structuring within studies. To improve normality and bound predictions between 0 and 1, I transformed FRic and FDis using an arcsin-square-root transformation. The best-fitting model was sought using backwards stepwise model selection, starting with the most complex model that included all two-way interactions among the specified main effects. Model fits were compared using likelihood-ratio tests at each iteration of the selection procedure.

Across vertebrates, the starting models included the effects of land use, land-use intensity and region

(temperate versus tropical). The best-fitting model for FRic was:

$$\text{arcsin}(\sqrt{\text{FRic}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use : Region}.$$

(Model 1a)

For FDis, the best-fitting model did not include interactions between land use and region, but the main effect of region was retained:

$$\text{arcsin}(\sqrt{\text{FDis}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity}.$$

(Model 1b)

To investigate differences in responses across classes, I pooled some of the land uses together, because otherwise, sample sizes would have been too low. Mature, intermediate and young secondary vegetation were grouped together as ‘Secondary vegetation’, and cropland and pasture were grouped together as ‘Agricultural land uses’. The starting models included the effects of land use, land-use intensity, region and taxonomic class. For FRic, the best model was:

$$\text{arcsin}(\sqrt{\text{FRic}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Class} + \text{Land use : Land-use intensity} + \text{Land use : Class} + \text{Land-use intensity : Region} + \text{Class : Region}.$$

(Model 2a)

For FDis, regional effects were dropped:

$$\text{arcsin}(\sqrt{\text{FDis}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Class} + \text{Land use : Land-use intensity} + \text{Land use : Class} + \text{Land-use intensity : Class}.$$

(Model 2b)

To assess whether the results were robust to imputation error, I used a subset of the PREDICTS data considering only species for which there were complete trait information (6,212 sites; 442 mammals; 1,975 birds; 78 reptiles; 9 amphibians), and I fitted models again to this data subset. I did not have enough complete trait data among amphibians to be able to consider this class separately, so I first considered amphibians and reptiles together (herptiles), and reptiles only. In addition, I complemented this validation with a sensitivity analysis to variation in imputed values. I calculated FDis and FRic using each of the eight imputed trait datasets and fitted the previous models to each set. I then qualitatively evaluated the congruence of the estimates from the different models. Finally, because there tended to be more sites sampled in primary vegetation than in other land uses (Figures S3.1, S3.1), I ran additional sensitivity tests to assess whether the

results were robust to resampling primary vegetation sites to a number equal to 50 (a sample size close to the median number of sites sampled in land uses other than primary vegetation in both regions (median = 37 for the temperate subset and 57 for the tropical subset, Figure S3.1)).

### **3.2.4 Investigating functional under-dispersion (Hypothesis 2)**

To assess whether effects of land use and land-use intensity on FDis differed from what would be expected by chance given changes in local species richness, I generated null expectations of FDis at each site. I randomised assemblage composition 500 times, drawing species from the corresponding study's species pool while maintaining local species richness. For each site, I thus obtained a null distribution for FDis. Then, I tested whether FDis differed from null expectations using Wilcoxon signed-rank tests. I created a binary variable which was assigned 1 if FDis was significantly lower than null expectations at a given site (significant under-dispersion), and 0 otherwise. I investigated how land use, land-use intensity, region and taxonomic class affected the probability of occurrence of under-dispersion using a generalised linear mixed-effects model with a binomial distribution of errors. The best-fitting model did not retain any effect of taxonomic class:

$$P_{\text{under-dispersion}} \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use : Region.}$$

(Model 3)

### **3.2.5 Functional loss and functional gain (Hypothesis 3)**

I calculated the proportion of trait space that was lost in disturbed land uses compared to reference land uses (functional loss) and the proportion of trait space that was gained in disturbed land uses (functional gain) (Figure 3.1(c)), across and within taxonomic classes. I selected studies where at least one site was sampled in primary vegetation. I then made within study pairwise comparisons between reference assemblages, sampled in primary vegetation, and disturbed assemblages. In addition, I considered all comparisons between pairs of primary-vegetation sites, to create reference pairs. I then investigated how land use, land-use intensity and region affected functional loss and gain across and within taxonomic classes using linear mixed-effects models, controlling for study identity in the random effects. Across vertebrates, the best-fitting model for functional loss was:

$$\text{arcsin}(\sqrt{\text{loss}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity} + \text{Land use : Region.}$$

(Model 4a)

For functional gain, one interaction term (land use with region) was dropped:

$$\text{arcsin}(\sqrt{\text{gain}}) \sim \text{Land use} + \text{Land-use intensity} + \text{Region} + \text{Land use : Land-use intensity}.$$

(Model 4b)

When considering the effects of taxonomic class, the best-fitting model for functional loss was:

$$\begin{aligned} \text{arcsin}(\sqrt{\text{loss}}) &\sim \text{Land use} + \text{Land-use intensity} + \text{Class} + \text{Region} + \text{Land use :} \\ &\quad \text{Land-use intensity} + \text{Land use : Class} + \text{Land use : Region} + \text{Land-use intensity : Class}. \end{aligned}$$

(Model 5a)

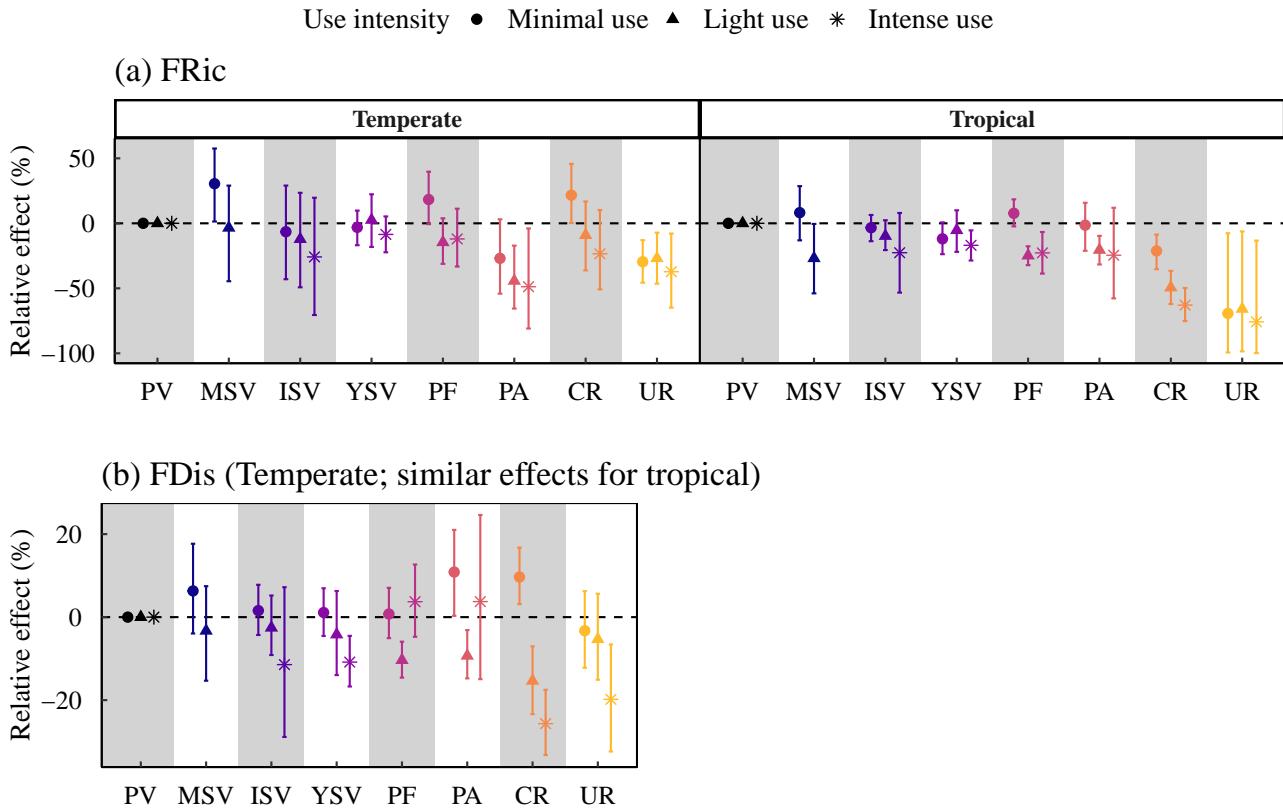
For functional gain (Model 5b), the fitted effects were the same as those of Model 2b. More details about the calculation of functional loss and gain can be found in Appendix S3.5.

All data analyses were conducted using R version 3.5.1 (R Core Team, 2018). I made the code available on figshare (DOIs: <https://doi.org/10.6084/m9.figshare.14161883> and <https://doi.org/10.6084/m9.figshare.15163926>), as well as the main result datasets (<https://doi.org/10.6084/m9.figshare.15163971>).

### 3.3 Results

#### 3.3.1 Effects of land use on FRic and FDis

Across all vertebrates, land use and land-use intensity significantly affected FRic and FDis (Figure 3.2). FRic tended to decrease with increasing disturbance level and higher intensity of land use. For FRic, relative effects differed between regions (Figure 3.2(a)). Although declines were overall more important for disturbed tropical assemblages, significant declines were observed for the temperate assemblages (e.g., a 37% average decline in intensely used urban areas; a 49% decline in pastoral areas of high land-use intensity). Nevertheless, tropical assemblages typically showed more important reductions in FRic. For instance, declines averaged 63% for intensely used tropical cropland and 76% for urban areas. For FDis, relative effects were similar in both regions (Figure 3.2(b)). The most important average declines were observed for urban assemblages of intense use (20% decline), and for lightly- and intensely used cropland (by 15% and 26%). Note that confidence intervals around the estimated average declines were large in some cases, highlighting some heterogeneity in the responses.

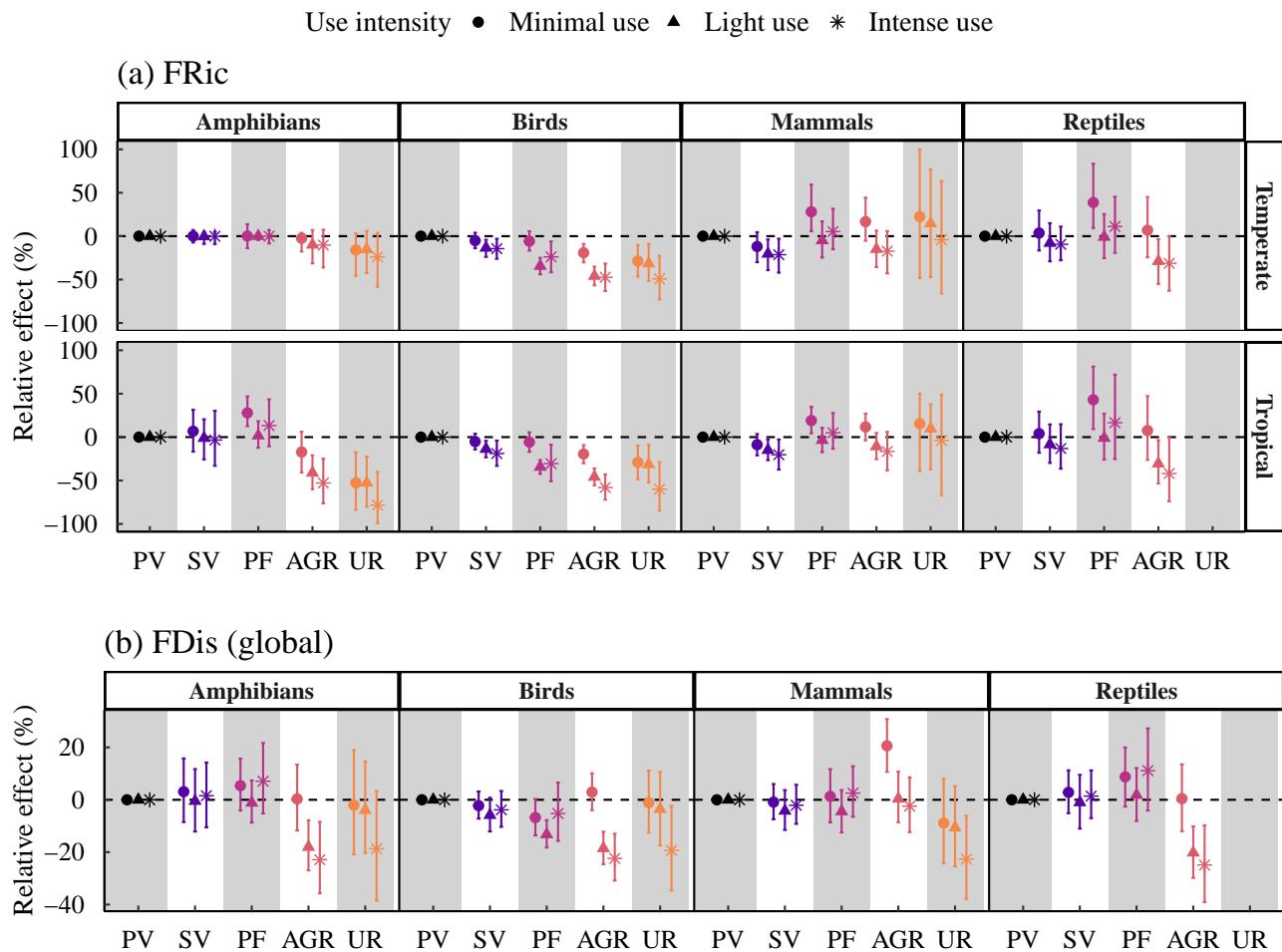


**Figure 3.2: Effects of land use, land-use intensity and region on FRic (a) and on FDis (b) across all PREDICTS vertebrates.** Effects were rescaled with reference to primary vegetation (PV) and are plotted as a % difference relative to PV within each land-use intensity. For FRic, the best-fitting model included interactions between land use and region, while these interactions were dropped for FDis, explaining the similar relative effects in both regions. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV, intermediate secondary vegetation; YSV, young secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Effects for intense use in MSV could not be estimated as there were not enough sampled sites. *Figure reproduced from Etard et al. (2021).*

Fitting the same models to the subset of species with complete trait data, I detected important declines in functional diversity in a number of land uses, showing that the conclusions are robust to trait imputation uncertainty (for example, FRic declined on average by 75% in intensely used temperate pastoral assemblages; by 48% for intensely used tropical cropland; and FDis declined by an average 37% in intensely used tropical urban assemblages; Figure S3.18). Furthermore, using the subset of species with complete trait data, I found that the results were not sensitive to the inclusion of geographical range size as an additional trait (Figure S3.19). Finally, the results were not sensitive to variation across imputed trait values (Figure S3.20) and were also robust to resampling in primary-vegetation sites (Figure S3.21).

Responses of FRic and FDis to land use and land-use intensity differed among taxonomic classes (Figure 3.3). Within-class effects for FDis were similar between regions. The most notable decreases were observed in lightly- and intensely used agricultural land uses in amphibians, birds and reptiles; and in intensely used

urban land uses for birds and mammals. For FRic, the effects in tropical and temperate regions were qualitatively similar in three out of four classes (birds, mammals and reptiles), although effect sizes tended to be bigger for tropical assemblages. Birds and reptiles showed reductions in disturbed land uses in both tropical and temperate regions, whereas I detected few significant effects for mammals. For birds, the most important average decline, of 50%, was observed in intensely used tropical urban land uses, while for reptiles I detected significant decreases in lightly- and intensely used agricultural sites (but I could not estimate effects for urban land uses due to the small sample size). Finally, the effects differed between tropical and temperate regions for amphibians, with no significant effects detected across temperate assemblages, but important reductions across tropical agricultural and urban assemblages.



**Figure 3.3: Effects of land use, land-use intensity, region and taxonomic class on FRic (a) and FDis (b).** Effects were rescaled with reference to primary vegetation (PV) and are plotted as a % difference relative to PV within each land-use intensity. Error bars represent 95% confidence intervals. Effects for FRic were estimated from Model 2a, and from Model 2b for FDis. SV, secondary vegetation; PF, plantation forest; AGR, agricultural land uses (pasture and cropland); UR, urban. Effects for reptiles in urban land uses could not be estimated as there were not enough sampled sites. *Figure reproduced from Etard et al. (2021).*

Fitting similar models only for species with complete trait data showed that these patterns are unlikely to be affected by imputation uncertainty for birds; for mammals and reptiles, the main results could even be conservative (Figure S3.22, S3.23). Indeed, although confidence intervals around the estimates were large, I typically observed larger decreases in functional diversity when using the complete data subset, including an 86% decline in FRic for mammals in intensely used tropical agricultural areas. The results were also unaffected by variation across replicate sets of imputed trait values (Figure S3.24).

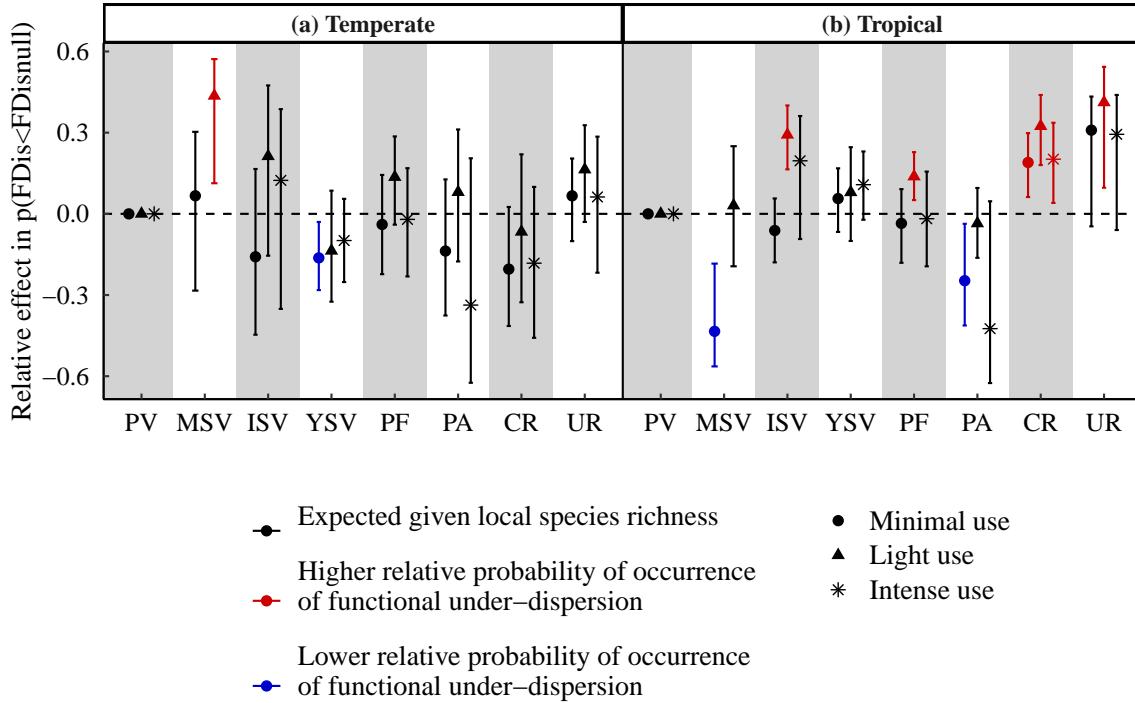
### **3.3.2 Changes in the probability of occurrence of functional under-dispersion**

Land use, land-use intensity and region significantly affected the probability of occurrence of functional under-dispersion across vertebrates. Functional under-dispersion was more likely to occur in tropical cropland of all land-use intensities (Figure 3.4(b)), as well as in some of the lightly-used land uses (notably urban and plantation forest). Contrary to my expectations, and with the exception of tropical cropland, functional under-dispersion was not more likely to occur in intensely-used land uses. For minimally-used sites, changes in FDis were mostly consistent with that expected given changes in species richness.

### **3.3.3 Functional loss and gain**

Across and within vertebrate classes, I detected high levels of functional loss, exceeding the natural turnover between primary-vegetation sites, both in temperate and tropical regions. Across vertebrates (Figure 3.5(a)), functional loss was notably high in temperate pastures (+27% above reference for minimal use; +73% for intense use), temperate urban sites (+27% for light use; +50% for intense use; effects for tropical urban sites could not be estimated), temperate and tropical cropland (+44% and +56% respectively for light use; effects for intense use could not be estimated). Important levels of functional loss were also observed in tropical plantation forest of light use intensity (+51%; effects for the intense use could not be estimated). High levels of functional loss were also observed within each class (Figure 3.6(a)) (although not all effects could be estimated because of limited sample sizes, Table S3.5). The highest losses were observed in agricultural areas for amphibians and reptiles, with important losses also observed in temperate urban areas for both birds and amphibians (+35% for minimal use; effects for tropical urban areas could not be estimated).

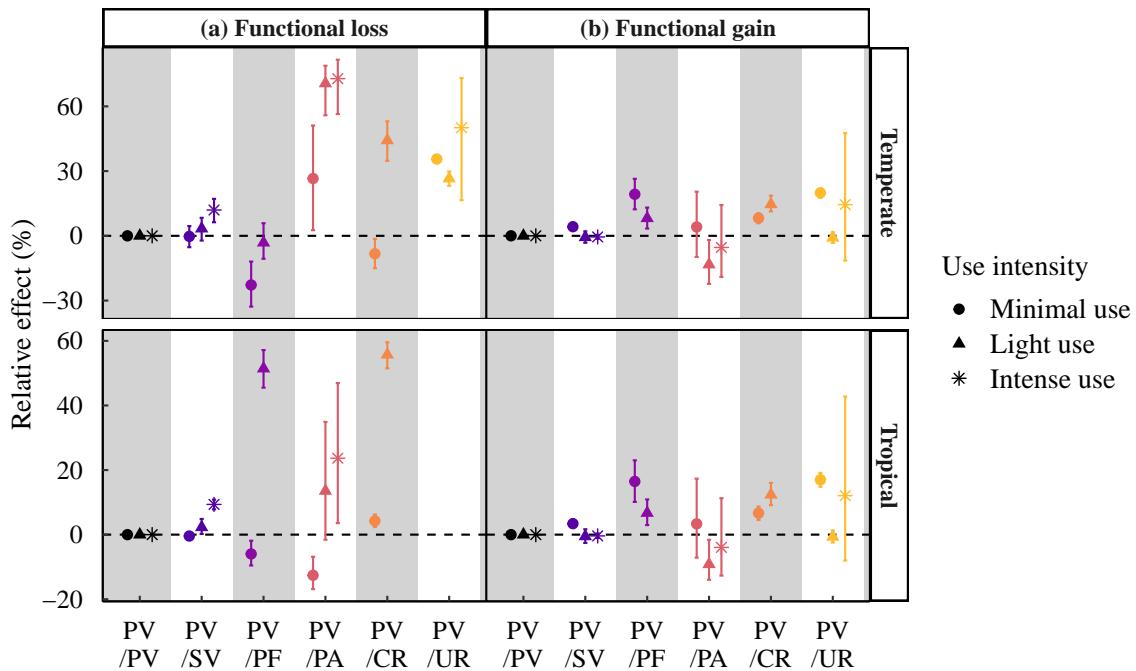
Across vertebrates (Figure 3.5(b)), average functional gain (average proportion of novel trait space in the disturbed assemblage) was moderate and on average did not exceed 20% in any disturbed land uses. Patterns of functional gain were similar in both regions. The highest functional gains were observed for minimally-used urban sites and plantation forest (range: +16% to +20%). On the other hand, important levels of functional gain were observed in some classes (Figure 3.6(b)), with the highest functional gain for



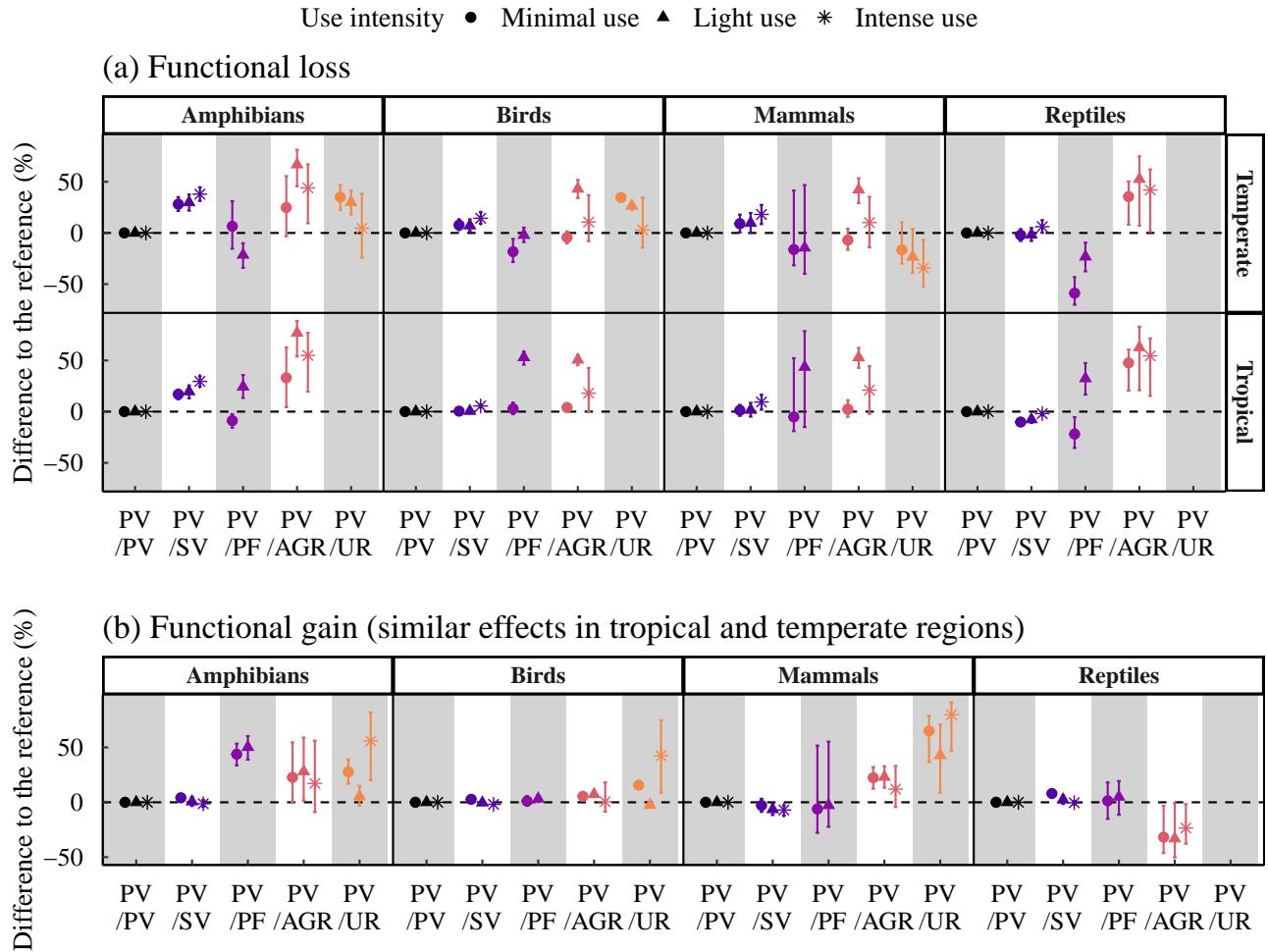
**Figure 3.4: Effects of land use, land-use intensity and region on the probability of occurrence of functional under-dispersion.** Error bars represent 95% confidence intervals. PV: primary vegetation; MSV, mature secondary vegetation; ISV, intermediate secondary vegetation; YSV, young secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Effects are rescaled and represent the average difference in the probability of occurrence of functional under-dispersion between the reference (PV, probability of functional under-dispersion set at 0 within each land-use intensity) and the disturbed land uses. *Figure reproduced from Etard et al. (2021).*

mammals (+80% in intensely used urban sites).

Diagnostic plots (qq-plots and residual distributions) for the models are shown in Figures S3.9–S3.17. Overall, the model residuals were appropriately distributed (but with some leptokurtic residual distributions, to which mixed-effect models are generally robust (Schielzeth et al., 2020)).



**Figure 3.5: Effects of land use and land-use intensity on (a) functional loss and (b) functional gain across pairs of sites in temperate areas and tropical areas.** PV: primary vegetation; SV, secondary vegetation; PF, plantation forest; PA, pasture; CR, cropland; UR, urban. Error bars represent 95% confidence intervals. Effects are rescaled and represent the average difference in functional loss and gain between the reference pairs (PV/PV, with loss and gain set at 0% within each land-use intensity), and pairs of primary vegetation with disturbed land uses. Negative effects mean that, on average, levels of functional loss (or gain) were lower than functional loss (or gain) observed between pairs of primary-vegetation sites, and not that there were absolute ‘negative losses’ or ‘negative gains’. *Figure reproduced from Etard et al. (2021).*



**Figure 3.6: Effects of land use, land-use intensity, region and taxonomic class on functional loss and functional gain across pairs of sites.** PV: primary vegetation; SV, secondary vegetation; PF, plantation forest; AGR, agricultural land uses (pasture and cropland); UR, urban. Error bars represent 95% confidence intervals. Effects are rescaled and represent the average difference in functional loss and gain between the reference pairs (PV/PV, with loss and gain set at 0% within each land-use intensity), and pairs of primary vegetation with disturbed land uses. Negative effects mean that, on average, levels of functional loss (or gain) were lower than functional loss (or gain) observed between pairs of primary-vegetation sites, and not that there were absolute ‘negative losses’ or ‘negative gains’. Figure reproduced from Etard et al. (2021).

### 3.4 Discussion

Here, I showed that the functional diversity of vertebrate assemblages is negatively impacted in human land uses, particularly in the most intensely used land types. The results of this Chapter extend previous studies that have been more taxonomically or geographically restricted (Flynn et al., 2009; Matuoka et al., 2020). Matuoka et al. (2020) found that the functional diversity of tropical bird assemblages was negatively affected by human disturbance, a pattern that did not appear in temperate assemblages. Yet, I found that functional diversity was negatively affected in both tropical and temperate, with important functional losses in all four vertebrate classes.

Using multiple metrics allowed me to explore different facets of functional diversity. For instance, functional gain could locally offset functional loss in some disturbed land uses. This could indicate that despite no apparent negative effect on FRic, some disturbed land uses (e.g. lightly-used temperate cropland) could experience important functional loss, and highlights the importance of using a variety of indicators. This mechanism could be at play in mammalian assemblages, for which important levels of functional gain were observed in agricultural and urban sites. Further, functional gain in disturbed land uses could indicate that disturbances facilitate the introduction of functionally novel species, falling into previously unoccupied parts of the trait space. This may be because non-native species are more likely to become established in disturbed assemblages. Previous work has shown that land-use disturbance facilitates biological invasions in island ecosystems (Jesse et al., 2018; Sanchez-Ortiz et al., 2019), but to my knowledge, this has not been tested specifically across continental areas for invasive vertebrates (but see Pyšek et al. (2010)). It is also possible that disturbed areas harbour synanthropic species that do not occur in primary vegetation, leading to substantial functional gain.

Overall, the negative effects of land use on functional richness tended to be more pronounced in the tropics. This is congruent with past studies that have found tropical biodiversity to be disproportionately sensitive to human pressures (Martins and Pereira, 2017; Newbold et al., 2020). There are a number of potential explanations for this. First, it could be that a long history of intense land-use disturbance at large scales in many temperate regions (e.g. Western Europe; Stephens et al. (2019)) means that biodiversity is now less sensitive to new disturbances, because the most sensitive species have been filtered out (Balmford, 1996; Krauss et al., 2010; Le Provost et al., 2020; Munteanu et al., 2020). Species unable to cope with such disturbances may have gone extinct in the past, while the remaining species would be more disturbance-tolerant (Betts et al., 2019). Tropical regions, historically less disturbed at large scales, would then contain a higher proportion of disturbance-sensitive species than temperate regions. Consequently, the functional

richness in undisturbed tropical sites could be less resilient to new disturbances. This also highlights that time since land-use conversion could have important impacts on local functional diversity. Although I did not consider the effects of time since land-use conversion in this work (notably because PREDICTS contained data only for about 22% of the sites), I expect that time since land-use conversion may affect assemblage composition, and thus, functional diversity, with potentially land-use-specific relationships between time since conversion and functional diversity (e.g., a positive relationship for recovering secondary vegetation or a negative relationship for urban areas; but I did not detect such effects when using the data subset for which I have information on time since land-use conversion [see Appendix 3, section S3.8]).

Second, it could be that tropical species are intrinsically more sensitive to disturbances than temperate species because of their evolutionary history. Natural climatic variability experienced by species as well as species history of exposure to disturbances have been proposed to influence sensitivity to disturbance. For instance, tropical species are, on average, nearer to their climatic limits than temperate species (Deutsch et al., 2008; Sunday et al., 2014). Tropical species could therefore experience more deleterious effects from interacting drivers of change, with land-use change bringing about novel climatic conditions pushing them beyond their tolerance limits (Frishkoff et al., 2016; Williams and Newbold, 2020).

In addition to filtering out sensitive species, land-use change is also expected to modify interactions among species, thereby influencing species persistence (Tylianakis et al., 2008; Valiente-Banuet et al., 2015). Although I detected a signal of functional under-dispersion (particularly in tropical cropland), which indicates that assemblages may be locally structured by environment filtering (Bregman et al., 2015), it is likely that several assembly rules underpin assemblage composition (Fournier et al., 2016). For instance, land-use changes could enhance competition among species, promoting over-dispersion by removing species that share similar resources. Such opposite signatures of environmental filtering and enhanced competition on functional dispersion could explain why I did not detect stronger effects of land use on functional under-dispersion occurrence.

Studies looking at impacts of global land use on functional diversity computed with species from all four terrestrial vertebrate classes remain rare. Lack of availability of standardised trait data across terrestrial vertebrates may have hindered such studies from being conducted in the past. To overcome this problem, I based the analyses on a large-scale collation of trait data (Chapter 2; Etard et al. (2020)), and I imputed missing trait values to obtain complete trait datasets in each class. I used random forest algorithms, currently thought to be one of the most robust technique for missing value imputations in trait datasets (Debastiani et al., 2021; Johnson et al., 2021; Penone et al., 2014). Replicating the analyses on complete trait data subsets showed that imputation uncertainty did not affect the main conclusions of this work and that the

negative effects of human land uses were in some cases even stronger when using the complete data subsets. Furthermore, the results were highly consistent across imputed datasets and so insensitive to variation across imputed values. Although missing value imputation can offer a robust filling of missing entries, this study highlights the existing taxonomic biases both in trait data availability and in PREDICTS studies, and thus stresses the need to pursue data compilation efforts, particularly for the least-sampled classes (reptiles and amphibians).

Another implication of trait data availability for vertebrates is that the choice of traits was constrained. Mouillot et al. (2021) showed that functional diversity metrics are sensitive to trait omission and that the sensitivity to trait omission decreases with increasing levels of correlation among traits. Here, I chose seven traits that were available across all classes at least for a subset of the species and that have been implicated in shaping species responses to environmental change. A notable omission was any metric of dispersal ability, which is likely to influence species' ability to respond to land-use change but is difficult to obtain for most species. In fact, past studies have shown that dispersal abilities can be predicted from ecological correlates, such as body mass, diet or geographical range size (Schloss et al., 2012; Sutherland et al., 2000). Since the results were robust to the omission of geographical range size, I am confident that the omission of dispersal abilities also does not affect the conclusions of this work.

Functional diversity metrics are often used as a proxy for ecosystem functioning because of the conceptual and mechanistic link between functional 'effect' traits and ecosystem processes (Lavorel and Garnier, 2002; Violle et al., 2007). In many studies focused on vertebrates, however, functional diversity metrics do not correlate with a given ecosystem function (Hatfield et al., 2018). Here, I did not explicitly target given ecosystem functions, but I argue that evidence of functional loss of vertebrate assemblages indicates that processes sustained by vertebrates are put at risk by land-use change. My results further show that some disturbed land uses are more likely to experience functional under-dispersion, particularly tropical cropland and tropical urban areas, which again indicates a potential imperilment of ecological processes. Indeed, in such cases, decreases in functional dispersion exceed changes expected from the chance removal of species; such non-random modifications indicate that certain areas of the functional trait space are more sensitive to land-use disturbance. Future work could investigate the impacts of land-use change on particular ecosystem functions. The integration of trophic information (beyond the trophic levels I used here) to the species-trait dataset could be an interesting step in that direction, as dietary traits relate to resource use and are, as such, probably the most straightforward traits to link with ecosystem functions. Furthermore, my results suggest that the functional loss experienced within a class is unlikely to be compensated for by the persistence of functionally similar species in other classes. Indeed, I detected negative effects of human land use on func-

tional richness in at least three out of four vertebrate classes (amphibians, birds, and reptiles), in accordance with past studies focusing on each of these groups (Gallmetzer and Schulze, 2015; Marcacci et al., 2021; Riemann et al., 2017; Sol et al., 2020). Although overall mammalian functional richness was less affected, high levels of functional gain suggest that the functional composition of mammalian assemblages is heavily modified in disturbed land uses.

To conclude, the results of this third chapter highlight the negative impacts of human land uses on multiple dimensions of functional diversity, within and across terrestrial vertebrate classes, at a global scale. In many disturbed sites, decreases in functional diversity exceed changes expected from species loss alone, showing that human activities non-randomly reshape ecological assemblages. By intensifying functional loss and promoting functional under-dispersion, land-use change could have deleterious effects on ecosystem functioning, highlighting the necessity of putting into place effective conservation measures in the face of anthropogenic change.

# **4 | Geographical range area, habitat breadth and specialisation on natural habitats predict land-use responses and climate-change sensitivity more consistently than life-history and dietary traits in terrestrial vertebrates**

## **Keywords**

## **Abstract**

### **4.1 Introduction**

### **4.2 Methods**

### **4.3 Results**

### **4.4 Discussion**

# **5 | Energetic constraints and trophic levels explain species persistence in disturbed land uses**

## **Keywords**

Land use; land-use intensity; metabolic rates; energetic constraints; terrestrial vertebrates; trophic level; occurrence.

## **Abstract**

Land-use change is the primary driver of global biodiversity loss. In terrestrial vertebrates, previous work has shown that sensitivity to land-use change depends on species traits, but the extent to which energetic constraints explain species responses to disturbed land uses remains largely unexplored. Here, I investigate relationships between the energetic requirements of terrestrial vertebrates (estimated from resting metabolic rates) and land-use change, at two levels of organisation. First, at the assemblage level, I hypothesize that total energetic requirements in disturbed land uses are lower than in undisturbed land uses (assuming that there is less energy available in these areas). Second, after controlling for the effects of body mass and taxonomy on metabolic rates, I predict that species with relatively lower energetic expenditure are favoured over species with relatively higher energetic expenditure in disturbed land uses, as resource efficiency may be beneficial in resource-poor environments. Because trophic level can influence species ability to assimilate various types of food, I investigate whether my predictions are consistent among trophic levels (here, omnivores, carnivores or herbivores).

The results challenged both hypotheses. I found that total assemblage-level energetic requirements did

not systematically decrease in disturbed land uses. For instance, I detected significant increases for urban areas in all trophic levels, highlighting that disturbed areas may not be as energy-poor as I initially assumed. Second, I found a positive effect of metabolic rates (after controlling for body mass and taxonomy) on species probability of occurrence across all trophic levels for at least one of the most disturbed land uses I consider (pasture, cropland and urban). Species with higher energetic expenditure may display a set of characteristics rendering them more able to cope with disturbances, such as higher activity levels or bigger brain sizes, which could help them make use of the resources available in disturbed environments. The findings of this Chapter highlight that land-use change has significant impacts on vertebrate community metabolism.

## 5.1 Introduction

## 5.2 Methods

### 5.2.1 Vertebrate assemblage composition

I obtained vertebrate assemblage composition in different land uses from the PREDICTS database (Hudson et al., 2014, 2017). The PREDICTS database is a large collection of published studies that measure biodiversity across different land uses and is one of the most comprehensive global databases of its type. In each PREDICTS study, species occurrence and often abundance were recorded across different sites. Each site was assigned to one of the following land-use types: primary vegetation (natural, undisturbed vegetation), secondary vegetation (recovering after complete destruction of primary vegetation), plantation forest (woody crops), pasture (areas grazed by livestock), cropland (herbaceous crops) and urban (built-up areas). The land-use categories were assigned based on habitat descriptions from the original studies (Hudson et al. (2014), sometimes in consultation with the original study authors. Each site was also classified in terms of land-use intensity as either minimal, light or intense. The land-use-intensity assignment was also made on the basis of the habitat description in the original studies, and depended on criteria specific to each land use (such as degree of mechanisation, yield or chemical inputs for cropland; or the amount of green space in urban areas; Hudson et al. (2014)).

I subset the PREDICTS database for studies that sampled terrestrial vertebrates, and for which both land use and land-use intensity had been characterised. I thus obtained 181 studies for 4,238 species sampled across 6,484 sites (Fig. 1a). Sample sizes varied across land uses and land-use intensities (Fig. 1b).

## **5.2.2 Energy availability by land-use type and land-use intensity**

The predictions of this Chapter rely on the assumption that there is less energy available in disturbed compared to undisturbed land uses. To test this assumption, I used terrestrial net primary productivity (NPP) across land uses as a proxy for available energy. NPP quantifies the amount of atmospheric carbon fixed by plants and accumulated as biomass. I derived NPP using imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) on board NASA's Terra satellite. NPP estimates are based on a yearly composite of measures made at 8-day intervals, captured at 500-m spatial resolution (Running & Zhou, 2015). NPP was obtained for 4,062 of the PREDICTS database sites used in the analysis (I matched the sites to the NPP data using the sampling year available in PREDICTS). I fit a linear mixed-effects model (*lme4* package, version 1.1-23, Bates et al. (2015)) explaining site-level NPP by land use and land-use intensity, with a random intercept accounting for study identity, to control for differences in experimental design across studies. Model predictions showed that NPP decreased significantly in several land uses (e.g., pasture and cropland) compared with the primary vegetation reference level, although the strength and in some cases direction of the difference varied among land-use and intensity combinations (e.g., increases in urban land uses; Fig. 1c).

## **5.2.3 Resting Metabolic Rates (RMR) & imputations of missing RMR values**

As a proxy for species-level energetic expenditure, I used estimates of the minimum amount of energy required for organismal maintenance, i.e., basal metabolic rates (BMR) for endotherms, and resting metabolic rates (RMR) for ectotherms. From the literature, I obtained estimates of BMR for 719 species of birds and 685 mammals, and estimates of RMR for 126 amphibians and 173 reptiles (Supporting information, Table S1). For endotherms, BMR are measured when species are in their thermoneutral zone, that is, when there is little to no energy expenditure allocated to thermoregulation. Thus, BMR estimates were derived from lab studies that mostly measured oxygen consumption of the organisms at rest under controlled conditions and in the thermoneutral zone of the species. For an ectotherm, there is no 'basal' metabolic rate, as body temperature mainly depends on environmental temperature. Their metabolic rates follow a hump-shaped relationship with environmental temperature, highest at an optimal temperature which corresponds to a performance peak. To be able to compare endotherms' BMR with ectotherms' RMR, Stark et al. (2020) used the metabolic rates that correspond to a performance peak for both groups (i.e., BMR in the thermoneutral zone for endotherms, and metabolic rates at optimal temperature for ectotherms). Thus, I used the data compiled in Stark et al. (2020) for ectotherms, and from the sources specified in Table S1 for endotherms. The units for BMR and RMR were standardized to mL of dioxygen consumed per hour (mLO<sub>2</sub>/h). As in Stark

et al. (2020), I henceforth refer to both basal and resting metabolic rates as RMR.

For the species occurring in PREDICTS, initial data coverage for RMR was poor (Table S1), necessitating imputation of missing values. To do so, I first measured the phylogenetic signal in BMR and RMR ( $\log_e$ -transformed), using Pagel's  $\lambda$  (Pagel, 1999), to assess whether metabolic rates were sufficiently phylogenetically conserved to be estimated from species phylogenetic position. I obtained class-specific phylogenetic trees from Jetz et al. (2012) for birds, from Faurby et al. (2018, 2020) for mammals, from Tonini et al. (2016) for reptiles (squamates), and from Jetz & Pyron (2018) for amphibians (all downloaded in April 2020). For each class, I randomly sampled 100 trees. To account for phylogenetic uncertainty, I calculated Pagel's  $\lambda$  for each sampled tree and reported the median value, as well as the 2.5th and 97.5th percentiles (Table S1).

In addition to being highly phylogenetically conserved (Table S1), RMR correlate strongly with body mass (Fig. 2a). Thus, I imputed missing values using body-mass information (see next section), phylogenetic relationships and taxonomic orders as predictors (Penone et al. 2014). For each class, I used a consensus phylogenetic tree from which I summarised phylogenetic relationships in the form of five phylogenetic eigenvectors. Including more eigenvectors had little impact on the imputed values (results not shown). Consensus trees were obtained with the TreeAnnotator programme of the BEAST software (Bouckaert et al. 2014). Missing RMR values were imputed using random forests algorithms implemented in R using the 'missForest' package (Version 1.4; Stekhoven & Bühlmann 2012; Penone et al. 2014; Stekhoven 2016).

#### 5.2.4 Trophic level and body mass information

I used body mass and trophic level information for terrestrial vertebrates compiled in Chapter 2. Body mass was compiled as a single measure at the species level, meaning I was unable to consider intraspecific variation. Trophic level described species as either carnivores, omnivores, or herbivores. Because there were gaps in the availability of the data, more so for trophic level than for body mass (see Chapter 2), I imputed the missing trait values (independently of RMR imputations), then used both imputed and empirical body mass values for imputations of missing RMR values. To impute missing body mass and trophic levels, I used random forests algorithms (using the missForest R package; Stekhoven & Bühlmann (2012); Stekhoven (2016)), including as additional predictors phylogenetic information, added in the form of 10 phylogenetic eigenvectors (Diniz-Filho et al. 2012) following Penone et al. (2014), and also taxonomic order. I considered a wider set of life-history traits in the missing values imputations: lifespan, litter/clutch size, habitat breadth and use of artificial habitats (also compiled in Chapter 2). Phylogenetic eigenvectors were extracted from

the class-specific phylogenies using the PVR package (Santos 2018).

### **5.2.5 Effects of land use, land-use intensity and trophic level on assemblage-level total RMR (prediction 1; Figure 2(a))**

Assemblage-level total RMR (tRMR) was obtained by summing abundance-weighted RMR for the species occurring in each site; abundance data were available for 125 of the 181 PREDICTS studies I considered (sampling 3,487 species across 4,644 sites). I fitted a linear mixed-effects model to explain  $\log_e$ -tRMR as a function of land use, land-use intensity and trophic level, with a random intercept accounting for study identity to control for differences in experimental design across studies. I started with a model allowing all two-way interactions among the predictors. I then tested whether adding the three-way interaction among land use, land-use intensity and trophic level improved the fit of the model, using a likelihood-ratio test. The model that included the three-way interaction was retained ( $P \ll 0.01$ ; model 1, Fig. 2). In addition, because it is well established that resting metabolic rates are influenced by temperature (Clarke & Fraser 2004), I checked whether including annual mean temperature in the model affected the conclusions. Annual mean temperature at each PREDICTS site was estimated from WorldClim version 2.1 (Fick & Hijmans 2017), using a 2.5 arc-minute resolution. Adding annual mean temperature did not improve model fit (likelihood-ratio test:  $P = 0.113$ ), thus I did not consider its effects any further.

#### **Model validation.**

To ensure that imputation uncertainty did not affect the conclusions, I refitted model 1 using the subset of species ( $n = 426$ ) from PREDICTS for which there were empirical RMR information (i.e., excluding imputed RMR values).

#### **Disentangling the effects of body mass and abundance on tRMR.**

Since RMR correlates strongly with body mass, changes in tRMR are likely to be driven in part by changes in the size-spectrum of ecological assemblages. I fitted an additional model to explain changes in species' abundance (given presence) by land use, land-use intensity, trophic level, body mass and their interactions, to understand the role of shifts in the body of species on observed changes in tRMR (see Fig. S1).

### **5.2.6 Effects of land use, land-use intensity, trophic level and residual RMR on species occurrence probability (Prediction 2; Figure 2(b))**

To control for the effects of body mass and taxonomy on RMR, I used the residual variation in RMR after accounting for these variables, from a linear mixed-effects model fitting  $\log_e$ -RMR as a function of  $\log_e$ -body mass with nested random taxonomic effects (1|Class/Order/Family; Fig. 2). Hence, I used a metric that describes how much more energy (positive deviations) or less energy (negative deviations) than expected from body mass and taxonomic position a species spends for organismal maintenance. Similar approaches have been used in previous papers (Furness & Speakman 2008; Naya et al. 2013). As detailed earlier, I expect species with lower residual RMR to do better in disturbed land uses than species with higher residual RMR (prediction 2; Fig. 2b) because, given any body mass, investing less energy in maintenance could contribute to persistence in a context of resource scarcity.

To test the second prediction, I fitted a binomial mixed-effects model explaining species occurrence with land use, land-use intensity, trophic level and residual RMR. I started with a complete model that included all two-way interactions among the main effects. Because I wanted to test whether the second prediction was valid for each trophic level, I needed to account for potential differences in the slope of the relationships between occurrence probability and residual RMR among trophic levels. Thus, I performed a forward stepwise selection procedure to test whether adding three-way interactions among (1) land use, trophic level and residual RMR and (2) among land-use intensity, trophic level and residual RMR improved model fit, using likelihood-ratio tests. The final model included both three-way interactions (Fig. 2b; model 2). I fitted random effects that accounted for species identity, as well as for study and site identity within PREDICTS.

#### **Model validation.**

I checked the phylogenetic signal in the model residuals using Pagel's  $\lambda$  (Pagel 1999). Non-significant phylogenetic signal in the residuals would indicate that fitting species identity in the model's random effects was sufficient to account for residual phylogenetic variation in RMR. Further, to assess the potential effects of imputation uncertainty on the results, I again fitted model 2 on the data subset for the 489 species with collected empirical RMR values, across 5,948 sites in 151 studies (i.e., excluding imputed values).

### **5.3 Results**

### **5.4 Discussion**

## **6 | General discussion**

## **7 | Conclusion**

# Bibliography

- Alex Smith, M and M. Green, D (2005). Dispersal and the metapopulation paradigm in amphibian ecology and conservation: are all amphibian populations metapopulations? *Ecography*, 28.1, 110–128.
- Balmford, A. (1996). Extinction filters and current resilience: The significance of past selection pressures for conservation biology. DOI: 10.1016/0169-5347(96)10026-4.
- 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.
- Barber, N. A., Mooney, K. A., Greenberg, R., Philpott, S. M., Van Bael, S. A., and Gruner, D. S. (2010). Interactions among predators and the cascading effects of vertebrate insectivores on arthropod communities and plants. *Proceedings of the National Academy of Sciences*. DOI: 10.1073/pnas.1001934107.
- Barlow, J., França, F., Gardner, T. A., Hicks, C. C., Lennox, G. D., Berenguer, E., Castello, L., Economo, E. P., Ferreira, J., Guénard, B., Gontijo Leal, C., Isaac, V., Lees, A. C., Parr, C. L., Wilson, S. K., Young, P. J., and Graham, N. A. (2018). The future of hyperdiverse tropical ecosystems. DOI: 10.1038/s41586-018-0301-1.
- Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. DOI: 10.18637/jss.v067.i01.
- Betts, M. G. et al. (2019). Extinction filters mediate the global effects of habitat fragmentation on animals. *Science*. DOI: 10.1126/science.aax9387.
- 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. 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. 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., García, 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.
- Bregman, T. P., Lees, A. C., Seddon, N., Macgregor, H. E., Darski, B., Aleixo, A., Bonsall, M. B., and Tobias, J. A. (2015). Species interactions regulate the collapse of biodiversity and ecosystem function in tropical forest fragments. *Ecology*, 96.10, 2692–2704. DOI: 10.1890/14-1731.1.
- Bruggeman, J., Heringa, J., and Brandt, B. W. (2009). PhyloPars: Estimation of missing parameter values using phylogeny. *Nucleic Acids Research*. DOI: 10.1093/nar/gkp370.
- Buuren, S van and Oudshoorn, C. G. M. (2007). MICE: multivariate imputation by chained equations. *R package version*.
- Cadotte, M. W., Carscadden, K., and Mirochnick, N. (2011). Beyond species: Functional diversity and the maintenance of ecological processes and services. DOI: 10.1111/j.1365-2664.2011.02048.x.
- Cadotte, M. W. and Tucker, C. M. (2017). Should Environmental Filtering be Abandoned? DOI: 10.1016/j.tree.2017.03.004.
- 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.
- Chapman, P. M., Tobias, J. A., Edwards, D. P., and Davies, R. G. (2018). Contrasting impacts of land-use change on phylogenetic and functional diversity of tropical forest birds. *Journal of Applied Ecology*. DOI: 10.1111/1365-2664.13073.
- Chaudhary, C., Saeedi, H., and Costello, M. J. (2016). Bimodality of Latitudinal Gradients in Marine Species Richness. *Trends in Ecology & Evolution*, 31.9, 670–676. DOI: <https://doi.org/10.1016/j.tree.2016.06.001>.
- Colin, N., Villéger, S., Wilkes, M., Sostoa, A. de, and Maceda-Veiga, A. (2018). Functional diversity measures revealed impacts of non-native species and habitat degradation on species-poor freshwater fish assemblages. *Science of the Total Environment*. DOI: 10.1016/j.scitotenv.2017.12.316.
- Collen, B., Dulvy, N. K., Gaston, K. J., Gärdenfors, U., Keith, D. A., Punt, A. E., Regan, H. M., Böhm, M., Hedges, S., Seddon, M., Butchart, S. H., Hilton-Taylor, C., Hoffmann, M., Bachman, S. P., and Akçakaya, H. R. (2016). Clarifying misconceptions of extinction risk assessment with the IUCN Red List. *Biology Letters*. DOI: 10.1098/rsbl.2015.0843.
- 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.
- Cornwell, W. K., Schwilk, D. W., and Ackerly, D. D. (2006). A trait-based test for habitat filtering: Convex hull volume. *Ecology*. DOI: 10.1890/0012-9658 (2006) 87 [1465:ATTFHF] 2.0.CO;2.
- Davison, C. W., Rahbek, C., and Morueta-Holme, N. (2021). Land-use change and biodiversity: Challenges for assembling evidence on the greatest threat to nature. *Global Change Biology*, 27.21, 5414–5429. DOI: 10.1111/gcb.15846.
- Debastiani, V. J., Bastazini, V. A. G., and Pillar, D (2021). Ecological Informatics Using phylogenetic information to impute missing functional trait values in ecological databases. 63.April. DOI: 10.1016/j.ecoinf.2021.101315.
- 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.
- Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghilambor, C. K., Haak, D. C., and Martin, P. R. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, 105.18, 6668–6672. DOI: 10.1073/pnas.0709472105.
- 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.
- Etard, A., Pigot, A. L., and Newbold, T. (2021). Intensive human land uses negatively affect vertebrate functional diversity. *Ecology Letters*, n/a.n/a. DOI: <https://doi.org/10.1111/ele.13926>.
- 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.

- Flynn, D. F., Gogol-Prokurat, M., Nogeire, T., Molinari, N., Richers, B. T., Lin, B. B., Simpson, N., Mayfield, M. M., and DeClerck, F. (2009). Loss of functional diversity under land use intensification across multiple taxa. *Ecology Letters*, 12.1, 22–33. DOI: [10.1111/j.1461-0248.2008.01255.x](https://doi.org/10.1111/j.1461-0248.2008.01255.x).
- Fournier, B., Mouly, A., and Gillet, F. (2016). Multiple assembly rules drive the co-occurrence of orthopteran and plant species in grasslands: Combining network, functional and phylogenetic approaches. *Frontiers in Plant Science*, 7.AUG2016, 1–12. DOI: [10.3389/fpls.2016.01224](https://doi.org/10.3389/fpls.2016.01224).
- Freeman, B. G. and Pennell, M. W. (2021). The latitudinal taxonomy gradient. *Trends in Ecology and Evolution*, 36.9, 778–786. DOI: [10.1016/j.tree.2021.05.003](https://doi.org/10.1016/j.tree.2021.05.003).
- Frishkoff, L. O., Karp, D. S., Flanders, J. R., Zook, J., Hadly, E. A., Daily, G. C., and M'Gonigle, L. K. (2016). Climate change and habitat conversion favour the same species. DOI: [10.1111/ele.12645](https://doi.org/10.1111/ele.12645).
- 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](https://doi.org/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](https://doi.org/10.1038/s41597-019-0059-9).
- Gallmetzer, N. and Schulze, C. H. (2015). Impact of oil palm agriculture on understory amphibians and reptiles: A Mesoamerican perspective. *Global Ecology and Conservation*, 4, 95–109. DOI: [10.1016/j.gecco.2015.05.008](https://doi.org/10.1016/j.gecco.2015.05.008).
- 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](https://doi.org/10.1111/j.1365-2656.2012.01999.x).
- 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](https://doi.org/10.1111/ele.12035).
- 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](https://doi.org/10.1111/geb.12769).
- Guralnick, R. P., Zermoglio, P. F., Wieczorek, J., LaFrance, R., Bloom, D., and Russell, L. (2016). The importance of digitized biocollections as a source of trait data and a new VertNet resource. *Database*. DOI: [10.1093/database/baw158](https://doi.org/10.1093/database/baw158).
- Hatfield, J. H., Harrison, M. L. K., and Banks-Leite, C. (2018). Functional Diversity Metrics: How They Are Affected by Landscape Change and How They Represent Ecosystem Functioning in the Tropics. DOI: [10.1007/s40823-018-0032-x](https://doi.org/10.1007/s40823-018-0032-x).
- Hooper, D. U., Adair, E. C., Cardinale, B. J., Byrnes, J. E., Hungate, B. A., Matulich, K. L., Gonzalez, A., Duffy, J. E., Gamfeldt, L., and Connor, M. I. (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature*. DOI: [10.1038/nature11118](https://doi.org/10.1038/nature11118).
- 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](https://doi.org/10.1146/annurev-ecolsys-112414-054400).
- Hudson, L. N. et al. (2014). The PREDICTS database: A global database of how local terrestrial biodiversity responds to human impacts. *Ecology and Evolution*. DOI: [10.1002/ece3.1303](https://doi.org/10.1002/ece3.1303).

- Hudson, L. N. et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*. DOI: 10.1002/ece3.2579.
- 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.
- Jesse, W. A., Behm, J. E., Helmus, M. R., and Ellers, J. (2018). Human land use promotes the abundance and diversity of exotic species on Caribbean islands. *Global Change Biology*, 24.10, 4784–4796. DOI: 10.1111/gcb.14334.
- 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.
- Johnson, T. F., Isaac, N. J., Paviolo, A., and González-Suárez, M. (2021). Handling missing values in trait data. *Global Ecology and Biogeography*. DOI: 10.1111/geb.13185.
- 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. et al. (2018). Towards global data products of Essential Biodiversity Variables on species traits. DOI: 10.1038/s41559-018-0667-3.
- 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.
- Krauss, J., Bommarco, R., Guardiola, M., Heikkinen, R. K., Helm, A., Kuussaari, M., Lindborg, R., Öckinger, E., Pärtel, M., Pino, J., Pöyry, J., Raatikainen, K. M., Sang, A., Stefanescu, C., Teder, T., Zobel, M., and Steffan-Dewenter, I. (2010). Habitat fragmentation causes immediate and time-delayed biodiversity loss at different trophic levels. *Ecology Letters*. DOI: 10.1111/j.1461-0248.2010.01457.x.
- La Sorte, F. A., Lepczyk, C. A., Aronson, M. F., Goddard, M. A., Hedblom, M., Katti, M., MacGregor-Fors, I., Mörtberg, U., Nilsson, C. H., Warren, P. S., Williams, N. S., and Yang, J. (2018). The phylogenetic and functional diversity of regional breeding bird assemblages is reduced and constricted through urbanization. *Diversity and Distributions*. DOI: 10.1111/ddi.12738.
- Laliberte, E. and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91.1, 299–305. DOI: 10.1890/08-2244.1.
- Laliberté, E., Legendre, P., and Shipley, B. (2015). FD: measuring functional diversity from multiple traits, and other tools for functional ecology. *R Package*.

- 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.
- Le Provost, G., Badenauer, I., Le Bagousse-Pinguet, Y., Clough, Y., Henckel, L., Violle, C., Bretagnolle, V., Roncoroni, M., Manning, P., and Gross, N. (2020). Land-use history impacts functional diversity across multiple trophic groups. *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1910023117.
- Legras, G., Loiseau, N., and Gaertner, J. C. (2018). Functional richness: Overview of indices and underlying concepts. *Acta Oecologica*. DOI: 10.1016/j.actao.2018.02.007.
- Letnic, M., Ritchie, E. G., and Dickman, C. R. (2012). Top predators as biodiversity regulators: The dingo *Canis lupus dingo* as a case study. *Biological Reviews*. DOI: 10.1111/j.1469-185X.2011.00203.x.
- Luck, G. W., Lavorel, S., McIntyre, S., and Lumb, K. (2012). Improving the application of vertebrate trait-based frameworks to the study of ecosystem services. *Journal of Animal Ecology*. DOI: 10.1111/j.1365-2656.2012.01974.x.
- Magioli, M., Ferraz, K. M. P. M. d. B., Chiarello, A. G., Galetti, M., Setz, E. Z. F., Paglia, A. P., Abrego, N., Ribeiro, M. C., and Ovaskainen, O. (2021). Land-use changes lead to functional loss of terrestrial mammals in a Neotropical rainforest. *Perspectives in Ecology and Conservation*, 19.2, 161–170. DOI: 10.1016/j.pecon.2021.02.006.
- Májeková, M., Paal, T., Plowman, N. S., Bryndová, M., Kasari, L., Norberg, A., Weiss, M., Bishop, T. R., Luke, S. H., Sam, K., Le Bagousse-Pinguet, Y., Lepš, J., Götzenberger, L., and De Bello, F. (2016). Evaluating Functional diversity: Missing trait data and the importance of species abundance structure and data transformation. *PLoS ONE*. DOI: 10.1371/journal.pone.0149270.
- Marcacci, G., Westphal, C., Wenzel, A., Raj, V., Nölke, N., Tscharntke, T., and Grass, I. (2021). Taxonomic and functional homogenization of farmland birds along an urbanization gradient in a tropical megacity. *Global Change Biology*, June, 1–15. DOI: 10.1111/gcb.15755.
- 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.
- Martins, I. S. and Pereira, H. M. (2017). Improving extinction projections across scales and habitats using the countryside species-area relationship. *Scientific Reports*. DOI: 10.1038/s41598-017-13059-y.
- Matuoka, M. A., Benchimol, M., Almeida-Rocha, J. M. de, and Morante-Filho, J. C. (2020). Effects of anthropogenic disturbances on bird functional diversity: A global meta-analysis. *Ecological Indicators*. DOI: 10.1016/j.ecolind.2020.106471.
- Maxwell, S. L., Fuller, R. A., Brooks, T. M., and Watson, J. E. M. (2016). Biodiversity: The ravages of guns, nets and bulldozers. *Nature*. DOI: 10.1038/536143a.
- 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.
- Menegotto, A. and Rangel, T. F. (2018). Mapping knowledge gaps in marine diversity reveals a latitudinal gradient of missing species richness. *Nature Communications*. DOI: 10.1038/s41467-018-07217-7.
- Millard, J., Outhwaite, C. L., Kinnersley, R., Freeman, R., Gregory, R. D., Adedoja, O., Gavini, S., Kioko, E., Kuhlmann, M., Ollerton, J., Ren, Z.-x., and Newbold, T. (2021). Global effects of land-use intensity on local pollinator biodiversity. *Nature Communications*, 1–11. DOI: 10.1038/s41467-021-23228-3.
- Mouillot, D., Loiseau, N., Grenié, M., Algar, A. C., Allegra, M., Cadotte, M. W., Casajus, N., Denelle, P., Guégan, M., Maire, A., Maitner, B., McGill, B. J., McLean, M., Mouquet, N., Munoz, F., Thuiller, W., Villéger, S., Violle, C., and Auber, A. (2021). The dimensionality and structure of species trait spaces. *Ecology Letters*, March, 1–22. DOI: 10.1111/ele.13778.
- Munteanu, C., Kamp, J., Nita, M. D., Klein, N., Kraemer, B. M., Müller, D., Koskina, A., Prishchepov, A. V., and Kuemmerle, T. (2020). Cold War spy satellite images reveal long-term declines of a philopatric keystone species in response to cropland expansion. *Proceedings of the Royal Society B: Biological Sciences*, 287.1927. DOI: 10.1098/rspb.2019.2897.
- 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., Hudson, L. N., Contu, S., Hill, S. L., Beck, J., Liu, Y., Meyer, C., Phillips, H. R., Scharlemann, J. P., and Purvis, A. (2018). Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide. *PLoS Biology*. DOI: 10.1371/journal.pbio.2006841.
- Newbold, T., Oppenheimer, P., Etard, A., and Williams, J. J. (2020). Tropical and Mediterranean biodiversity is disproportionately sensitive to land-use and climate change. *Nature Ecology and Evolution*. DOI: 10.1038/s41559-020-01303-0.
- 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.
- Newbold, T. et al. (2015). Global effects of land use on local terrestrial biodiversity. *Nature*. DOI: 10.1038/nature14324.
- 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.
- Nowakowski, A. J., Thompson, M. E., Donnelly, M. A., and Todd, B. D. (2017). Amphibian sensitivity to habitat modification is associated with population trends and species traits. *Global Ecology and Biogeography*. DOI: 10.1111/geb.12571.

- 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.
- Oliver, T. H., Heard, M. S., Isaac, N. J., Roy, D. B., Procter, D., Eigenbrod, F., Freckleton, R., Hector, A., Orme, C. D. L., Petchey, O. L., Proença, V., Raffaelli, D., Suttle, K. B., Mace, G. M., Martín-López, B., Woodcock, B. A., and Bullock, J. M. (2015). Biodiversity and Resilience of Ecosystem Functions. DOI: 10.1016/j.tree.2015.08.009.
- Pacifci, 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.
- Pakeman, R. J. (2014). Functional trait metrics are sensitive to the completeness of the species' trait data? *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.12136.
- Paradis, E., Baillie, S. R., Sutherland, W. J., and Gregory, R. D. (1998). Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology*, 67.4, 518–536. DOI: 10.1046/j.1365-2656.1998.00215.x.
- 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.
- Penone, C., Davidson, A. D., Shoemaker, K. T., Di Marco, M., Rondinini, C., Brooks, T. M., Young, B. E., Graham, C. H., and Costa, G. C. (2014). Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.12232.
- Perronne, R., Munoz, F., Borgy, B., Reboud, X., and Gaba, S. (2017). How to design trait-based analyses of community assembly mechanisms: Insights and guidelines from a literature review. DOI: 10.1016/j.ppees.2017.01.004.
- Peugh, J. L. and Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. DOI: 10.3102/00346543074004525.
- Purvis, A., Gittleman, J. L., Cowlishaw, G., and Mace, G. M. (2000). Predicting extinction risk in declining species. *Proceedings of the Royal Society B: Biological Sciences*. DOI: 10.1098/rspb.2000.1234.
- Pyšek, P., Bacher, S., Chytrý, M., Jarošík, V., Wild, J., Celesti-Grapow, L., Gassó, N., Kenis, M., Lambdon, P. W., Nentwig, W., Pergl, J., Roques, A., Sádlo, J., Solarz, W., Vilà, M., and Hulme, P. E. (2010). Contrasting patterns in the invasions of European terrestrial and freshwater habitats by alien plants, insects and vertebrates. *Global Ecology and Biogeography*, 19.3, 317–331. DOI: 10.1111/j.1466-8238.2009.00514.x.
- Quesnelle, P. E., Lindsay, K. E., and Fahrig, L. (2014). Low reproductive rate predicts species sensitivity to habitat loss: A meta-analysis of wetland vertebrates. *PLoS ONE*. DOI: 10.1371/journal.pone.0090926.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Ramírez-Bautista, A., Thorne, J. H., Schwartz, M. W., and Williams, J. N. (2020). Trait-based climate vulnerability of native rodents in southwestern Mexico. *Ecology and Evolution*. DOI: 10.1002/ece3.6323.

- Ratto, F., Simmons, B. I., Spake, R., Zamora-Gutierrez, V., MacDonald, M. A., Merriman, J. C., Tremlett, C. J., Poppy, G. M., Peh, K. S., and Dicks, L. V. (2018). Global importance of vertebrate pollinators for plant reproductive success: a meta-analysis. DOI: [10.1002/fee.1763](https://doi.org/10.1002/fee.1763).
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- Riemann, J. C., Ndriantsoa, S. H., Rödel, M. O., and Glos, J. (2017). Functional diversity in a fragmented landscape — Habitat alterations affect functional trait composition of frog assemblages in Madagascar. *Global Ecology and Conservation*, 10, 173–183. DOI: [10.1016/j.gecco.2017.03.005](https://doi.org/10.1016/j.gecco.2017.03.005).
- Ripple, W. J., Wolf, C., Newsome, T. M., Hoffmann, M., Wirsing, A. J., and McCauley, D. J. (2017). Extinction risk is most acute for the world's largest and smallest vertebrates. *Proceedings of the National Academy of Sciences*. DOI: [10.1073/pnas.1702078114](https://doi.org/10.1073/pnas.1702078114).
- 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](https://doi.org/10.1038/s41559-017-0332-2).
- Salo, P., Banks, P. B., Dickman, C. R., and Korpimäki, E. (2010). Predator manipulation experiments: Impacts on populations of terrestrial vertebrate prey. *Ecological Monographs*. DOI: [10.1890/09-1260.1](https://doi.org/10.1890/09-1260.1).
- Sanchez-Ortiz, K., Taylor, K. J., Palma, A. de, König, C., Pyšek, P., Weigelt, P., Essl, F., and Purvis, A. (2019). Effects of land-use change and related pressures on alien and native subsets of island communities. DOI: [10.1101/2019.12.16.878041](https://doi.org/10.1101/2019.12.16.878041).
- 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](https://doi.org/10.1111/1749-4877.12268).
- Sayol, F., Sol, D., and Pigot, A. L. (2020). Brain Size and Life History Interact to Predict Urban Tolerance in Birds. *Frontiers in Ecology and Evolution*, 8, 58. DOI: [10.3389/fevo.2020.00058](https://doi.org/10.3389/fevo.2020.00058).
- 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](https://doi.org/10.1111/geb.12244).
- Schielzeth, H., Dingemanse, N. J., Nakagawa, S., Westneat, D. F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N. A., Garamszegi, L. Z., and Araya-Ajoy, Y. G. (2020). Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*, 11.9, 1141–1152. DOI: [10.1111/2041-210X.13434](https://doi.org/10.1111/2041-210X.13434).
- Schloss, C. A., Nunez, T. A., and Lawler, J. J. (2012). Dispersal will limit ability of mammals to track climate change in the Western Hemisphere. *Proceedings of the National Academy of Sciences*. DOI: [10.1073/pnas.1116791109](https://doi.org/10.1073/pnas.1116791109).
- 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](https://doi.org/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](https://doi.org/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.
- Sol, D., Trisos, C., Múrria, C., Jeliazkov, A., González-Lagos, C., Pigot, A. L., Ricotta, C., Swan, C. M., Tobias, J. A., and Pavoine, S. (2020). The worldwide impact of urbanisation on avian functional diversity. *Ecology Letters*, 23.6, 962–972. DOI: 10.1111/ele.13495.
- 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.
- Stekhoven, D. J. (2016). Nonparametric Missing Value Imputation using Random Forest. *R Package version 1.4*. DOI: 10.1093/bioinformatics/btr597.
- Stekhoven, D. J. and Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. DOI: 10.1093/bioinformatics/btr597.
- Stephens, L. et al. (2019). Archaeological assessment reveals Earth's early transformation through land use. *Science*, 365.6456, 897 LP –902. DOI: 10.1126/science.aax1192.
- Sunday, J. M., Bates, A. E., Kearney, M. R., Colwell, R. K., Dulvy, N. K., Longino, J. T., and Huey, R. B. (2014). Thermal-safety margins and the necessity of thermoregulatory behavior across latitude and elevation. *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1316145111.
- Sutherland, G. D., Harestad, A. S., Price, K., and Lertzman, K. P. (2000). Scaling of natal dispersal distances in terrestrial birds and mammals. *Ecology and Society*, 4.1. DOI: 10.5751/es-00184-040116.
- Tilman, D. and Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature*. DOI: 10.1038/367363a0.
- Tinoco, B. A., Santillán, V. E., and Graham, C. H. (2018). Land use change has stronger effects on functional diversity than taxonomic diversity in tropical Andean hummingbirds. *Ecology and Evolution*. DOI: 10.1002/ece3.3813.
- 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.
- Todd, B. D., Nowakowski, A. J., Rose, J. P., and Price, S. J. (2017). Species traits explaining sensitivity of snakes to human land use estimated from citizen science data. *Biological Conservation*. DOI: 10.1016/j.biocon.2016.12.013.
- 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.

- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*. DOI: 10.1093/bioinformatics/17.6.520.
- Tylianakis, J. M., Didham, R. K., Bascompte, J., and Wardle, D. A. (2008). Global change and species interactions in terrestrial ecosystems. *Ecology Letters*, 11.12, 1351–1363. DOI: 10.1111/j.1461-0248.2008.01250.x.
- United Nations Educational Scientific and Cultural and Organization (2015). *UNESCO Global Science Report: Towards 2030*. Tech. rep.
- Valiente-Banuet, A., Aizen, M. A., Alcántara, J. M., Arroyo, J., Cocucci, A., Galetti, M., García, M. B., García, D., Gómez, J. M., Jordano, P., Medel, R., Navarro, L., Obeso, J. R., Oviedo, R., Ramírez, N., Rey, P. J., Traveset, A., Verdú, M., and Zamora, R. (2015). Beyond species loss: The extinction of ecological interactions in a changing world. *Functional Ecology*, 29.3, 299–307. DOI: 10.1111/1365-2435.12356.
- van Rossum, G. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica (CWI).
- Verberk, W. C. E. P., Noordwijk, C. G. E. van, and Hildrew, A. G. (2013). Delivering on a promise: integrating species traits to transform descriptive community ecology into a predictive science. *Freshwater Science*. DOI: 10.1899/12-092.1.
- 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.
- Villéger, S., Mason, N. W. H., and Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*.
- 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.
- Weiss, K. C. and Ray, C. A. (2019). Unifying functional trait approaches to understand the assemblage of ecological communities: synthesizing taxonomic divides. DOI: 10.1111/ecog.04387.
- Whitmee, S. and Orme, C. D. L. (2013). Predicting dispersal distance in mammals: A trait-based approach. *Journal of Animal Ecology*, 82.1, 211–221. DOI: 10.1111/j.1365-2656.2012.02030.x.
- Williams, J. J. and Newbold, T. (2020). Local climatic changes affect biodiversity responses to land use: A review. DOI: 10.1111/ddi.12999.
- 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.

Zhang, J., Qian, H., Girardello, M., Pellissier, V., Nielsen, S. E., and Svenning, J. C. (2018). Trophic interactions among vertebrate guilds and plants shape global patterns in species diversity. *Proceedings of the Royal Society B: Biological Sciences*. DOI: 10.1098/rspb.2018.0949.

# **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, synonymous 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 not directly related 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 synonymous binomial names from the

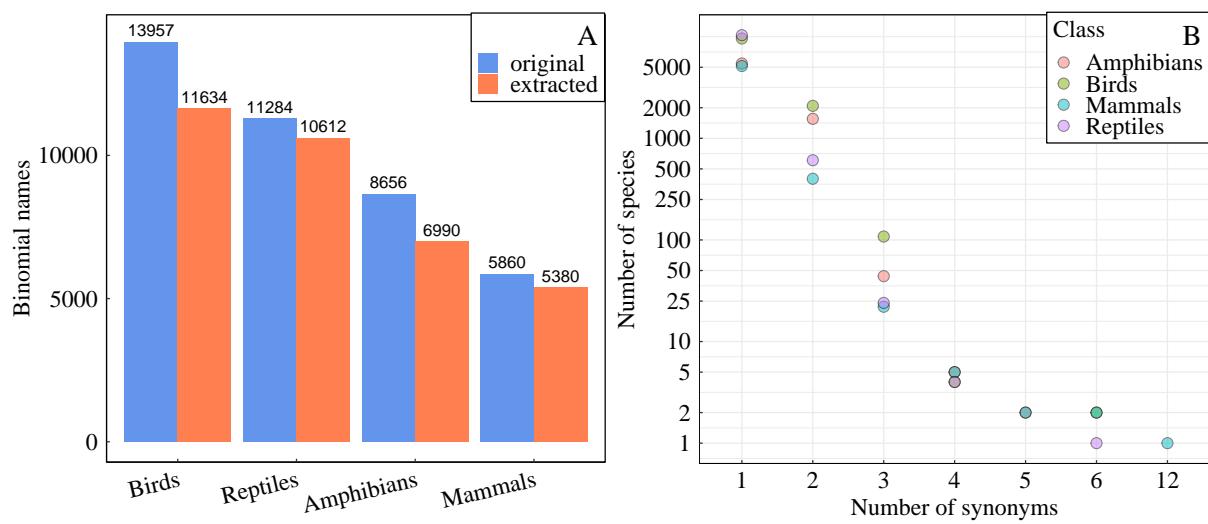
IUCN Red List or from ITIS, using the rredlist (Chamberlain, 2018) and taxize (Chamberlain and Szöcs, 2013) R packages; and (2) identify the status of each name (accepted or not accepted). We started by generating a list of all names featuring in any of the sources. These ‘original’ names were corrected for typos (using gnr\_resolve function in the 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 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 binomial names (14,124 recorded names for birds; 8,743 for amphibians; 6,090 for mammals; and 11,678 for reptiles), and for each we recorded 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). The species presenting the highest number of synonyms was the East African mole rat (*Tachyoryctes splendens*), for which we found 12 synonymous names (Figure S2.1B).

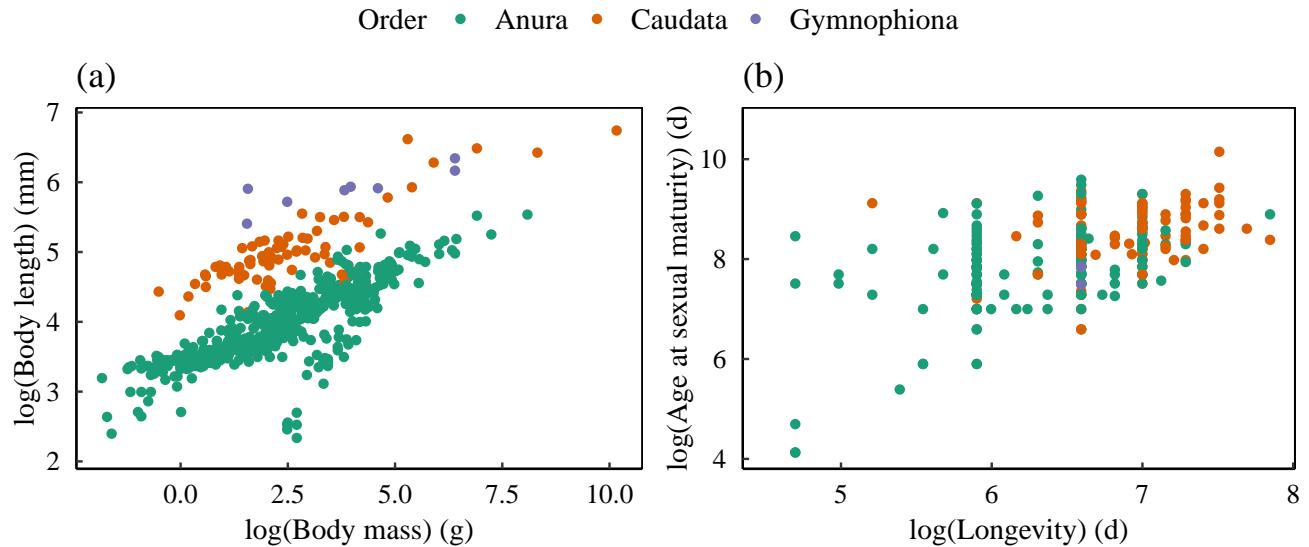
The automated procedure was not perfect, and taxonomic errors are likely to have persisted in the trait datasets. The Red List and the ITIS were not comprehensive taxonomic sources, and for clades with high degrees of synonymy in names, such as reptiles or amphibians, neither the Red List or the ITIS contained enough information. Taxonomy may be further improved by using class-specific sources in an automated procedure.



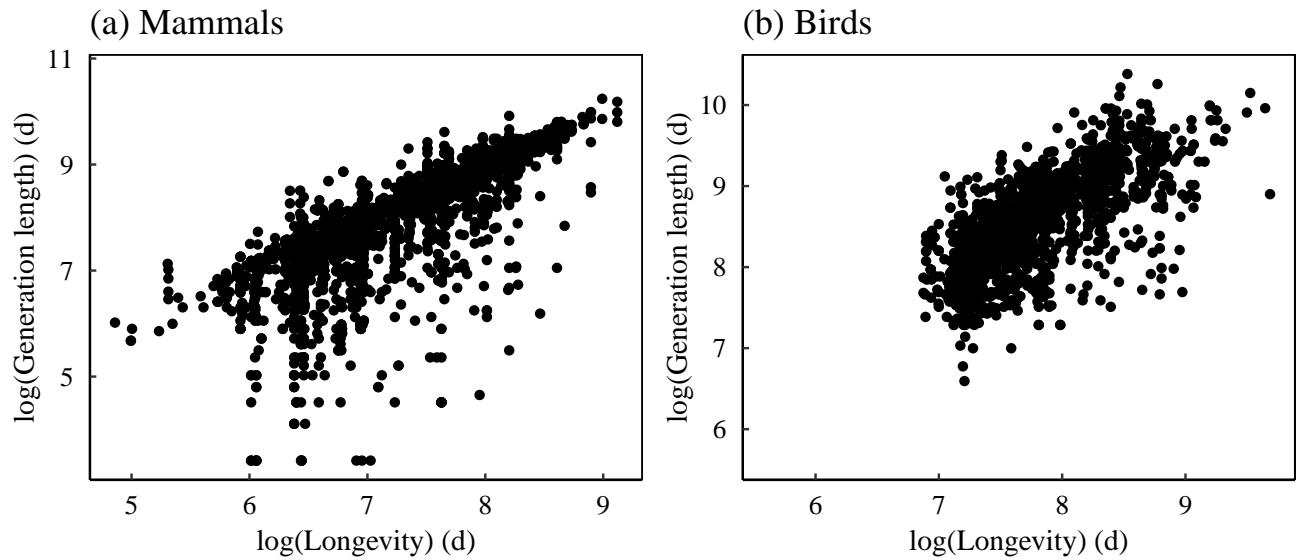
**Figure S2.1: Difference in species number before and after taxonomic correction (A) and distribution of number of synonyms across datasets (B).** (A) shows the number of species (binomial names) extracted from all sources (blue bars), and the number of uniquely identified accepted names (in red). Replacing 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 synonymous 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

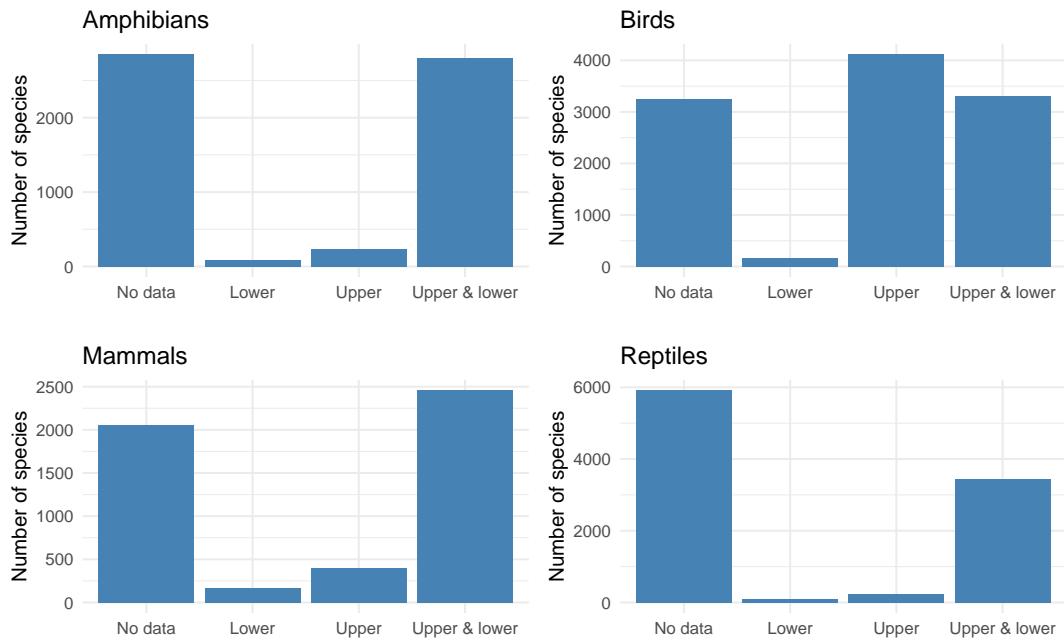


**Figure S2.2:** (a) Body mass versus body length and (b) longevity versus age at sexual maturity in amphibians. The Pearson's correlation coefficient was 0.71 in (a) and 0.55 in (b) (order was not included in these coefficients).

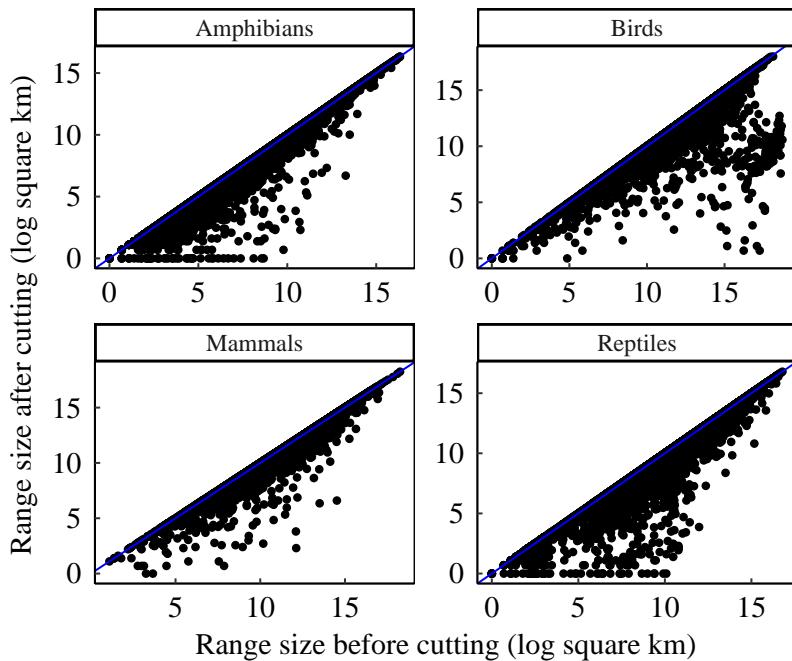


**Figure S2.3:** Generation length versus longevity data in mammals and birds. The Pearson's correlation coefficient was 0.74 in (a) and 0.70 in (b).

### S2.3 Cutting distribution maps by altitudinal limits

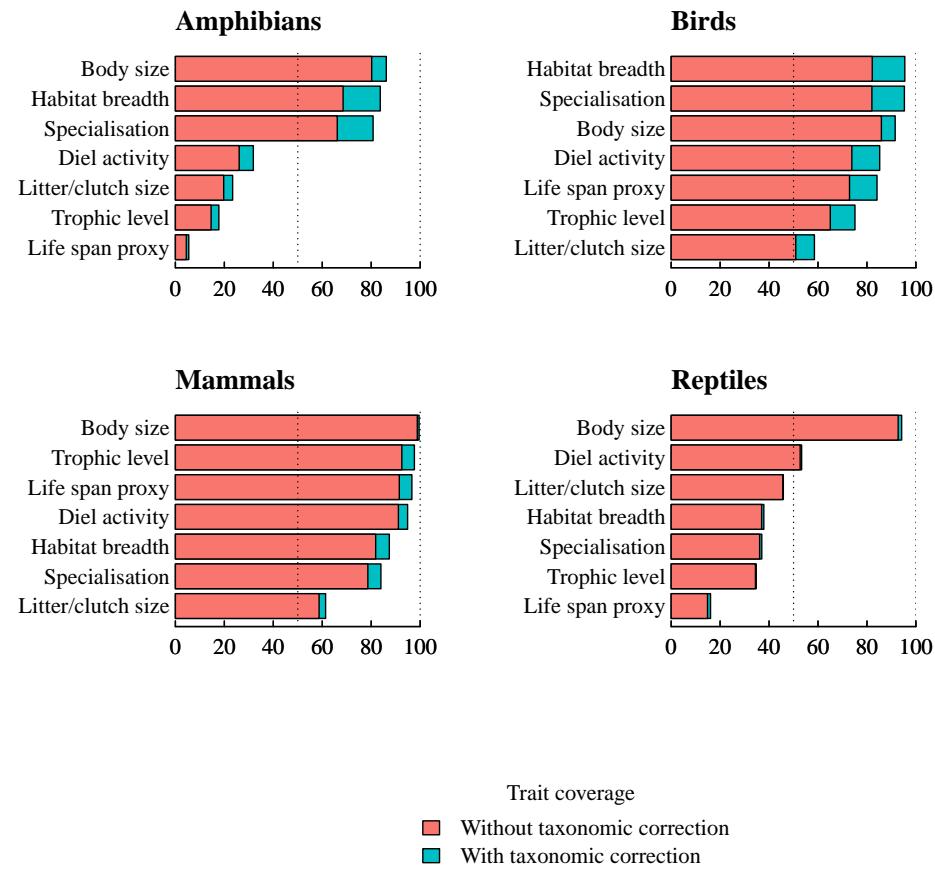


**Figure S2.4: Availability of altitudinal limits across species.** Upper and lower altitudinal limits were extracted from the IUCN Red List.



**Figure S2.5: Range sizes before versus after cutting by altitudinal limits.**

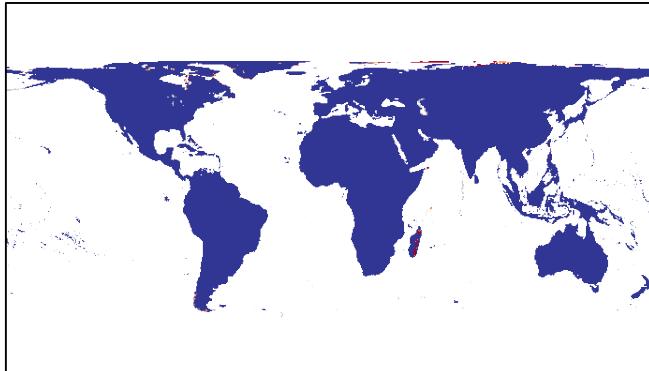
## S2.4 Impact of taxonomic corrections on trait coverage



**Figure S2.6: Trait coverage when no taxonomic correction is applied before matching sources by binomial names, versus when we applied 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)**

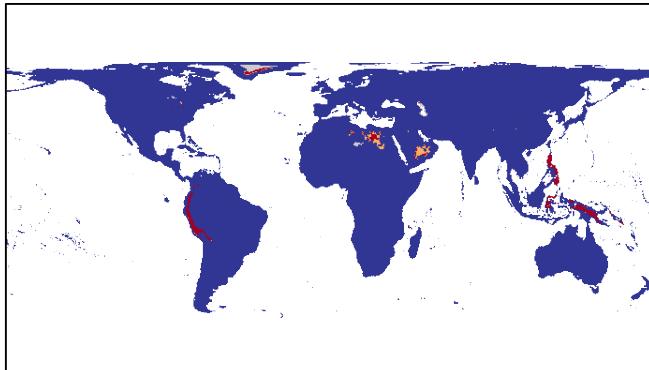
**(a) Mammals**



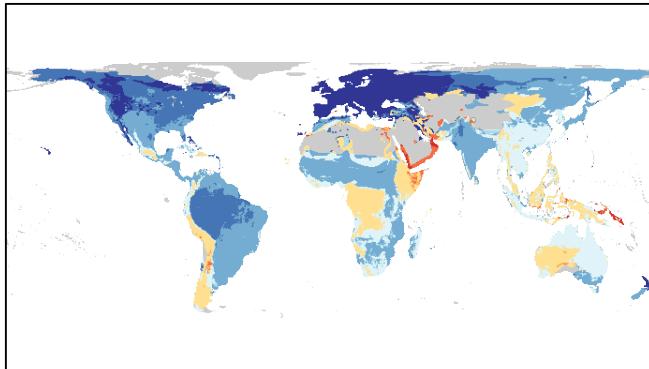
Median completeness  
(birds and mammals):

- NA
- 80 – 90%
- 90 – 91%
- 91 – 92%
- 92 – 93%
- 93 – 94%
- 94 – 95%
- 95 – 96%
- 96 – 97%
- 97 – 98%
- 98 – 99%
- 99 – 100%

**(b) Birds**



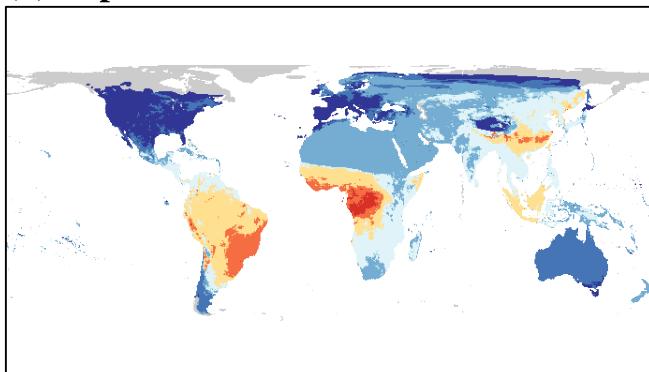
**(c) Amphibians**



Median completeness  
(herptiles):

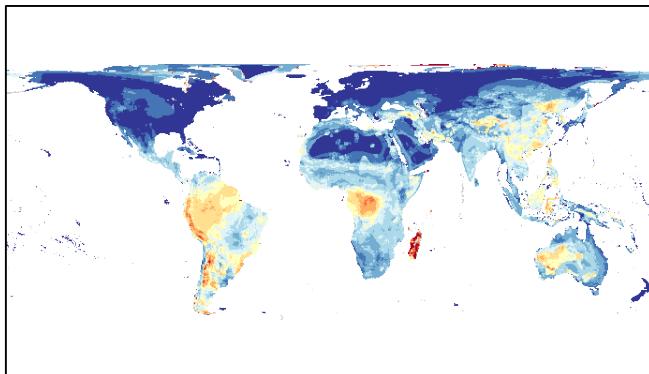
- NA
- 0 – 10%
- 10 – 20%
- 20 – 30%
- 30 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 70%
- 70 – 80%
- 80 – 90%
- 90 – 100%

**(d) Reptiles**



**Figure S2.7: Spatial distribution of assemblage-level median trait completeness in each class.** Note that the color breaks differ for mammals and birds and for herptiles.

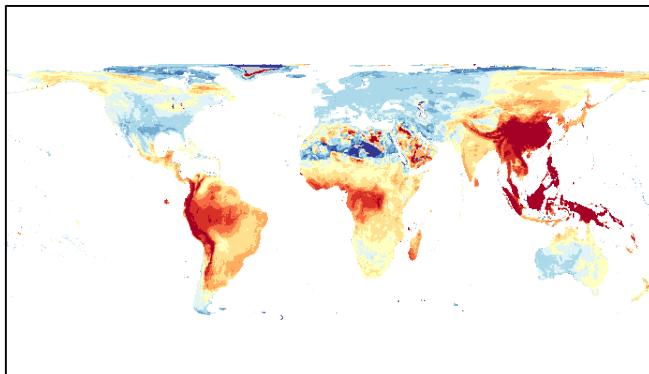
**(a) Mammals**



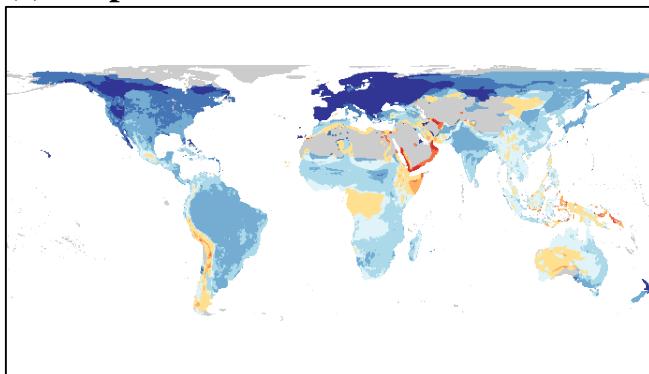
Mean completeness  
(birds and mammals):

- NA
- 80 – 90%
- 90 – 91%
- 91 – 92%
- 92 – 93%
- 93 – 94%
- 94 – 95%
- 95 – 96%
- 96 – 97%
- 97 – 98%
- 98 – 99%
- 99 – 100%

**(b) Birds**



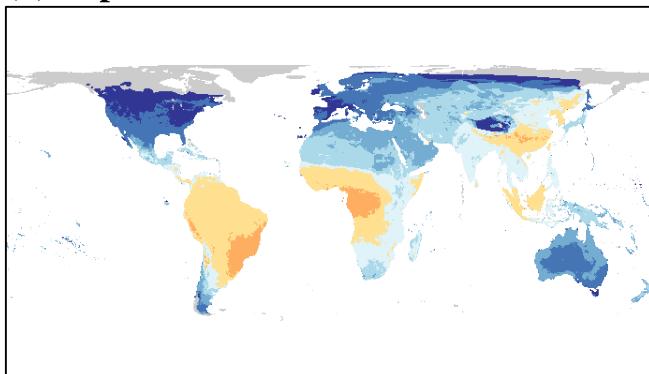
**(c) Amphibians**



Mean completeness  
(herptiles):

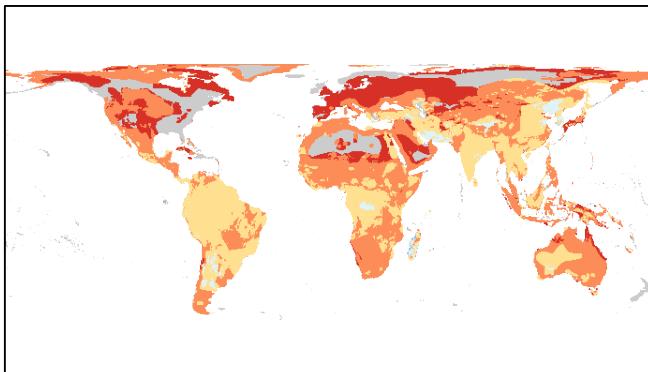
- NA
- 0 – 10%
- 10 – 20%
- 20 – 30%
- 30 – 40%
- 40 – 50%
- 50 – 60%
- 60 – 70%
- 70 – 80%
- 80 – 90%
- 90 – 100%

**(d) Reptiles**



**Figure S2.8:** Spatial distribution of assemblage-level mean trait completeness in each class. Note that the color breaks differ for mammals and birds and for herptiles.

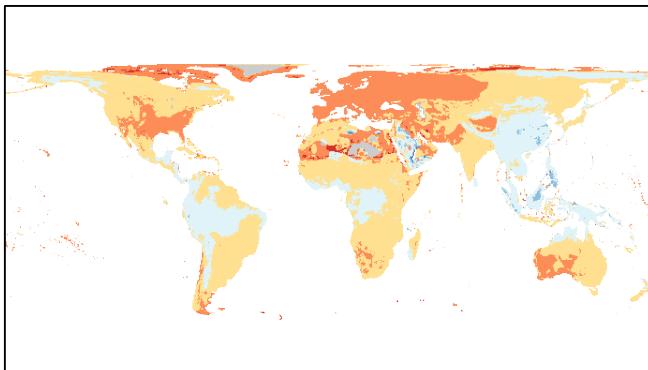
**(a) Mammals**



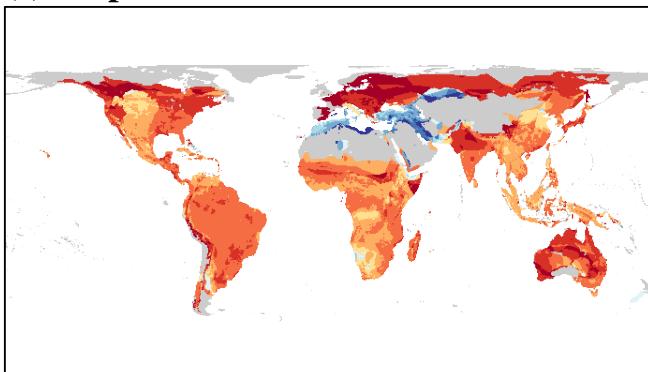
Standard deviation  
of completeness  
(birds and mammals):

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

**(b) Birds**



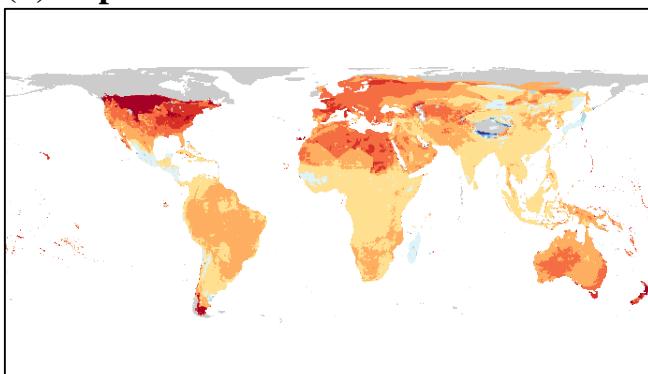
**(c) Amphibians**



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 in each class.** Note that the color breaks differ for mammals and birds and for herptiles.

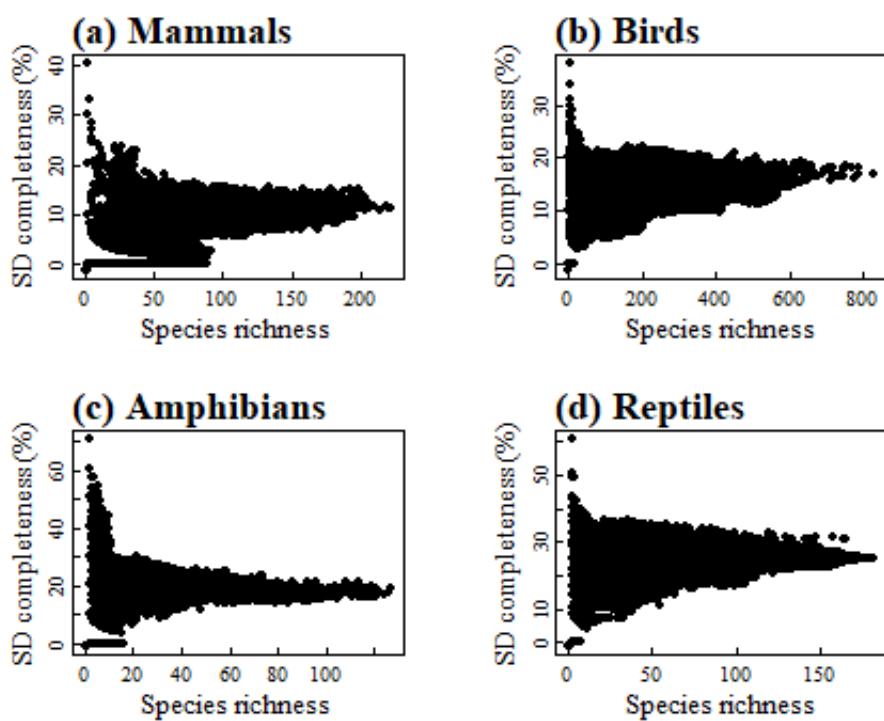
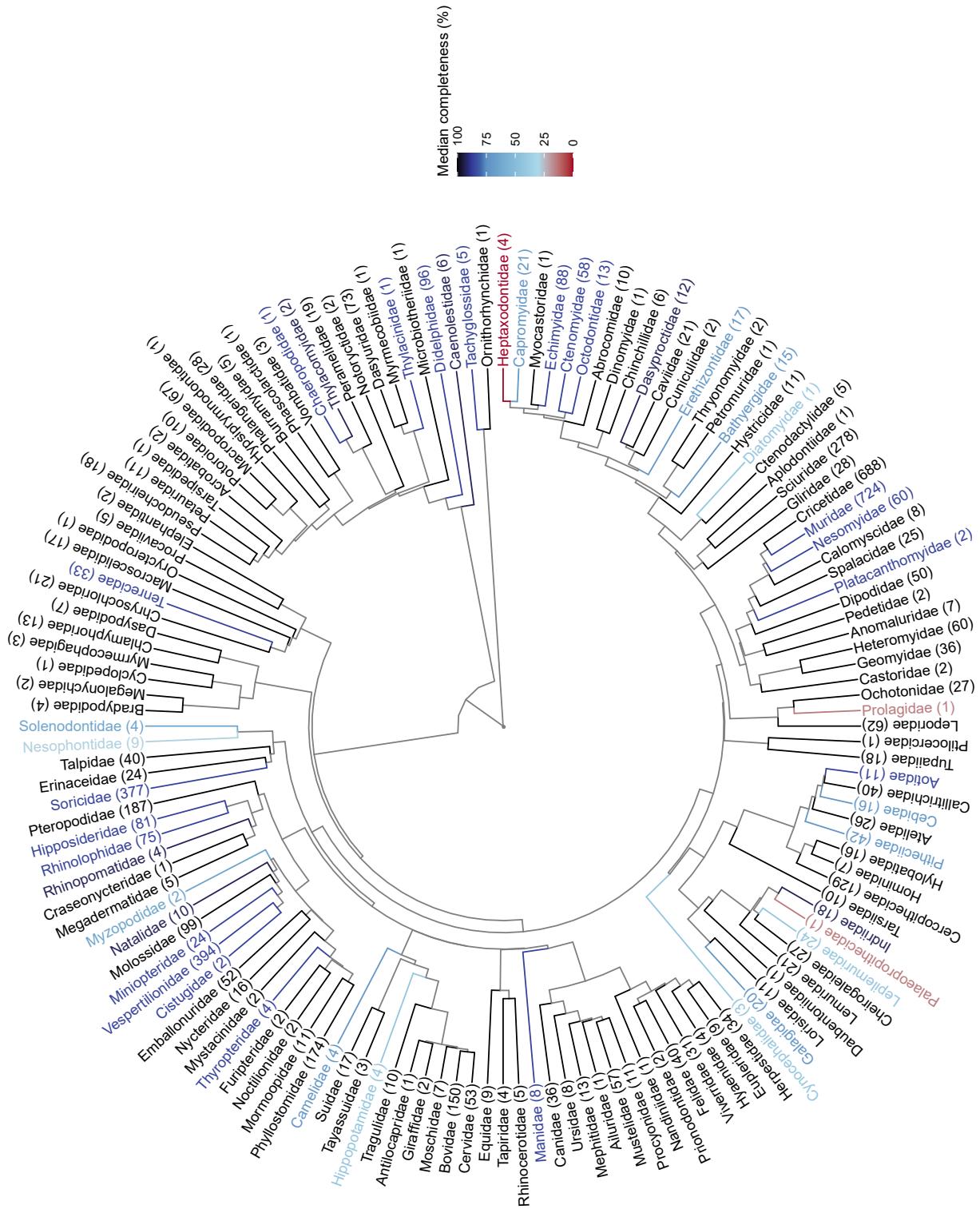
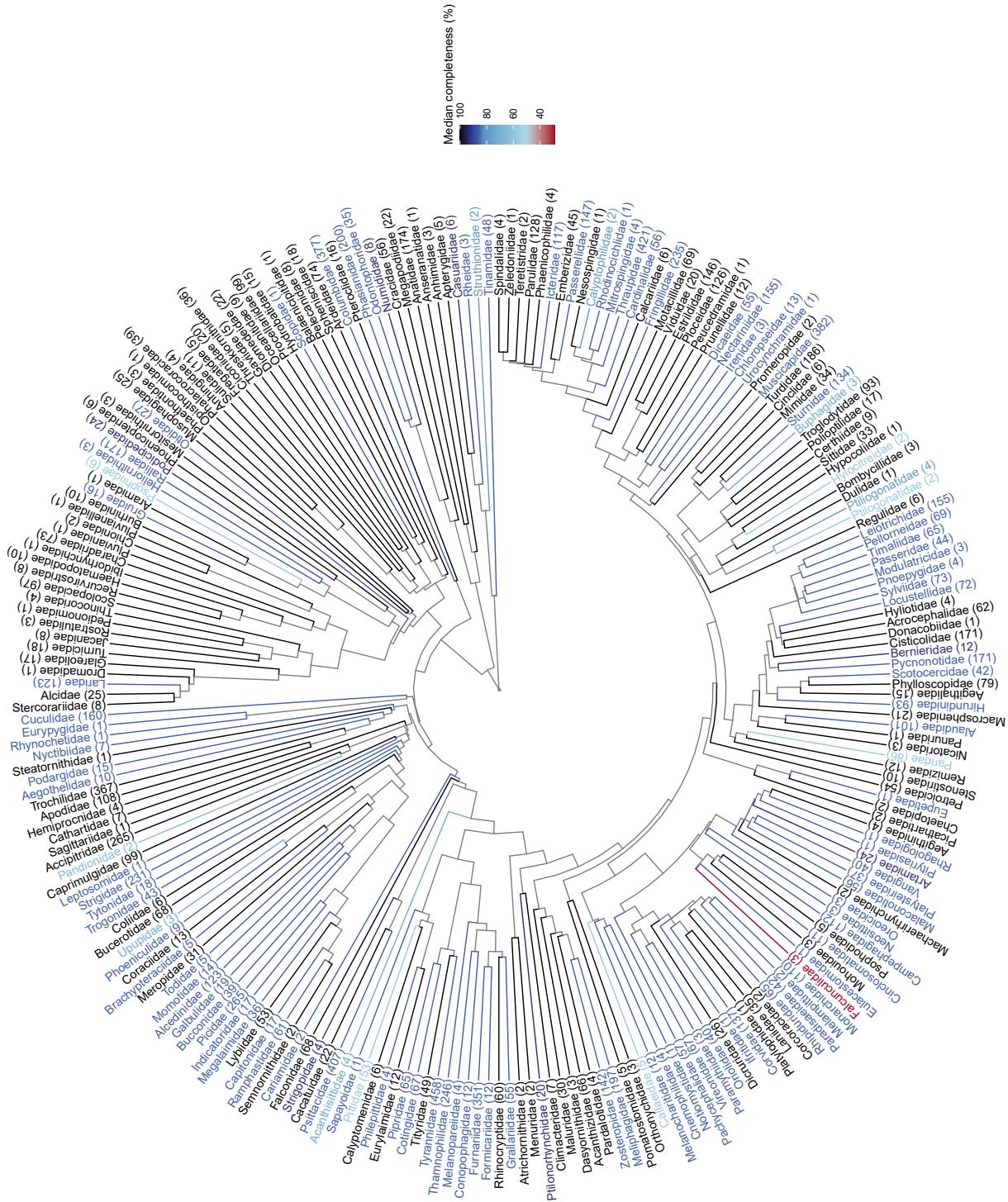


Figure S2.10: Assemblage-level species richness against standard deviation in completeness in each class.

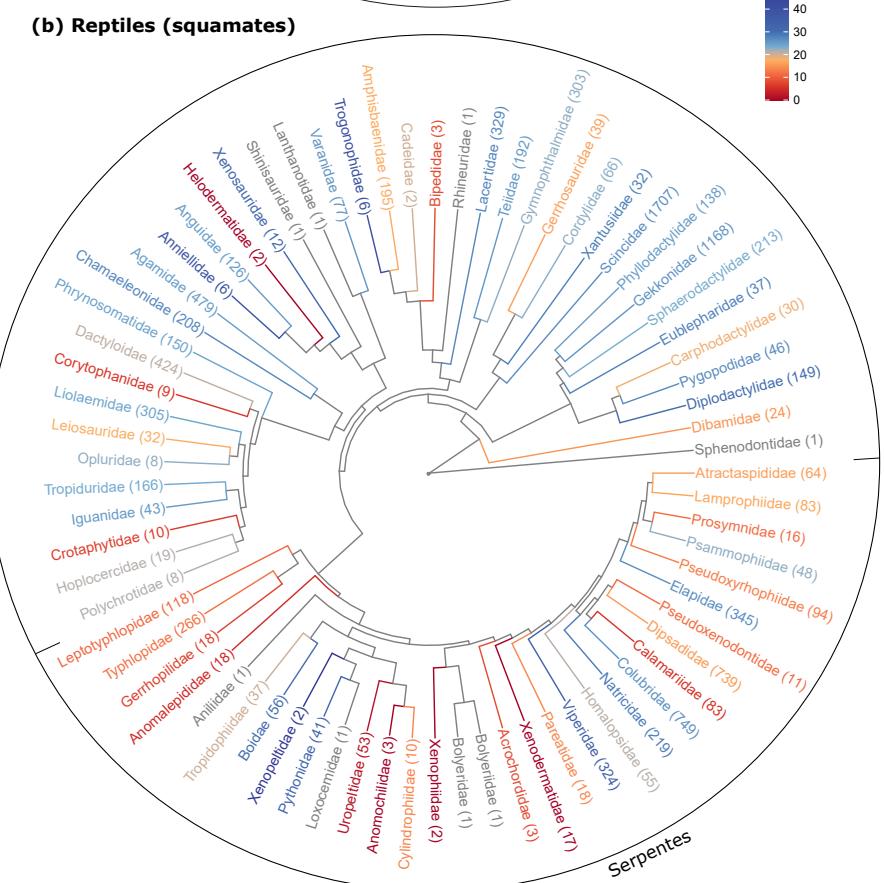
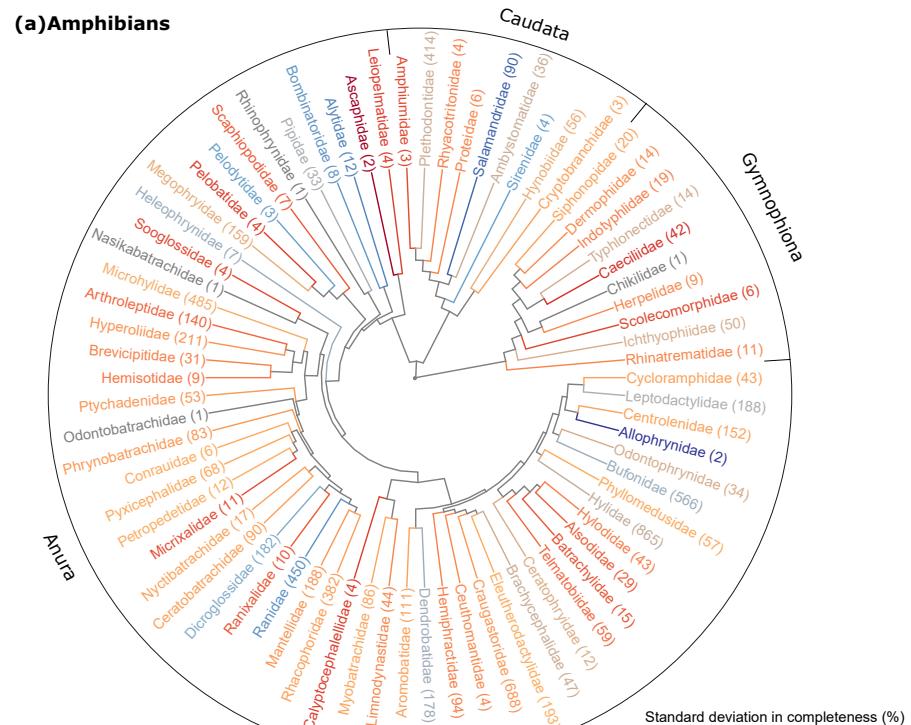
## **S2.6 Phylogenetic patterns in trait completeness**



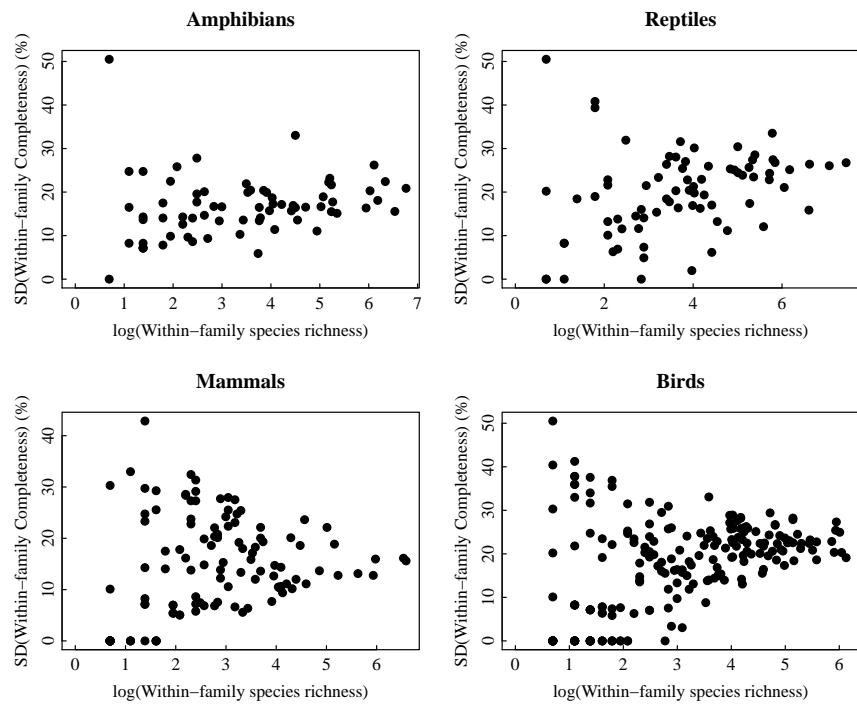
**Figure S2.11:** Within-family median trait completeness in mammals.



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



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



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

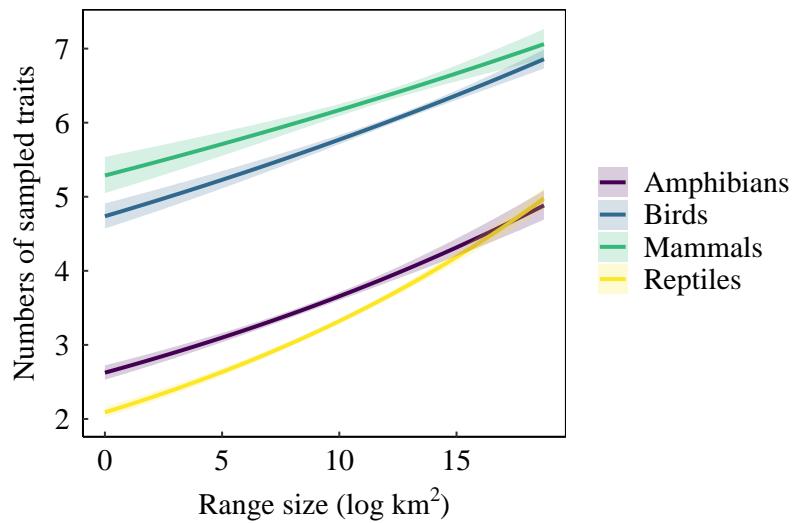
## S2.7 Model coefficients for Range size against Number of sampled traits (Poisson model).

**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 (see Figure 2.5 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.

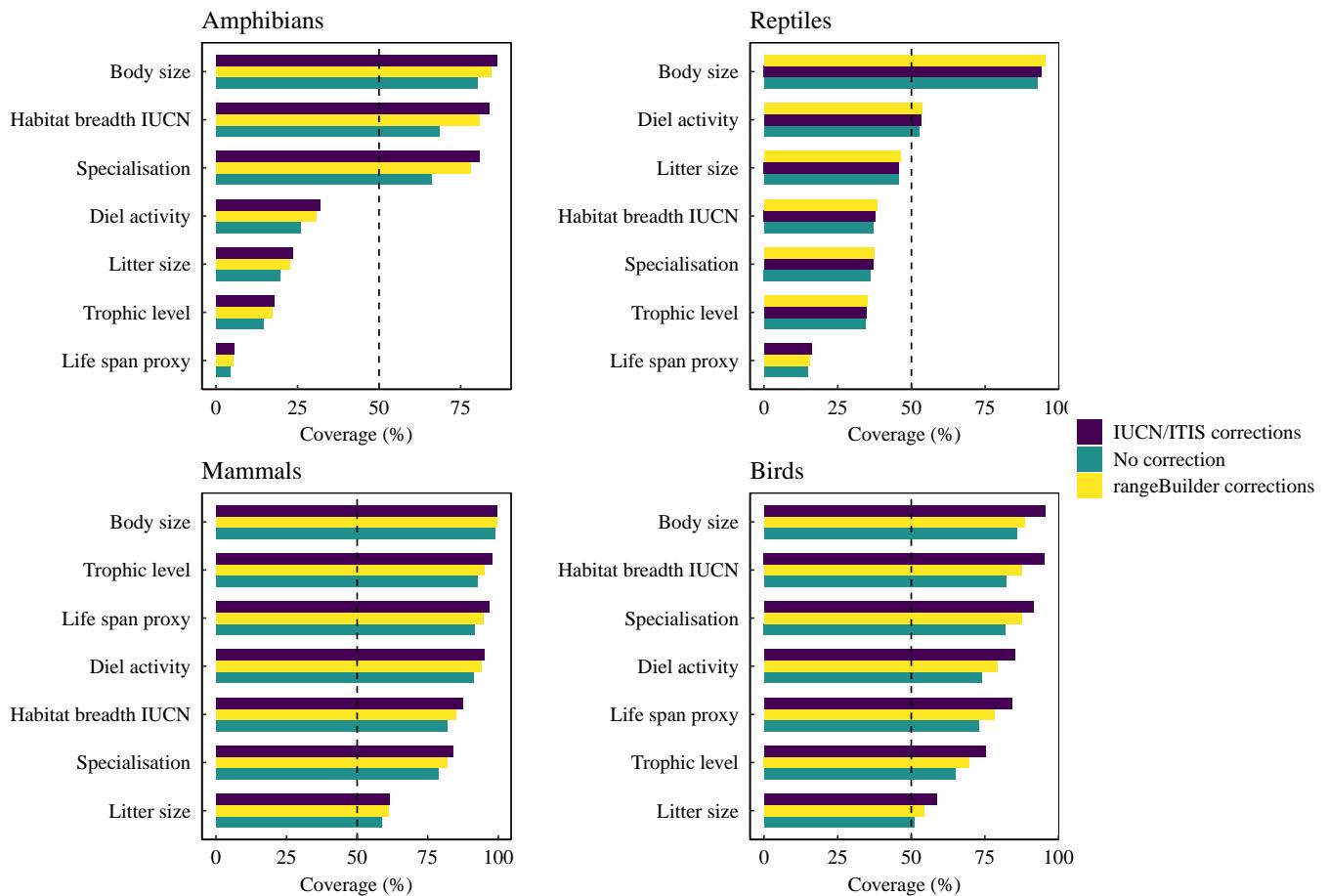
	Estimate	Std. Error	z value	Pr(> z )
Intercept - Realm: Afrotropic	0.0738	0.0064	11.4908	0
log(Species richness)	-0.0025	0.0017	-1.4261	0.1538
Realm: Australasia	-0.0109	0.0095	-1.1453	0.2521
Realm: Indo-Malay	0.0455	0.0119	3.8294	0.0001
Realm: Nearctic	0.0441	0.0082	5.3905	0.000000
Realm: Neotropic	-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: Nearctic	-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):Nearctic	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 (Rabosky et al., 2016), 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.



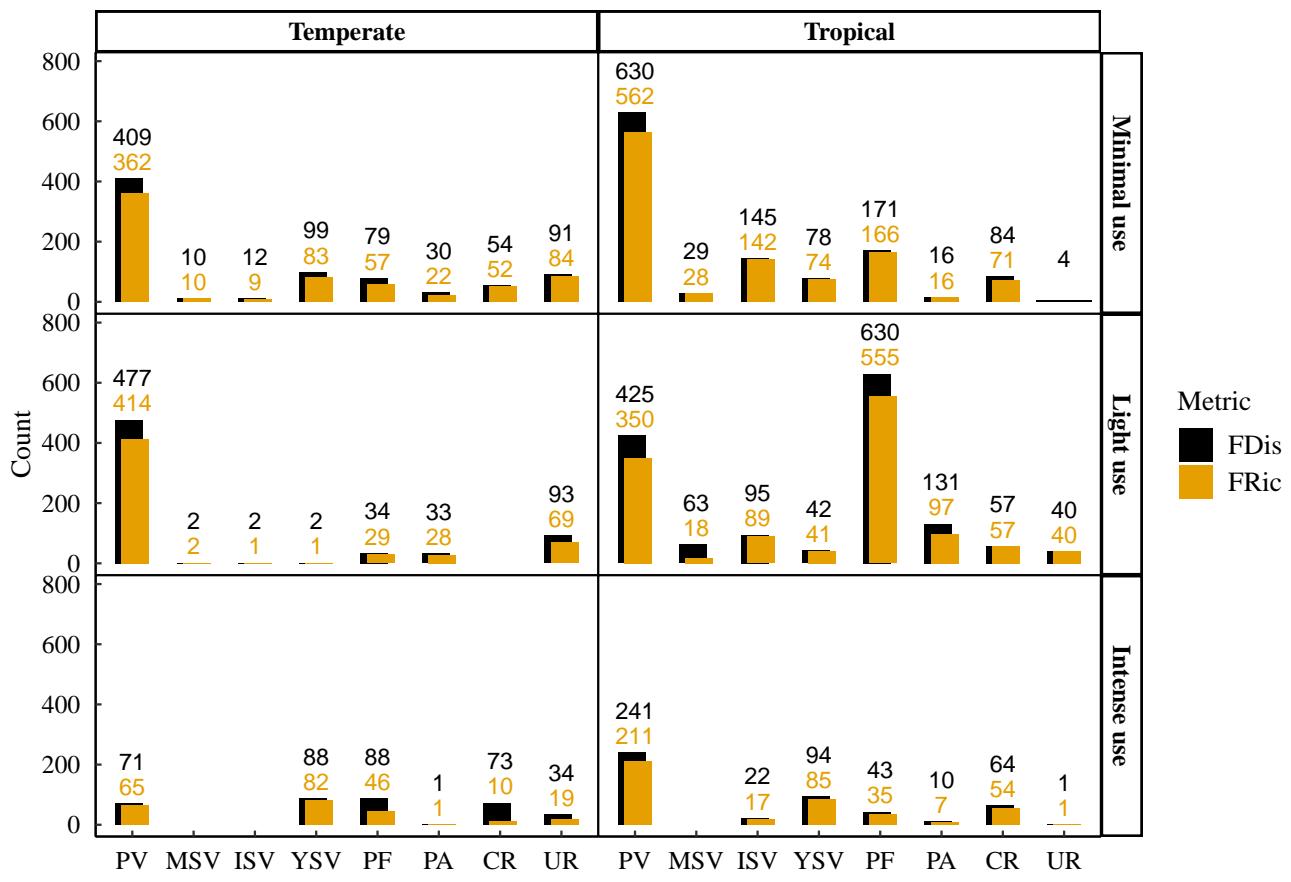
**Figure S2.16: Comparison of trait coverage among datasets corrected for taxonomy in different ways:** using the described procedure (purple bars), datasets corrected using the rangeBuilder package (extraction of synonyms from class-specific sources, yellow bars; Rabosky et al. (2016)) and datasets where no taxonomic correction was applied when matching sources (green bars).

# Appendix 3: Supporting information for Chapter 3

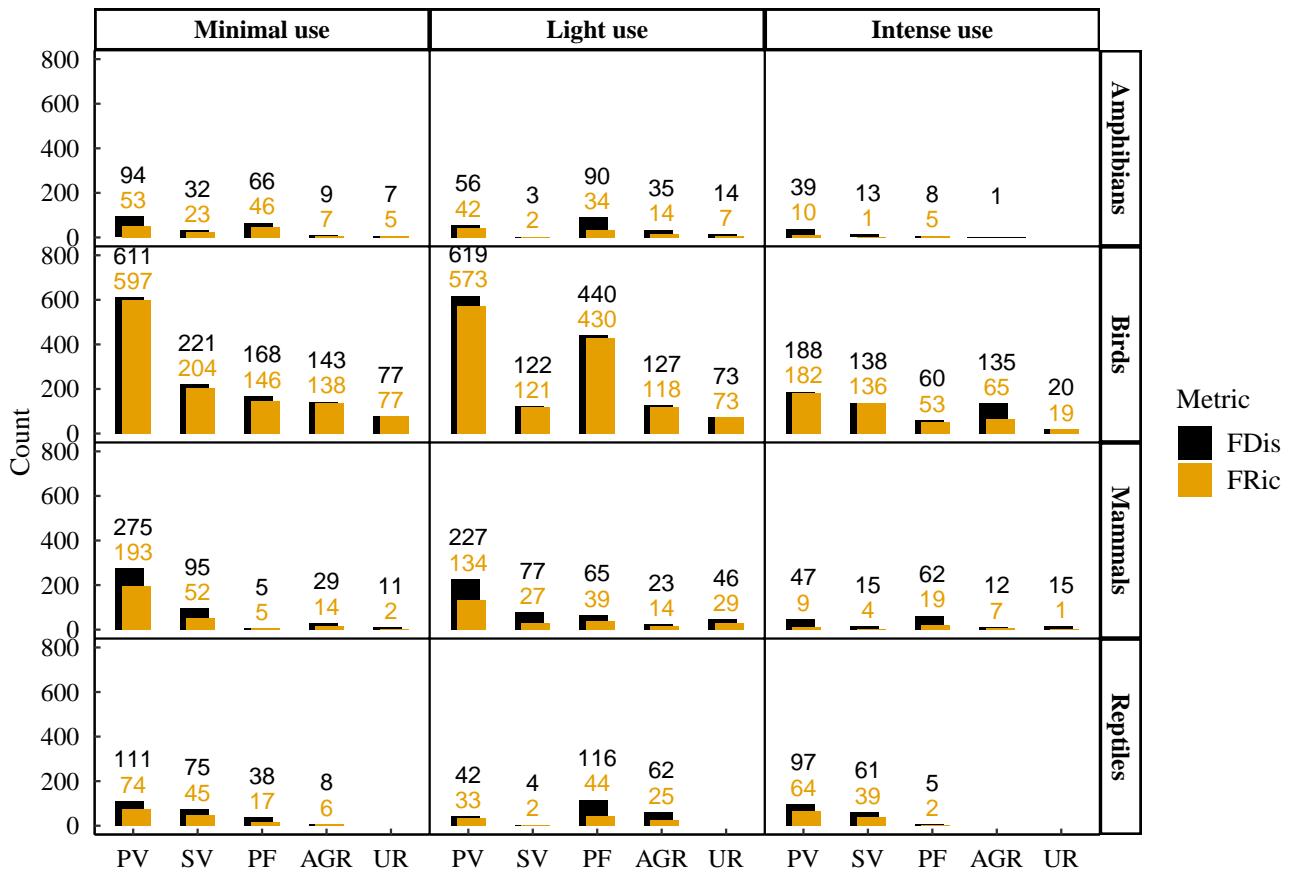
## S3.1 Land-use categories in PREDICTS and sample sizes

**Table S3.1: Land-use categories in the PREDICTS database.** See Hudson et al. (2014, 2017) for more details.

Land-use category	Definition
Primary vegetation	Native vegetation, undisturbed since its establishment under current climatic conditions. No known alterations due to human activities or to extreme natural events.
Mature secondary vegetation	Vegetation recovering after complete destruction of primary vegetation & where succession is near complete – the structure approaches that of primary vegetation.
Intermediate secondary vegetation	Vegetation recovering after complete destruction of primary vegetation at a mid-successional stage.
Young secondary vegetation	Vegetation recovering after complete destruction of primary vegetation at an early successional stage.
Plantation forest	Previously cleared areas planted with crop trees or shrubs grown and harvested for human consumption or for commercial purposes (includes wood, fruit, oil, biofuel, rubber, etc).
Pasture	Areas grazed by livestock, permanently or regularly. Can be improved through cultivation techniques.
Cropland	Previously cleared areas planted with herbaceous crops and harvested for human or animal consumption (including animal feed and crops used in the food industry), or for commercial purposes (e.g., crops grown for the textile industry).
Urban	Previously cleared areas built up by humans. Vegetation is managed for civic or personal purposes.



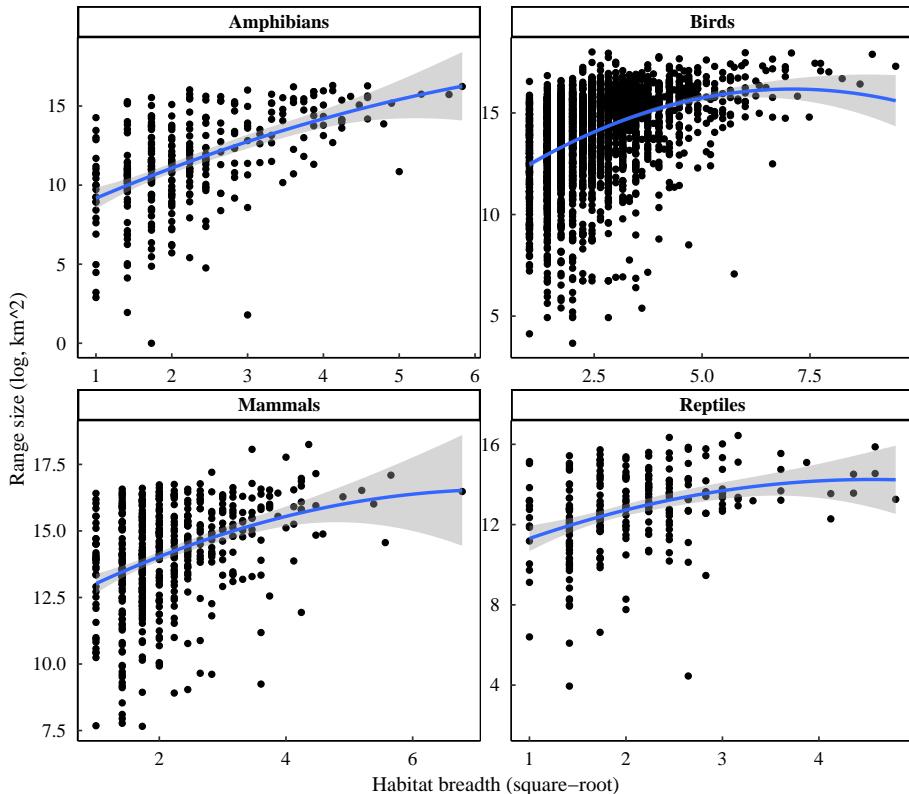
**Figure S3.1: Number of sites in each land use and land-use intensity for which FRic and FDis were calculated, across all species.** The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).



**Figure S3.2: Number of sites in each land use, land-use intensity and for which FRic and FDis were calculated, within each class.** The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).

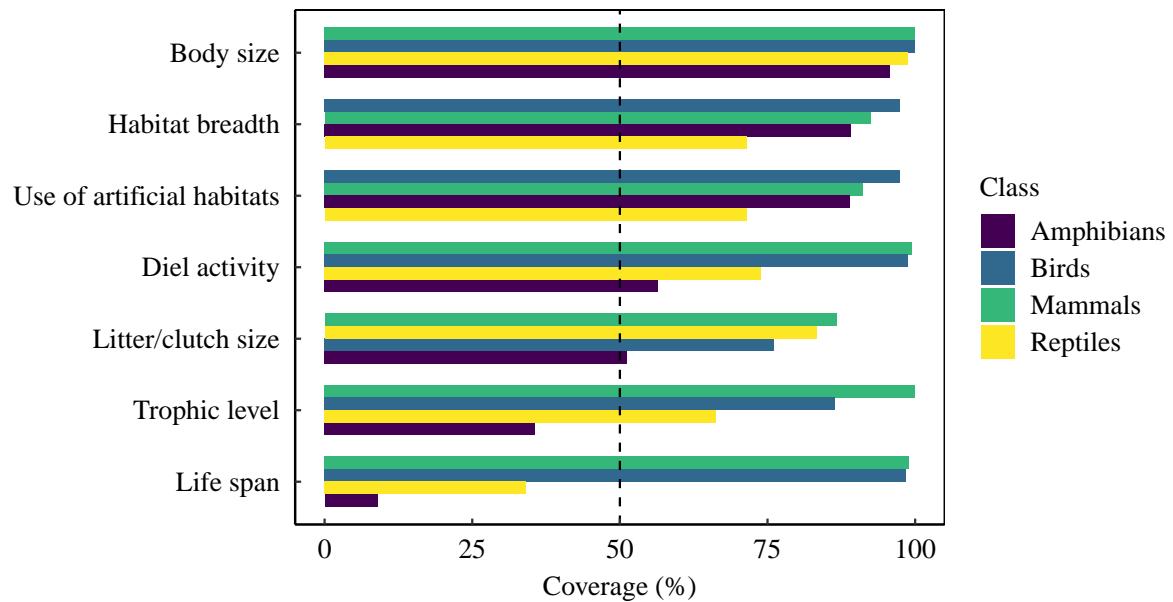
## S3.2 Trait data & imputation of missing trait values

I used the compilation of trait data across terrestrial vertebrates from Chapter 2 (and published in Etard et al. (2020)). I selected seven traits that were available at least for a subset of the species in each class: body mass, trophic level, lifespan, litter/clutch size, diel activity, habitat breadth and use of artificial habitats. These traits related to life-history, habitat specialisation and use of geographical space (for instance, habitat breadth is a significant predictor of geographical range size in all classes, Figure S3.3).

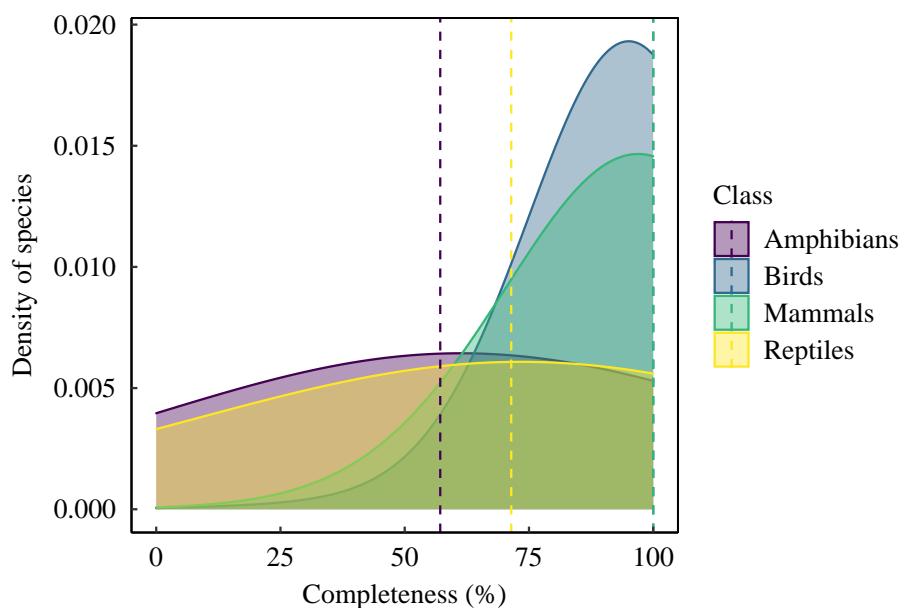


**Figure S3.3: Relationship between habitat breadth and geographical range size across species in each class.** The derivation of geographical range sizes is described in Chapter 2.

Trait-data coverage was highly variable among classes and traits, with important geographical and phylogenetic biases in trait data for reptiles and amphibians (Chapter 2; Etard et al., 2020; Figure S3.4, Figure S3.5). To obtain complete species-trait datasets, I imputed missing trait values. Further, in order to assess the sensitivity of our models to variation in imputed values, I imputed the missing trait values eight independent times. This allowed me to assess the congruence of the model predictions when using the different imputed trait datasets in the analyses.



**Figure S3.4:** Trait coverage for the vertebrate species sampled in the PREDICTS database. For a given trait, coverage is calculated as the percentage of species for which an estimate was available.



**Figure S3.5:** Distribution of trait completeness across the vertebrate species sampled in the PREDICTS database. For a given species, trait completeness is calculated as the proportion of traits for which an estimate was available. Dashed lines represent the median trait completeness.

### S3.2.1 Choice of imputation technique

There exist several imputation techniques (Debastiani et al., 2021; Etard et al., 2020; Johnson et al., 2021; Penone et al., 2014), such as K-nearest neighbour (Troyanskaya et al., 2001), multivariate imputation by chained equations (Buuren and Oudshoorn, 2007), random forest algorithms (implementable in R with missForest; Stekhoven, 2016; Stekhoven and Bühlmann, 2012), and phylogenetic imputations (implementable in R with PhyloPars, Bruggeman et al., 2009). Penone et al. (2014) assessed the performance of these four techniques and showed that missForest and PhyloPars performed better when traits were phylogenetically conserved, and when the species phylogenetic position was included as a predictor of missing trait values. PhyloPars can only handle continuous data, while missForest is compatible with mixed-type (including categorical) data. When no phylogenetic information was included, mice was found to be the best method, with fast imputations of mixed-type data (Penone et al., 2014). Therefore, to assess whether missForest or mice was more appropriate here, I measured the phylogenetic signal in trait data. For continuous traits, I used Pagel's  $\lambda$  (Pagel, 1999), and for categorical traits I used Borges'  $\delta$  (Borges et al., 2018). Strong phylogenetic signal would indicate that traits are phylogenetically conserved, and hence missForest would be the most suited approach for imputing missing trait values, with the inclusion of species' phylogenetic positions as a predictor.

### S3.2.2 Phylogenetic signal in traits

Across all classes, similar traits were used for calculating functional diversity metrics: body mass, litter/clutch size, lifespan (using different proxies in different vertebrate classes: generation length for birds and mammals, longevity for reptiles, and age at sexual maturity for amphibians), trophic level, diel activity, habitat breadth and use of artificial habitats. In addition, I included some class-specific traits for the imputations, as certain class-specific traits could be useful predictors of other traits (such as body length for instance in amphibians (Santini et al., 2018)). Table S3.2 details the traits that were included for the imputations in each class and the phylogenetic signal for each of these traits. Continuous traits were log-10 transformed before assessing Pagel's  $\lambda$  to improve normality. Pagel's  $\lambda$  was estimated using the phylosig function of the phytools package (Revell, 2012), and Borges'  $\delta$  was assessed using code provided by Borges et al. (2018), available at : [https://github.com/mrborges23/delta\\_statistic](https://github.com/mrborges23/delta_statistic). To test for the significance of  $\delta$ , I generated null distributions of  $\delta$  for each categorical trait by randomising trait vectors 50 times, and calculating  $\delta$  for each randomised vector – following the guidelines proposed by Borges et al. (2018). I then tested whether the observed medians were greater than the null distributions using one-sided Wilcoxon rank sum tests.

I used class-specific phylogenies to estimate phylogenetic signal, all downloaded on 13th April 2020. Trees from Faurby et al. (2018, 2020) were used for mammals (downloaded from <https://zenodo.org/record/3690867#.Xyc5wyhKhPZ>). For amphibians, birds and reptiles (squamates only), I downloaded trees from <https://data.vertlife.org/>. Trees were from Jetz et al. (2012) for birds, from Jetz and Pyron (2018) for amphibians and from Tonini et al. (2016) for squamates. For each class, I downloaded a distribution of 1,000 trees, from which I obtained consensus trees to estimate phylogenetic signal (to that end, I used the TreeAnnotator programme of the BEAST software (Bouckaert et al., 2014)).

**Table S3.2: Phylogenetic signal in continuous and categorical traits.** BM: body mass; BL: body length; GL: generation length; MA: age at sexual maturity; ML: maximum longevity; L: longevity; LCS: litter/clutch size; HB: habitat breadth; TL: trophic level; DA: diel activity; UA: use of artificial habitats. Continuous traits were log-10 transformed to improve normality before estimating Pagel's  $\lambda$  – except for habitat breadth which was square-rooted. A star indicates a significant signal (p-value<0.05 for the log-likelihood ratio test in the case of  $\lambda$ ; and a significant difference from the simulated null distribution of  $\delta$  for categorical traits). 'NA' indicates traits that were not considered for a given class. All traits showed significant phylogenetic signal, with signals for BM, BL, L, GL, MA and LCS being particularly strong (above 0.8) across the four classes.

Class	Pagel's $\lambda$								Borges' $\delta$		
	BM	BL	GL	MA	ML	L	LCS	HB	TL	DA	UA
<b>Amphibians</b>	0.98*	0.94*	NA	0.85*	0.82*	NA	0.93*	0.99*	18*	3.4*	4.5*
<b>Birds</b>	0.99*	NA	0.97*	NA	NA	NA	0.95*	0.60*	13*	32e3*	1.8*
<b>Mammals</b>	0.99*	NA	0.97	NA	NA	NA	0.99	0.71	26*	17*	1.3*
<b>Reptiles</b>	1.0*	NA	NA	NA	0.94*	0.98*	1.0*	0.52*	6.3*	6.4*	1.4*

### S3.2.3 Implementation of missForest imputations

As phylogenetic signals were strong in many categorical and continuous traits (Table S3.2), I imputed missing trait values using random forest algorithms, as implemented in R with missForest (Stekhoven, 2016; Stekhoven and Bühlmann, 2012). Another advantage of missForest was that, being a non-parametric approach, no prior assumption about data distribution was required. The data were therefore not transformed prior to imputations. In addition, Penone et al. (2014) showed that including phylogenetic information did not decrease the accuracy of imputations for traits that were less phylogenetically conserved, such as habitat breadth in this work.

Phylogenetic relationships were included as additional predictors in the form of phylogenetic eigenvectors (Diniz-Filho et al., 2012), extracted from the phylogenies using the PVR package (Santos, 2018). Following Penone et al. (2014), I included the first ten phylogenetic eigenvectors as additional predictors of missing trait values in each class, enough to minimise imputation error. As not all species were represented in the phylogenies, I also added taxonomic order as a predictor for all species. All traits in Table S3.2 were included in the imputations. Tuning parameters of missForest were set to ten maximum iterations and to one

hundred trees grown in each forest.

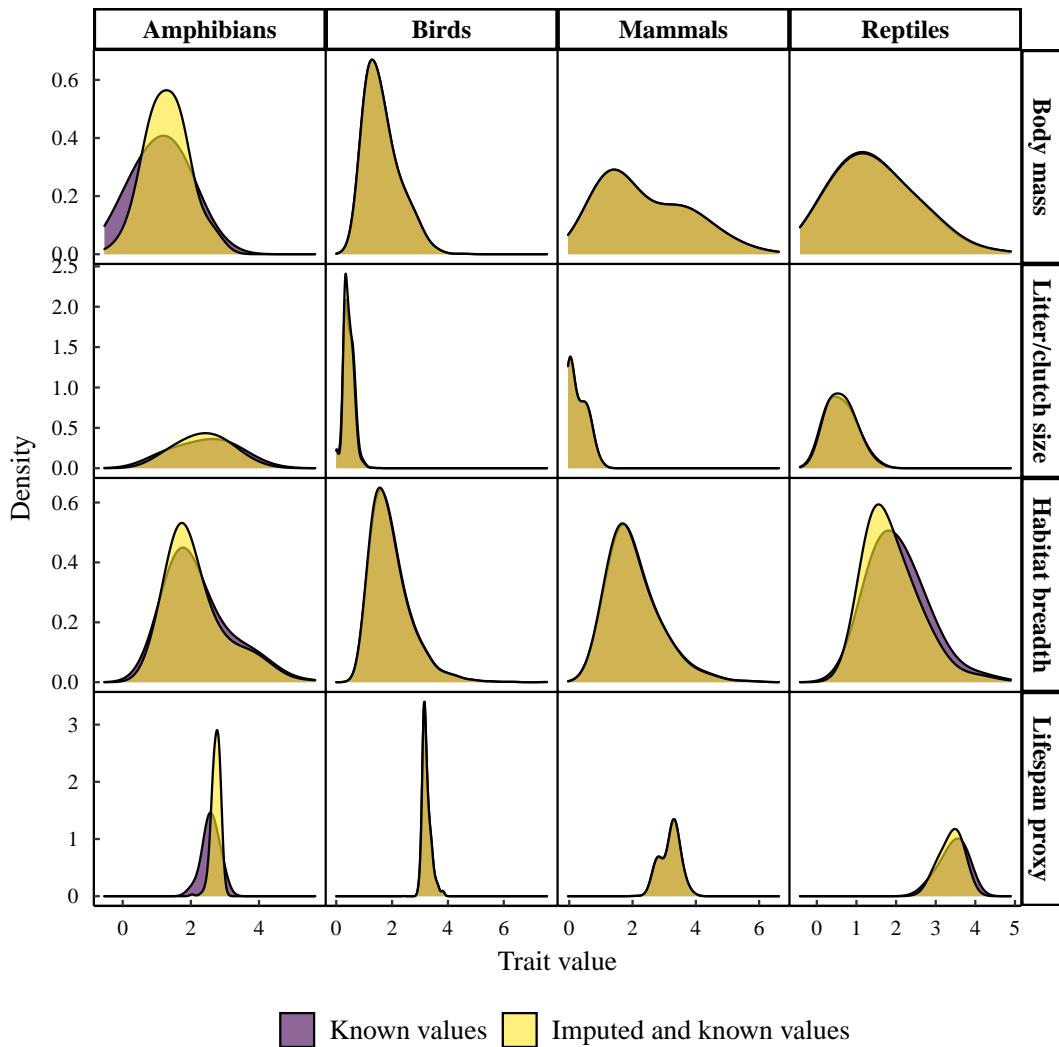
### S3.3 Degree of multicollinearity among traits

Multicollinearity among traits can be problematic when calculating functional diversity indices (Cadotte et al., 2011). After imputing missing trait values and before estimating functional metrics, I hence assessed whether the degree of multicollinearity among categorical and continuous traits was not problematically high. To that end, I used generalised variance inflation factors (Fox and Monette, 1992). Given a regression model, variance inflation factors quantify the overestimation in the variance of estimated regression coefficients due to multicollinearity among the predictors. A GVIF value of 5 or 10 is commonly used as a threshold to select out collinear predictors (Dormann et al., 2013). I used the stepwise.vif function of the Rnalytica package (<https://github.com/awsm-research/Rnalytica>), with a threshold of 5, to determine the GVIF of each trait. I used the imputed traits from the 8th imputation iteration to assess whether multicollinearity was problematically high. Continuous traits were log-10 transformed (except for habitat breadth which was square-rooted). Multicollinearity across traits was not detected to be problematically high, as all traits had a GVIF value below 2 (Table S3.3). As such, all seven traits were included in the calculation of functional diversity indices.

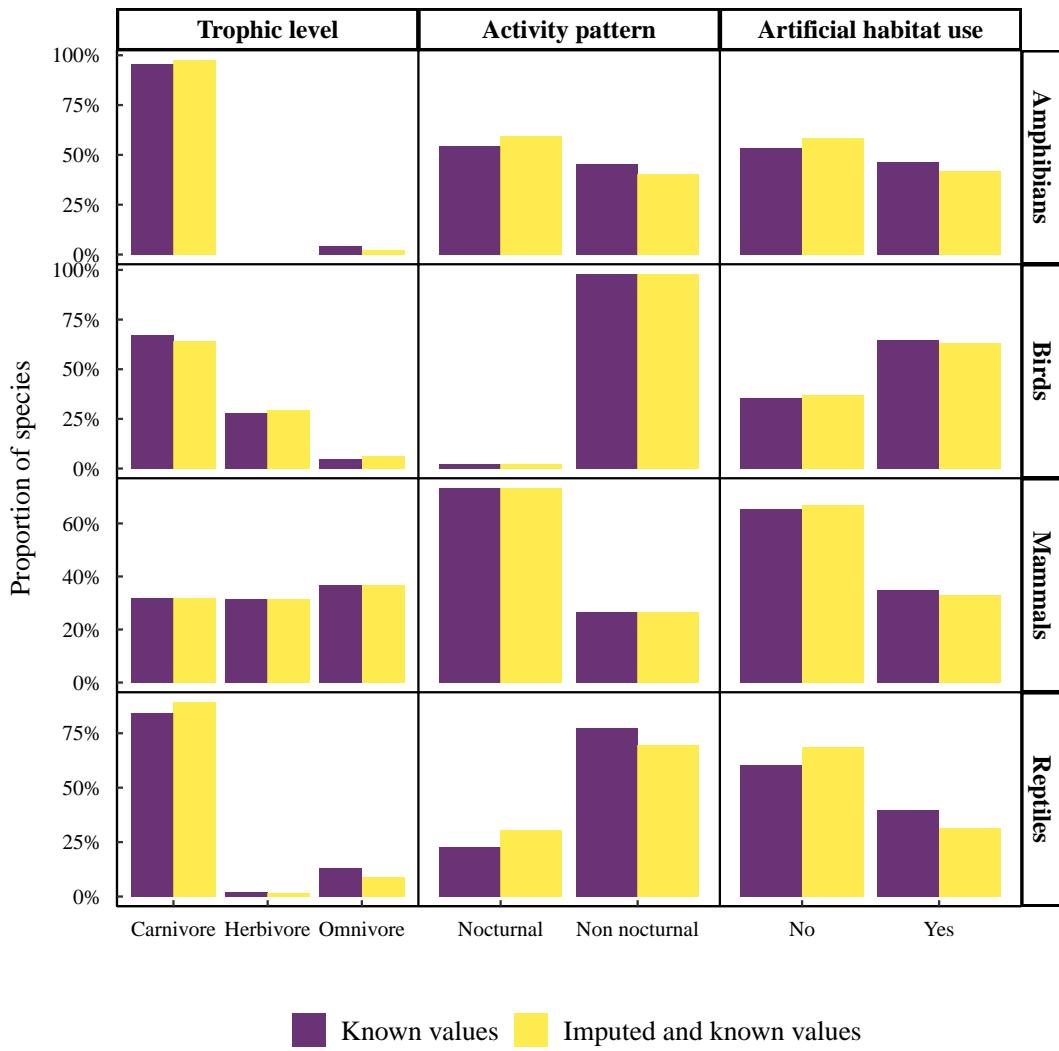
**Table S3.3: Variance inflation factors across considered (imputed) traits.**

Trait	GVIF
Diel activity	1.1
Trophic level	1.3
Use of artificial habitats	1.4
Body mass (log10)	1.5
Habitat breadth (square-root)	1.5
Litter/clutch size (log10)	1.6
Lifespan proxy (log10)	1.7

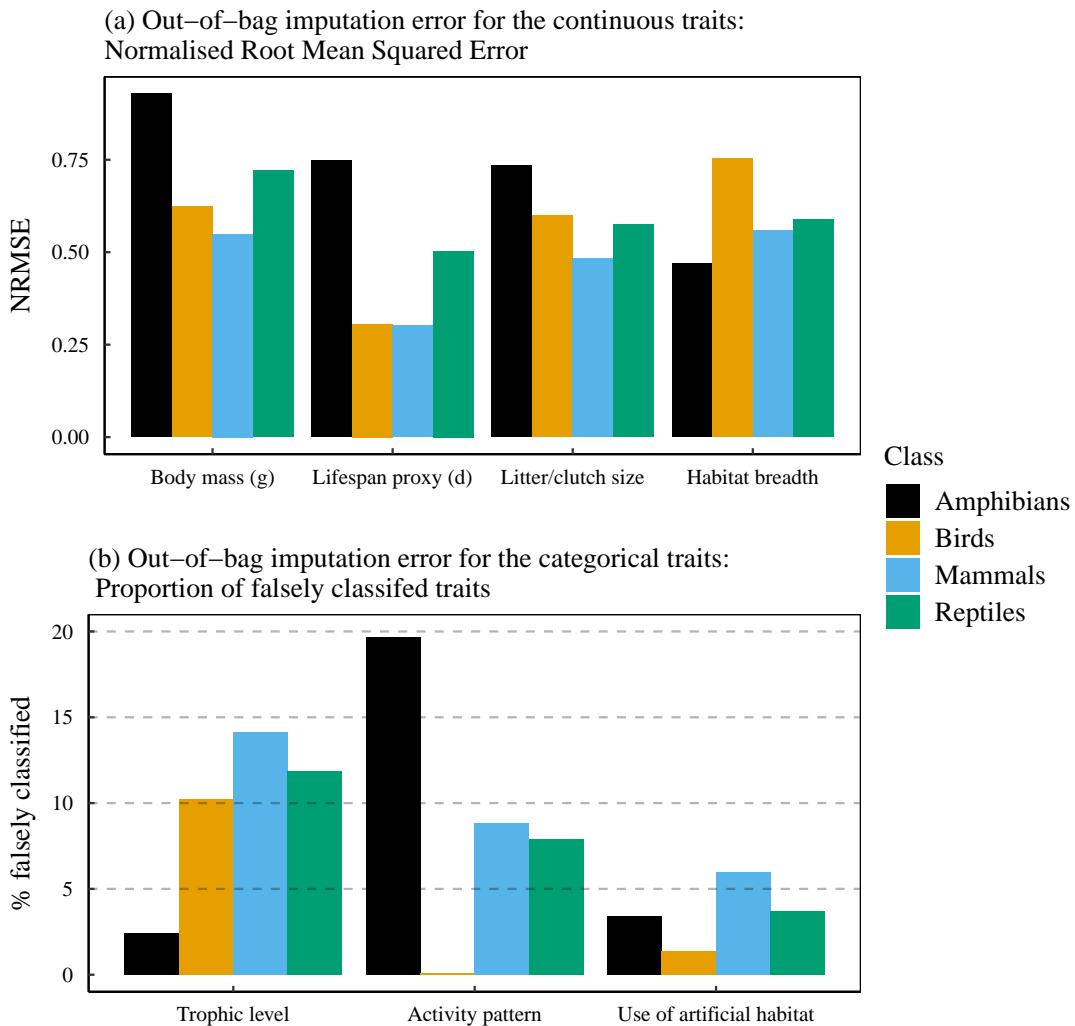
## S3.4 Imputation performance



**Figure S3.6:** Distribution of continuous traits considered in the calculation of the functional diversity metrics (shown as density plots), before and after missing value imputations, in each class and for the species occurring in the PREDICTS database. All traits were log<sub>10</sub>-transformed except Habitat breadth, which was square-rooted.



**Figure S3.7:** Frequency distribution for categorical traits considered in the calculation of the functional diversity metrics (shown as % of total species in each category) before and after missing value imputations, in each class, for the species occurring in the PREDICTS database.



**Figure S3.8: Out-of-bag imputation errors for the continuous traits (a) and categorical traits (b).** For continuous traits, the Normalised Root Mean Squared Error (NRMSE) was obtained by dividing the Mean Squared Error (MSE) by the variance of the known trait distribution, then square-rooting the result. The MSE was returned for each trait by the missForest function (missForest package, Stekhoven, 2016; Stekhoven and Bühlmann, 2012) and corresponds to an out-of-bag error. For categorical traits, the error was estimated as the out-of-bag proportion of falsely classified traits.

### S3.5 Functional loss and functional gain

Across all vertebrates, I estimated functional loss and gain using 84 studies for the tropical subset and 39 studies for the temperate subset (51,514 and 30,470 pairwise comparisons between sites respectively, Table S3.4). Because of this large number of pairwise comparisons, I did not develop a null modelling approach (if I used 100 randomisations per pair of sites, I would need to compute functional loss and gain for more than 8 million pairs, which would be very computationally demanding). I grouped mature, intermediate and young secondary vegetation together in this analysis. I could not estimate the effects in all land uses (for instance, sample sizes for tropical urban sites were too small).

Within classes, I used 18 tropical studies and 1 temperate study for amphibians; 38 and 21 for birds (respectively); 28 and 9 for mammals; and 11 and 7 for reptiles. As sample sizes differed among pairs of land uses and use I was not able to estimate all effects, notably for the intensely-used land uses.

To calculate functional loss and functional gain, the Gower distance matrix was first subsetted to the species occurring in a given pair of sites (see Chapter 3, Methods, ‘Functional traits and diversity indices’). Cailliez corrections were applied when the distance matrix was not Euclidian (Cailliez corrections consist of applying the smallest positive constant to the distances so as to make them Euclidian (Cailliez, 1983); ade4 R package (Dray et al., 2007)). I then performed a principal coordinates analysis on the (corrected) Gower distance matrix, retaining the first two axes to reduce the computational load in the calculation of convex hulls. Sites that contained fewer than three functionally different species were excluded (the computation of a convex hull requiring more species in the assemblage than PCoA axes). Then I estimated the volume of trait space occupied by each assemblage of a given pair, as well as the volume of the shared trait space (intersection), from which I derived functional loss and functional gain.

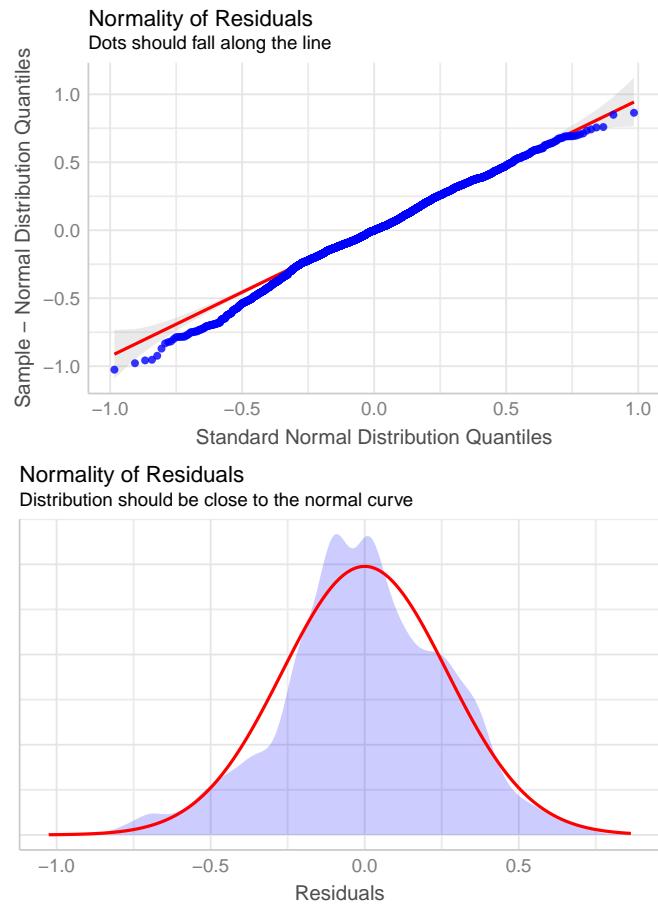
**Table S3.4: Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain across all species.**

Region	Pairs	Minimal use	Light use	Intense use
Temperate	PV-PV	7626	22546	492
	PV-SV	511	72	–
	PV-PF	9	166	–
	PV-PA	8	40	–
	PV-CR	150	–	–
	PV-UR	6306	1197	7
Tropical	PV-PV	8547	4016	16722
	PV-SV	6584	1124	9713
	PV-PF	580	1378	–
	PV-PA	36	20	22
	PV-CR	1700	1088	–
	PV-UR	–	–	–

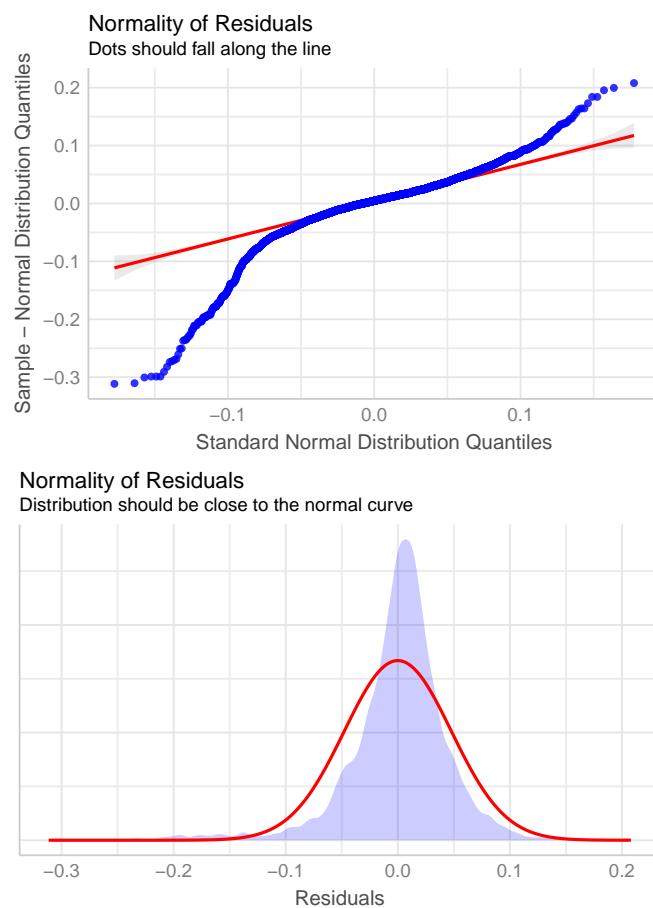
**Table S3.5: Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain within each class.**

Class	Region	Pair of land uses	Minimal use	Light use	Intense use
Amphibians	Temperate	PV/PV	–	45	–
Amphibians	Temperate	PV/SV	8	70	–
Amphibians	Temperate	PV/AGR	3	–	–
Amphibians	Temperate	PV/UR	6	100	–
Amphibians	Tropical	PV/PV	501	241	307
Amphibians	Tropical	PV/SV	838	–	90
Amphibians	Tropical	PV/PF	422	91	–
Amphibians	Tropical	PV/AGR	1	3	1
Birds	Temperate	PV/PV	7,382	19,300	491
Birds	Temperate	PV/SV	150	1	–
Birds	Temperate	PV/PF	9	166	–
Birds	Temperate	PV/AGR	145	40	–
Birds	Temperate	PV/UR	6,300	992	–
Birds	Tropical	PV/PV	5,059	3,117	9,014
Birds	Tropical	PV/SV	3,491	1,058	5,225
Birds	Tropical	PV/PF	156	994	–
Birds	Tropical	PV/AGR	1,626	1,085	–
Mammals	Temperate	PV/PV	110	3,030	–
Mammals	Temperate	PV/SV	25	–	–
Mammals	Temperate	PV/AGR	5	–	–
Mammals	Temperate	PV/UR	–	105	7
Mammals	Tropical	PV/PV	1,989	637	64
Mammals	Tropical	PV/SV	230	65	8
Mammals	Tropical	PV/PF	2	–	–
Mammals	Tropical	PV/AGR	109	20	21
Reptiles	Temperate	PV/PV	132	2	1
Reptiles	Temperate	PV/SV	250	1	–
Reptiles	Temperate	PV/AGR	5	–	–
Reptiles	Tropical	PV/PV	989	137	5,140
Reptiles	Tropical	PV/SV	1,760	1	3,456
Reptiles	Tropical	PV/PF	–	190	–

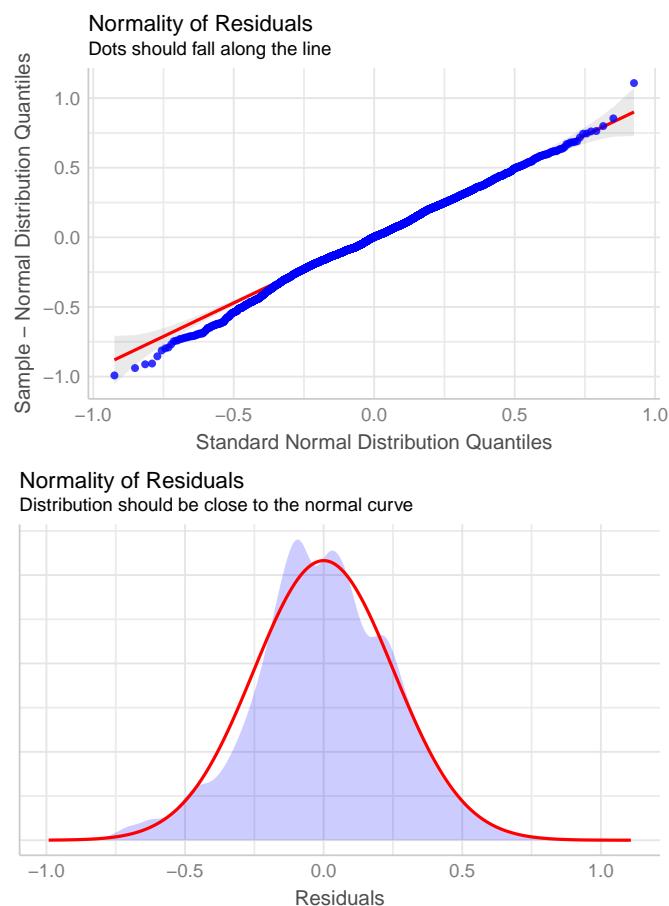
## S3.6 Diagnostic plots



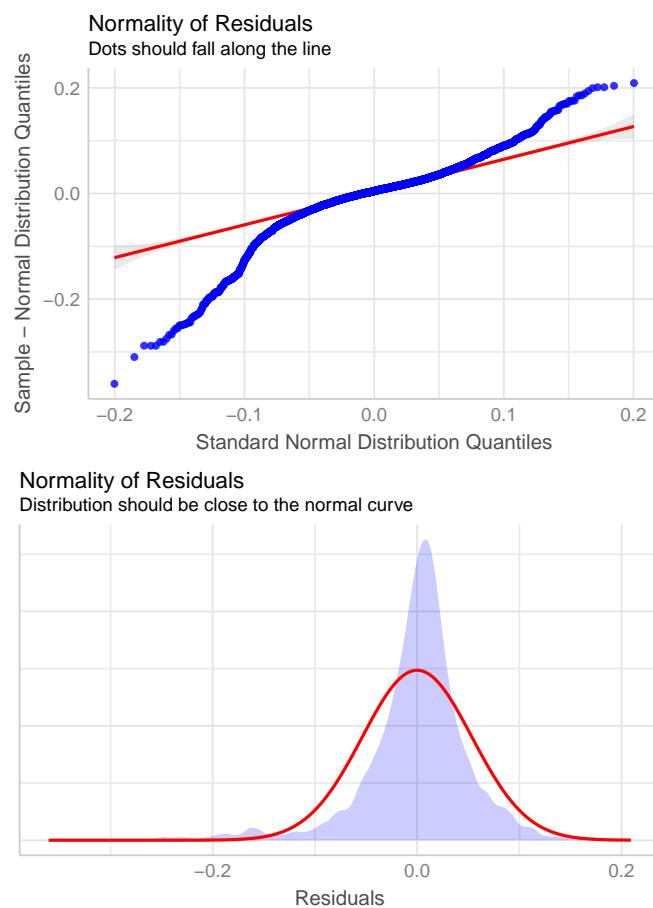
**Figure S3.9:** Diagnostic plots for Model 1a, obtained using the 'performance' R package (Lüdecke et al., 2021).



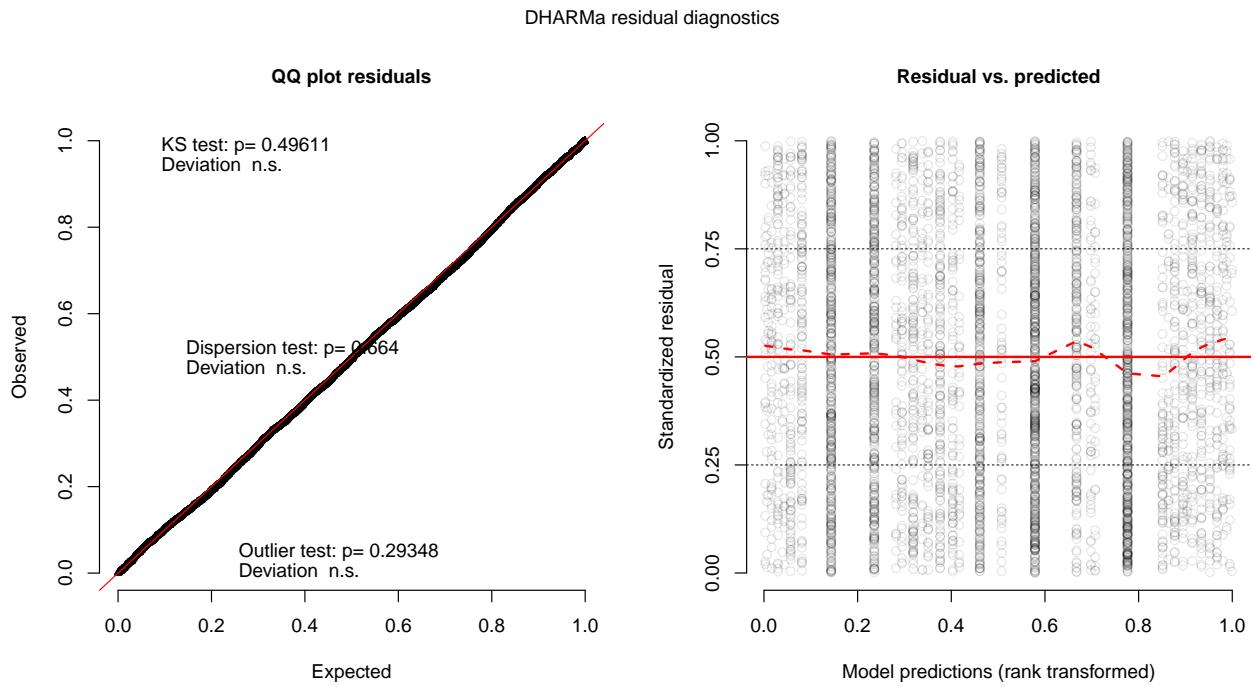
**Figure S3.10:** Diagnostic plots for **Model 1b**, obtained using the ‘performance’ R package (Lüdecke et al., 2021).



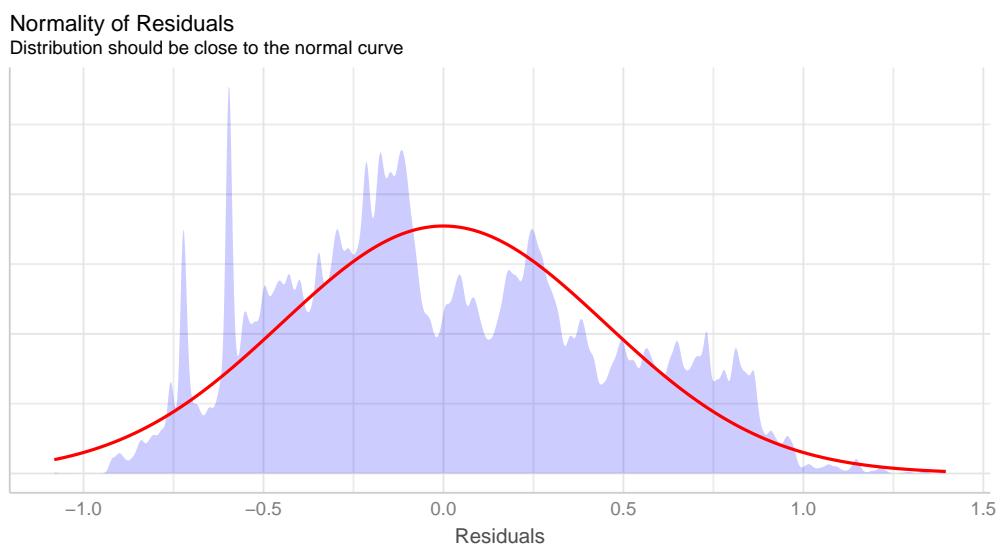
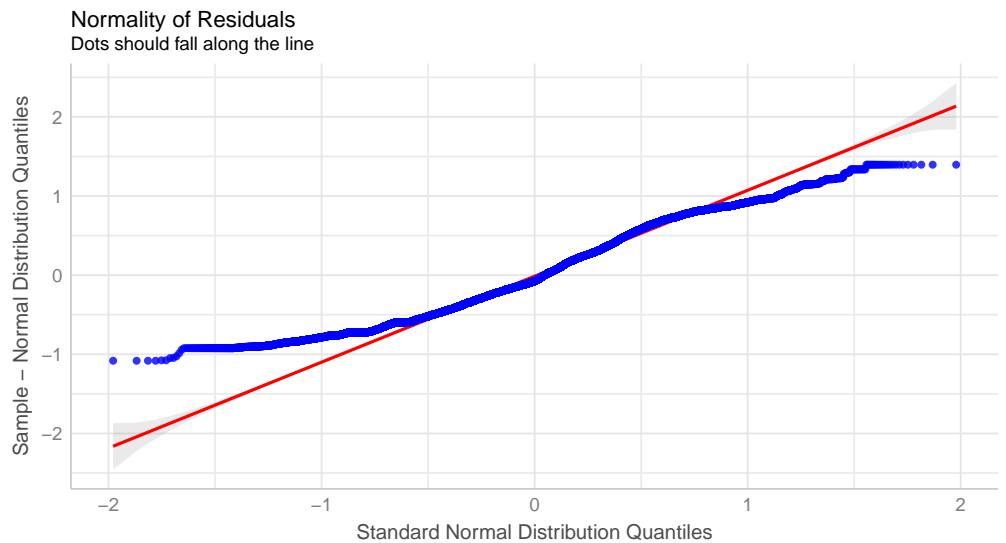
**Figure S3.11:** Diagnostic plots for Model 2a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).



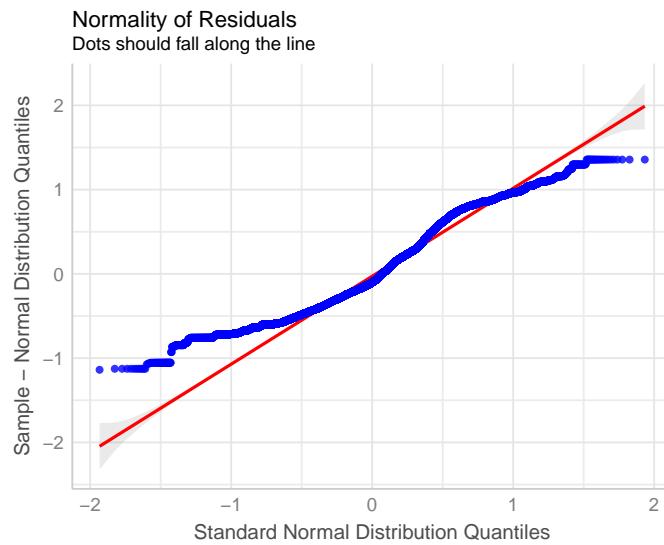
**Figure S3.12:** Diagnostic plots for Model 2b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).



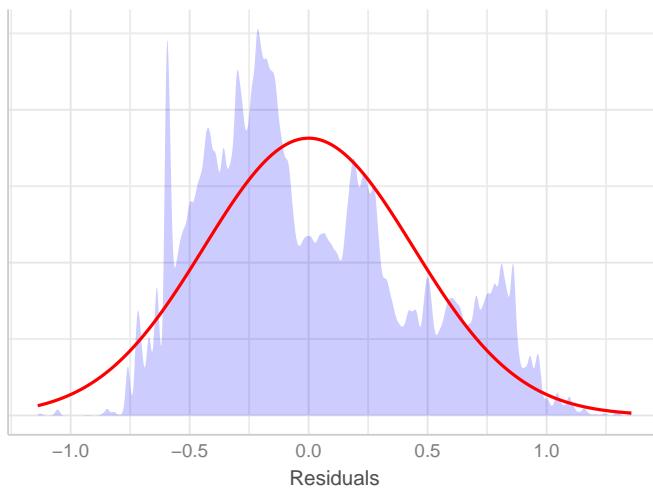
**Figure S3.13: Diagnostic plots for Model 3**, obtained using the ‘DHARMA’ R package (Hartig, 2021).



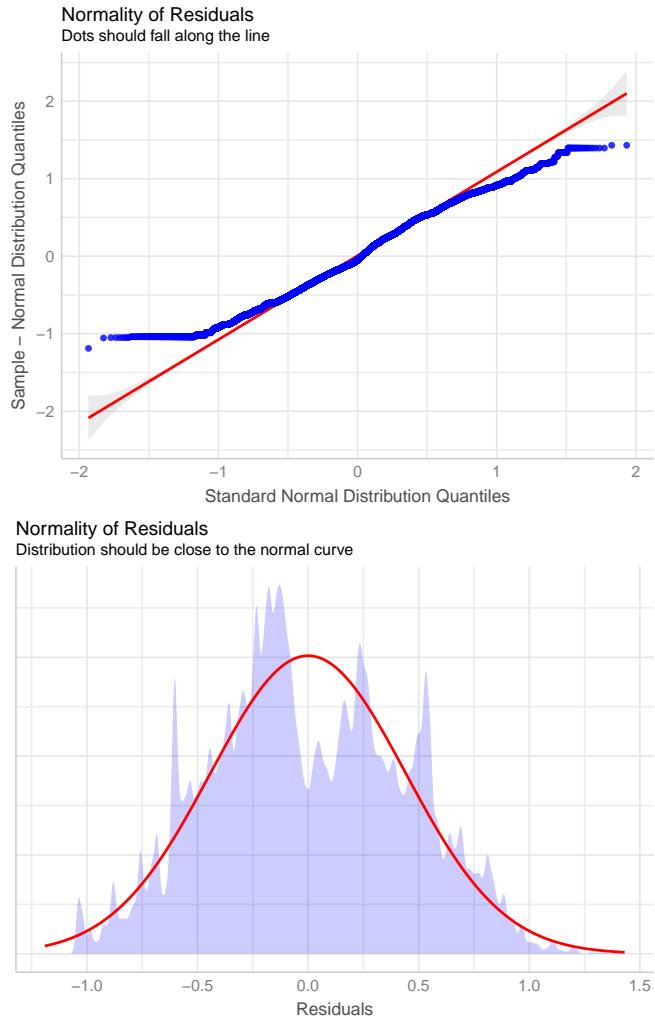
**Figure S3.14:** Diagnostic plots for Model 4a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).



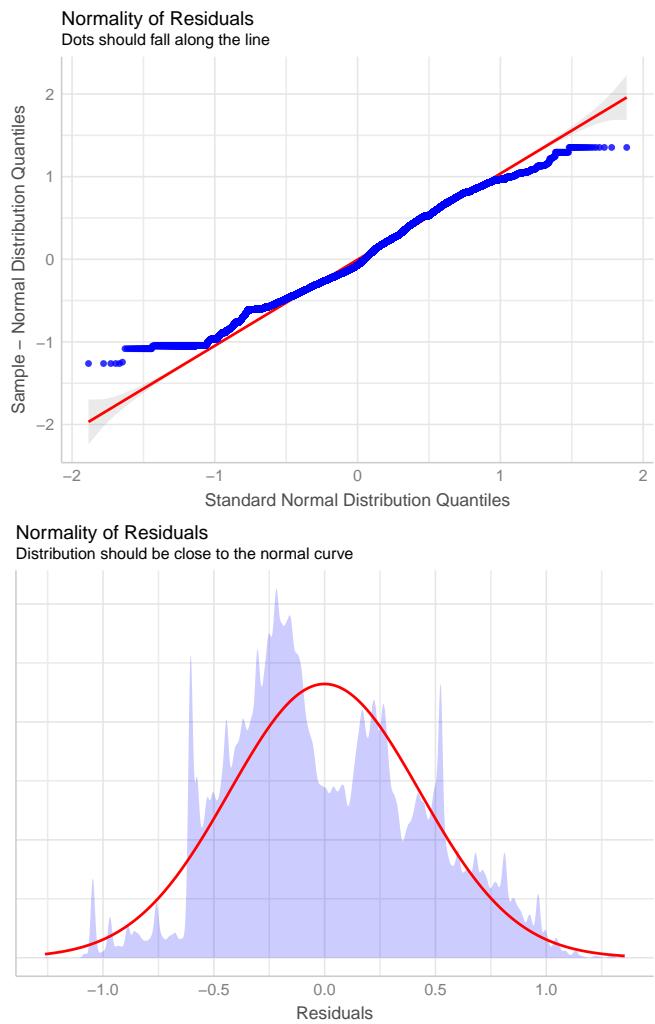
Normality of Residuals  
Distribution should be close to the normal curve



**Figure S3.15:** Diagnostic plots for Model 4b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

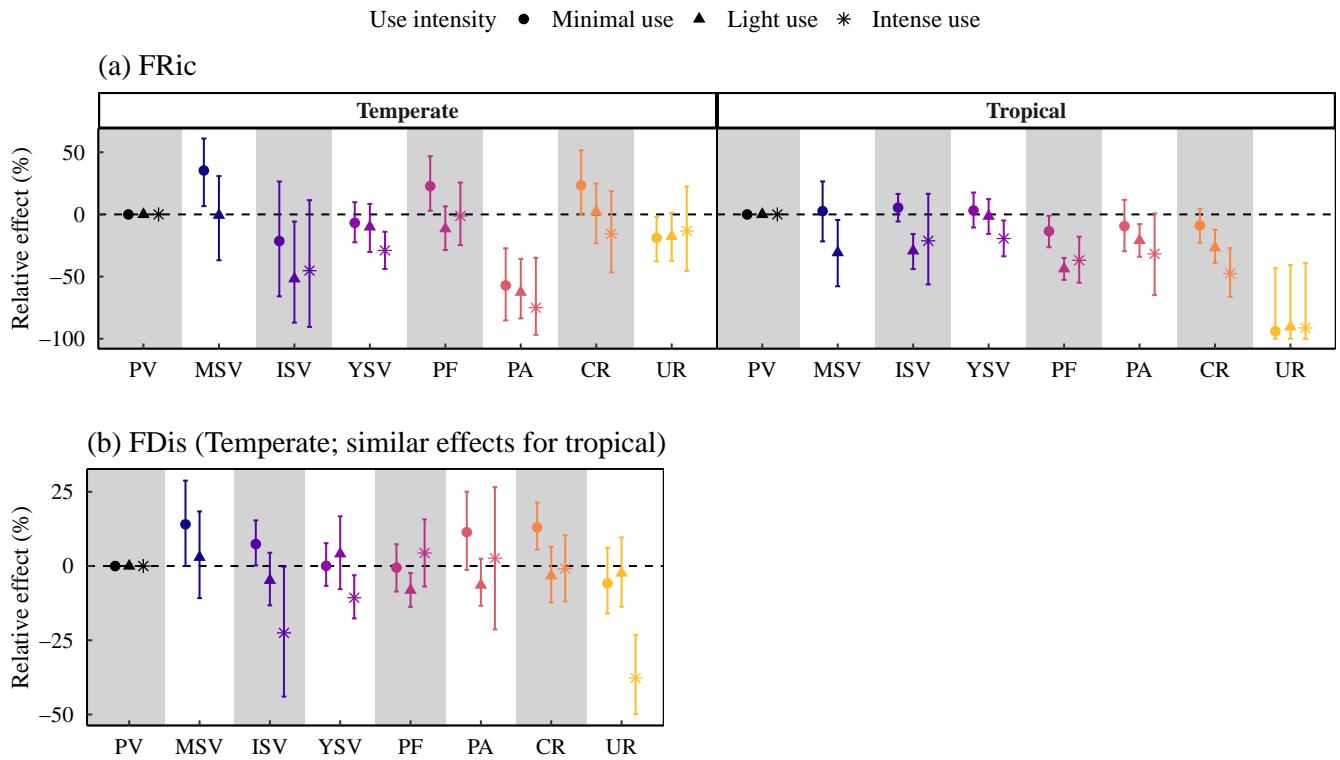


**Figure S3.16:** Diagnostic plots for Model 5a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

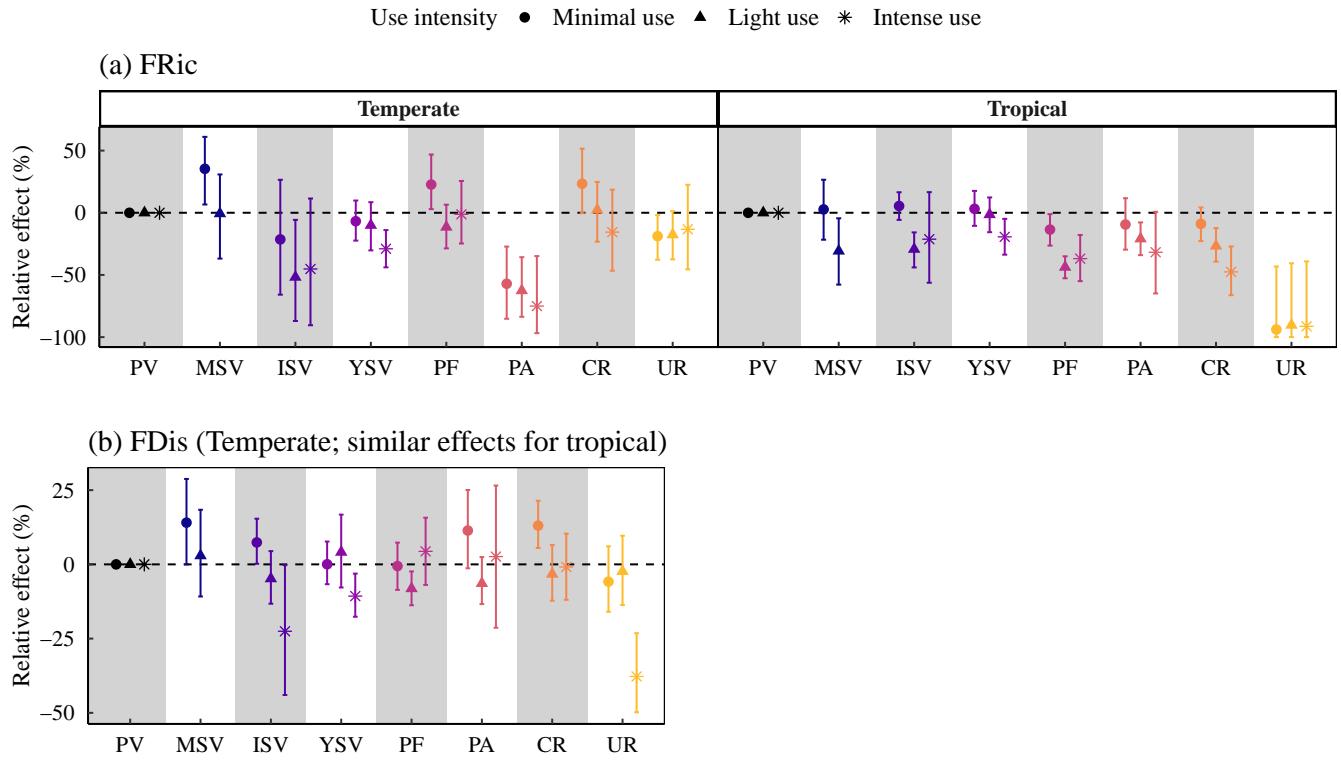


**Figure S3.17:** Diagnostic plots for Model 5b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

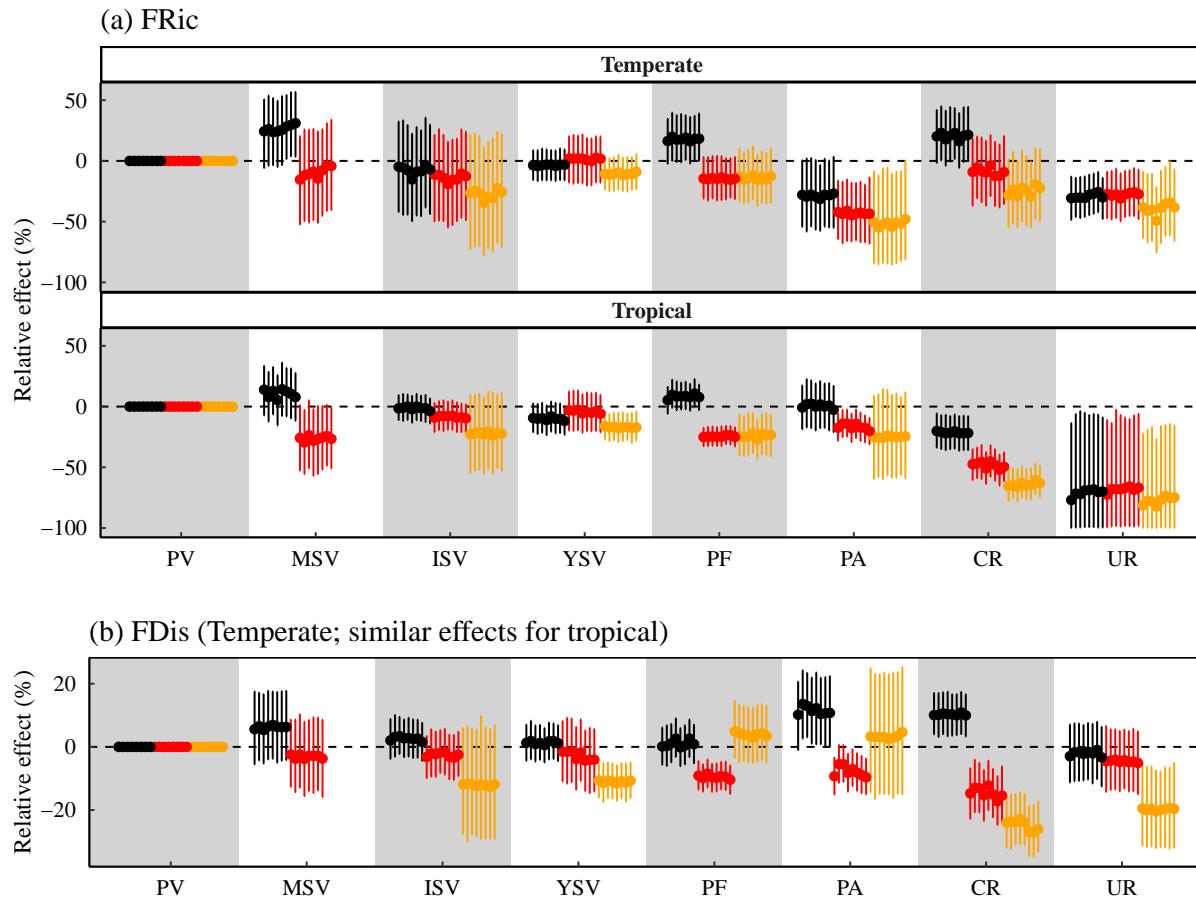
### S3.7 Model robustness



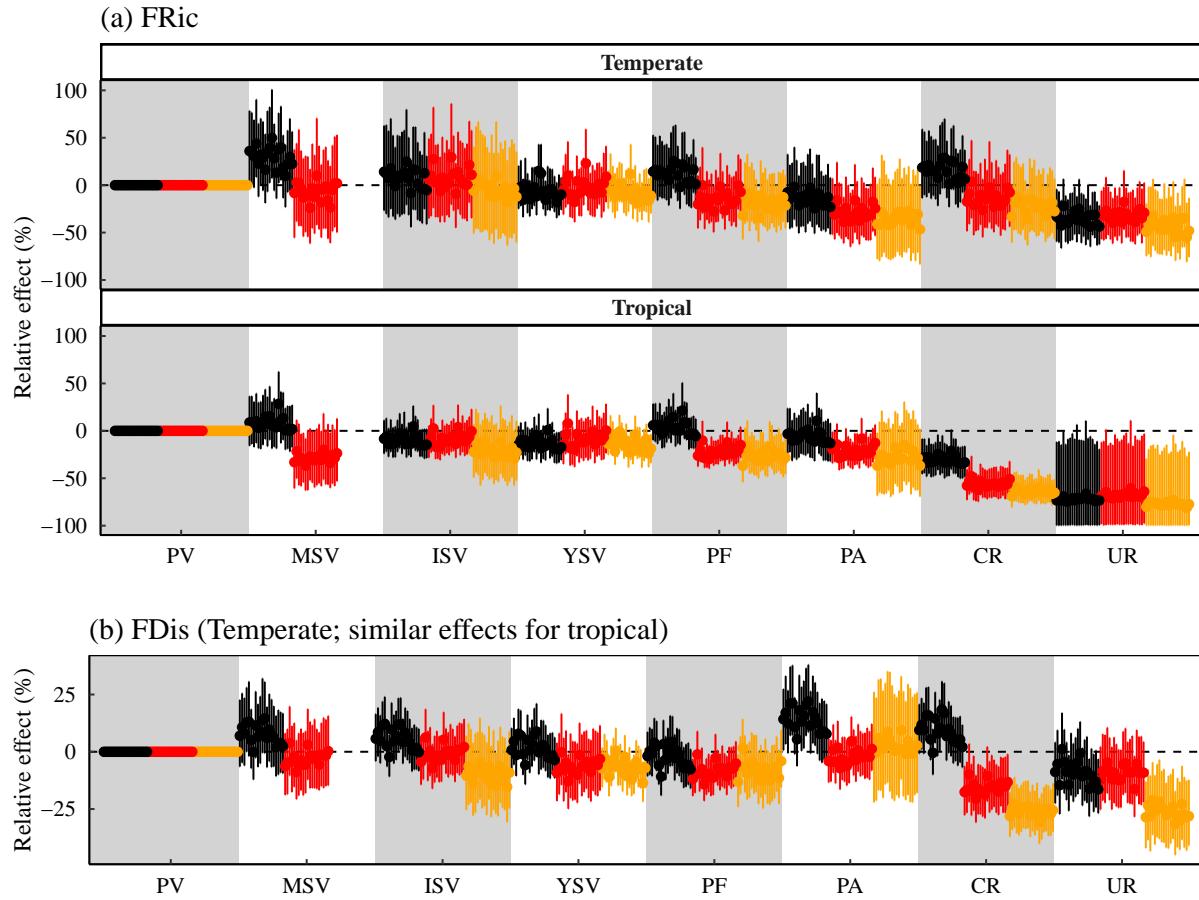
**Figure S3.18: Effects of land use, land-use intensity and region on FRic (a) and FDis (b) across vertebrates, for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



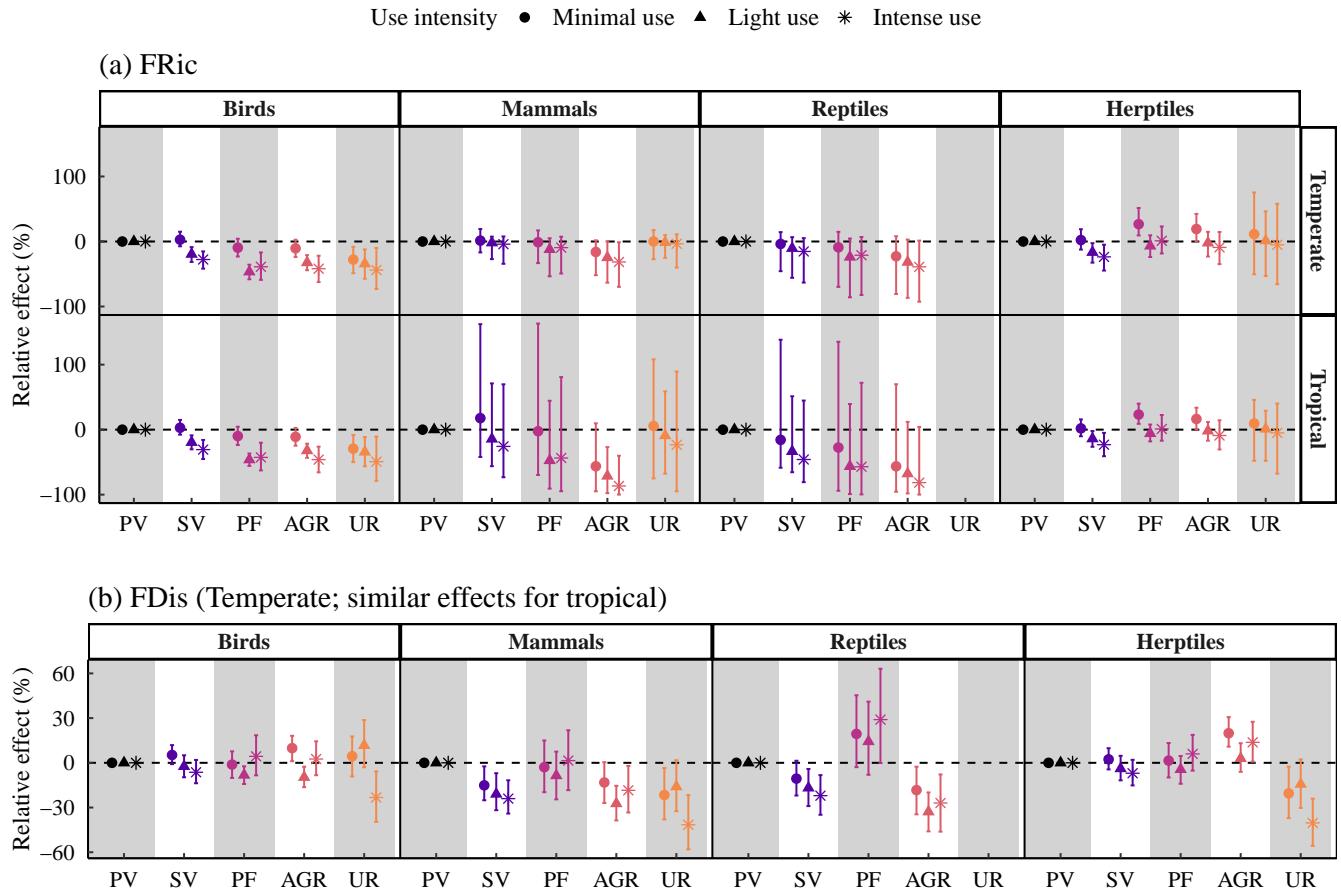
**Figure S3.19: Effects of land use, land-use intensity and region on FRic (a) and FDis (b), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness), with geographical range size as an additional trait considered in the calculation of functional diversity metrics.** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b , which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



**Figure S3.20: Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when calculating FRic and FDis with each set of imputed traits (eight in total). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.**



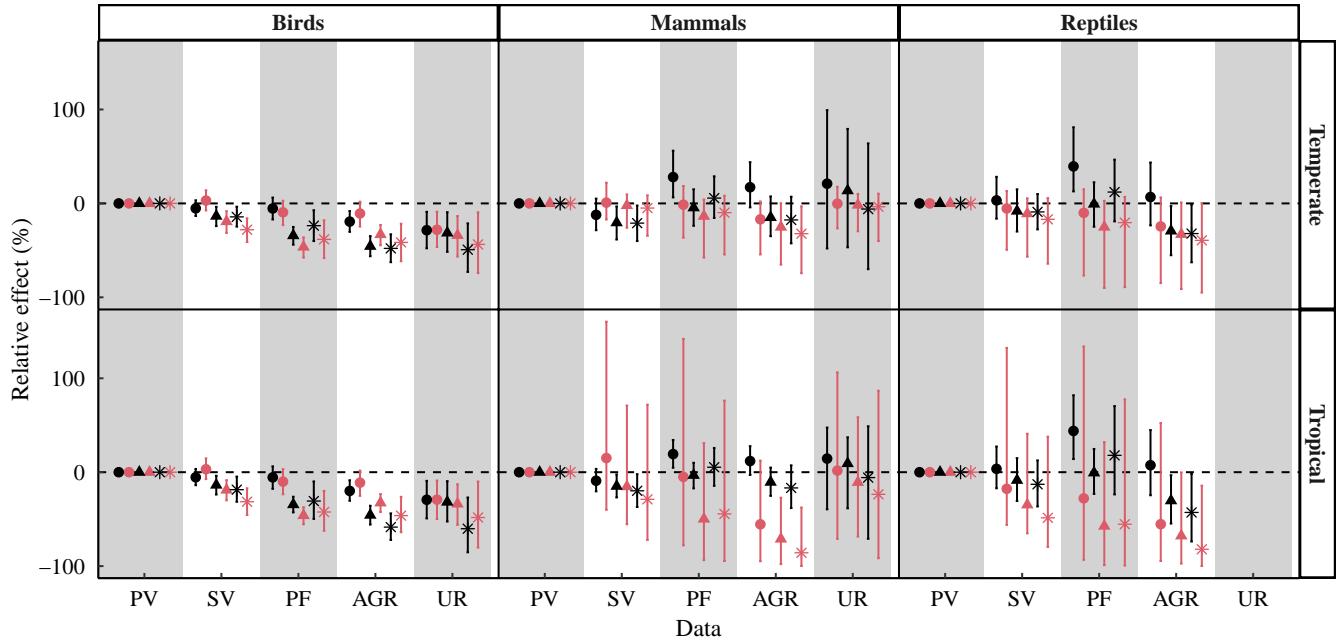
**Figure S3.21: Effects of land use and land-use intensity on FRic (a) and FDis (b), obtained when re-sampling primary vegetation sites twenty independent times.** I fixed the sample size for primary vegetation sites at 50. For FRic, we fitted Model 1a (see main text, Chapter 3), which included the effects of land use, land-use intensity and region, as well as interactions between land use and land-use intensity and between land use and region. For FDis, I fitted Model 1b, which included effects of land use, land-use intensity and region, and interactions between land use and land-use intensity. Error bars represent 95% confidence intervals. MSV: mature secondary vegetation; ISV: intermediate secondary vegetation; YSV: young secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for Intense use in MSV could not be estimated as there weren't enough sampled sites.



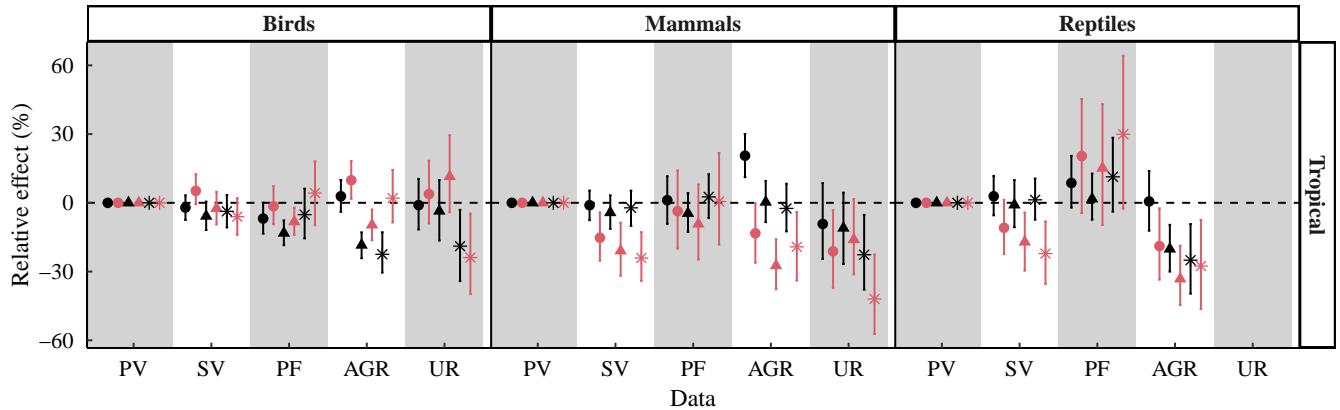
**Figure S3.22: Effects of land use, land-use intensity and taxonomic class on FRic (a) and FDis (c), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). I did not include the effects of region here as sample sizes were not large enough for some classes. For FRic, the model included the effects of land use, land-use intensity and class, and interactions between land use and land-use intensity as well as land use and class. For FDis, the model included an additional interaction between land-use intensity and class. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; PA: pasture; CR: cropland; UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.

—●— complete and imputed trait data —●— complete trait data subset      Use intensity    ● Minimal use    ▲ Light use    \* Intense use

(a) FRic

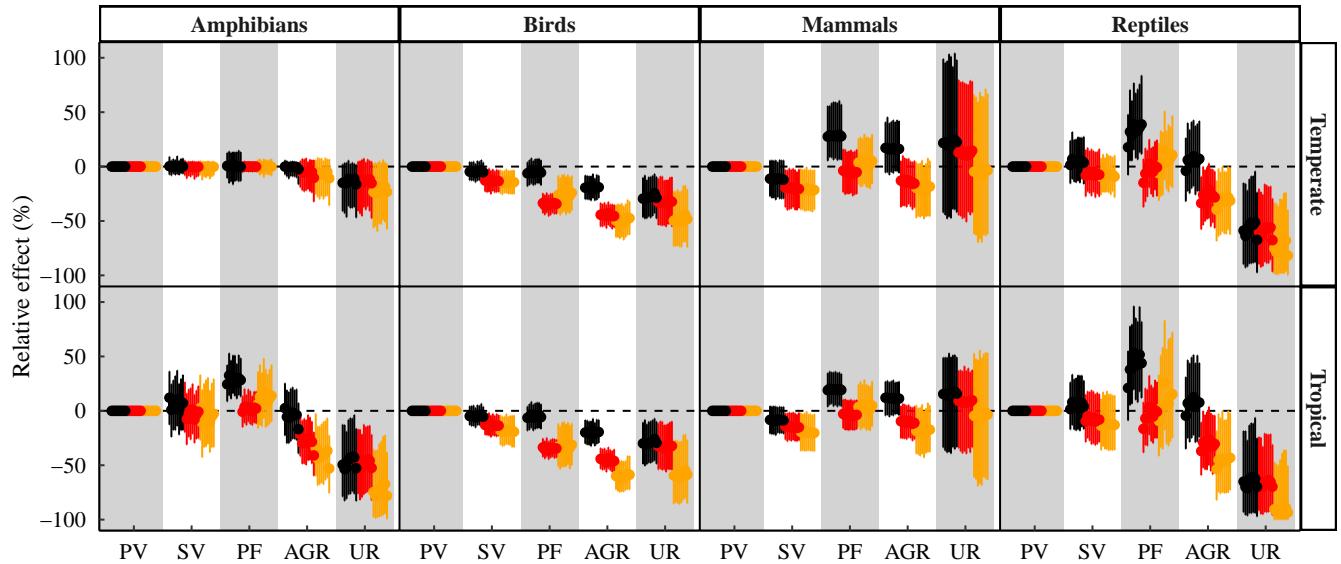


(a) FDis

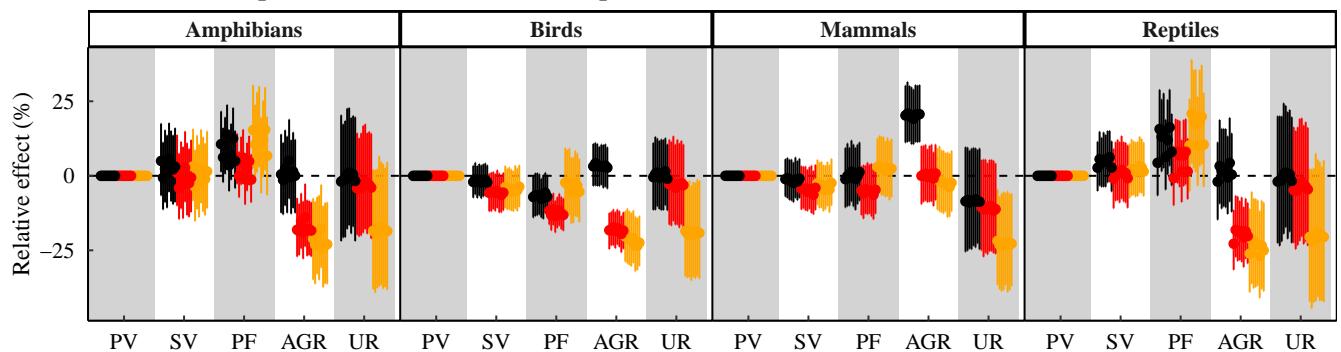


**Figure S3.23: Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis obtained with the imputed trait data (black points) or with the complete data subsets (red points).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 2a (see main text, Chapter 3), and I fitted Model 2b for FDis. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; AGR: agricultural (cropland and pasture); UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.

(a) FRic



(b) FDis (Temperate; similar effects for tropical)



**Figure S3.24: Effects of land use, region, land-use intensity and taxonomic class on FRic and FDis, obtained when calculating FRic and FDis with each set of imputed traits (eight in total).** Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, I fitted Model 2a (see main text, Chapter 3), and Model 2b for FDis. Error bars represent 95% confidence intervals. SV: secondary vegetation; PF: plantation forest; AGR: agricultural (cropland and pasture); UR: urban. Effects for reptiles in urban land uses could not be estimated as there weren't enough sampled sites.

## S3.8 Model robustness – time since land-use conversion

Time since land-use conversion could have important impacts on assemblage composition and thus, on local functional diversity. I did not investigate these effects because PREDICTS contained data on time since land-use conversion only for about 22% of the sites, considerably reducing samples sizes. Here, I investigated whether the results are likely robust to the inclusion of time since land-use conversion using the subset of sites for which time since land-use conversion was provided. To this end, I found the best-fitting models explaining FRic and FDis, using backwards stepwise selection, starting with complete models that included the effects of land use, time since land-use conversion, region, land-use intensity (for FRic only) and all two-way interactions among these predictors.

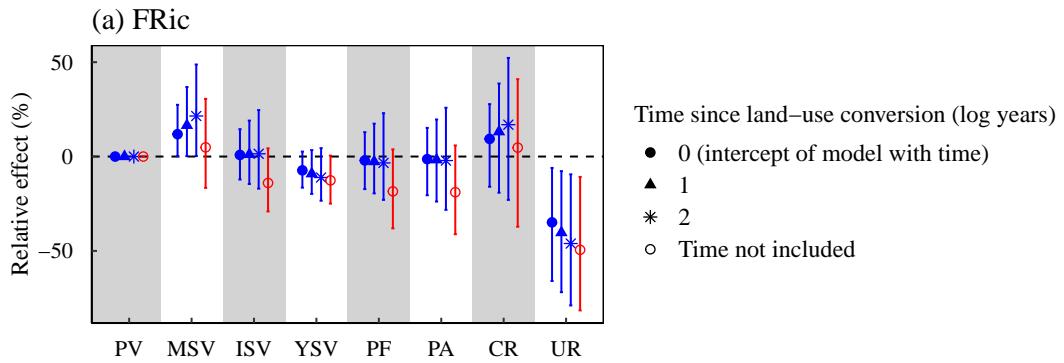
- For FRic, the best-fitting model includes the main effects of land use and time since land-use conversion, but no interaction between these predictors. The model's summary (Table S3.6) show that time since conversion has a significant negative effect on FRic, but the relationship between FRic and time since land-use conversion is similar in different land uses (as there are no interactions between land use and time since conversion, such that the slopes are similar in different land uses, and so the rate at which FRic decreases with time is similar in different land uses). The intercept is only different for urban land uses (significantly lower). Thus, based on this data subset, I expect time since land-use conversion to have a similar effect in different land uses.

**Table S3.6:** Summary of the model explaining FRic by land use and time since land-use conversion, fitted on the subset of data for which there are information on time since land-use conversion.

	Estimate	Std. Error	t value
Intercept: Primary vegetation	1.156	0.073	15.921
Mature secondary vegetation	0.178	0.093	1.907
Intermediate secondary vegetation	0.018	0.072	0.249
Young secondary vegetation	-0.078	0.051	-1.532
Plantation forest	-0.018	0.082	-0.224
Pasture	-0.005	0.093	-0.054
Cropland	0.133	0.152	0.875
Urban	-0.316	0.133	-2.368
log_Years	-0.094	0.021	-4.566

I then compare this model's predictions with a simpler model that does not account for time since land-use conversion ( $\text{FRic} \sim \text{Land use}$ ). The predictions (Fig. S3.25) show that including time since land-use conversion does not bias the results, as I find a similar significant effect with both models in

urban land uses, and elsewhere the effects are congruent. Thus, given this data subset, I argue that the results are robust to the inclusion of time since land-use conversion.



**Figure S3.25: Effects of land use on FRic for the model that includes time since land-use conversion (blue points) versus the model that does not take time since land-use conversion into account (red points).**

- For FDis, the best-fitting model includes the main effects of land use, time since land-use conversion as well as interactions between land use and time since land-use conversion (I didn't consider land-use intensity in the starting model because of sample size issues). Nevertheless, the main effect of time since land-use conversion is not significant (Table S3.7), and the relationship between time since land-use conversion and FDis is not significant in most land uses (except for plantation forest). Thus, I argue the available data do not allow us to properly investigate the relationship between time since land-use conversion and FDis.

**Table S3.7:** Summary of the model explaining FDis by land use and time since land-use conversion, fitted on the subset of data for which I have information on time since land-use conversion.

	Estimate	Std. Error	t value
Intercept: Primary vegetation	0.366	0.011	32.219
Mature secondary vegetation	0.032	0.055	0.577
Intermediate secondary vegetation	-0.015	0.050	-0.298
Young secondary vegetation	0.020	0.015	1.386
Plantation forest	0.074	0.023	3.213
Pasture	-0.017	0.048	-0.346
Cropland	-0.013	0.042	-0.317
Urban	0.031	0.054	0.573
log_Years	-0.004	0.004	-1.186
Mature secondary vegetation:log_Years	-0.005	0.015	-0.335
Intermediate secondary vegetation:log_Years	0.011	0.016	0.650
Young secondary vegetation:log_Years	-0.008	0.007	-1.170
Plantation forest:log_Years	-0.023	0.007	-3.077
Pasture:log_Years	0.010	0.015	0.688
Cropland:log_Years	0.007	0.012	0.620
Urban:log_Years	-0.016	0.022	-0.714

## **Appendix 4: Supporting information for Chapter 4**

## **Appendix 5: Supporting information for Chapter 5**

# Bibliography for the Appendices

- Borges, R., Machado, J. P., Gomes, C., Rocha, A. P., and Antunes, A. (2018). Measuring phylogenetic signal between categorical traits and phylogenies. *Bioinformatics*. DOI: 10.1093/bioinformatics/bty800.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., and Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1003537.
- Bruggeman, J., Heringa, J., and Brandt, B. W. (2009). PhyloPars: Estimation of missing parameter values using phylogeny. *Nucleic Acids Research*. DOI: 10.1093/nar/gkp370.
- Buuren, S van and Oudshoorn, C. G. M. (2007). MICE: multivariate imputation by chained equations. *R package version*.
- Cadotte, M. W., Carscadden, K., and Mirochnick, N. (2011). Beyond species: Functional diversity and the maintenance of ecological processes and services. DOI: 10.1111/j.1365-2664.2011.02048.x.
- Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*. DOI: 10.1007/BF02294026.
- 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.
- Debastiani, V. J., Bastazini, V. A. G., and Pillar, D (2021). Ecological Informatics Using phylogenetic information to impute missing functional trait values in ecological databases. 63.April. DOI: 10.1016/j.ecoinf.2021.101315.
- Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., Morales-Castilla, I., Olalla-Tárraga, M. Á., Rodríguez, M. Á., and Hawkins, B. A. (2012). On the selection of phylogenetic eigenvectors for ecological analyses. *Ecography*. DOI: 10.1111/j.1600-0587.2011.06949.x.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*. DOI: 10.1111/j.1600-0587.2012.07348.x.
- Dray, S., Dufour, A.-B., and Chessel, D. (2007). The ade4 Package – II: Two-Table and K-Table Methods. *R News*, 7.2, 47–52.

- 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., Antonelli, 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.
- Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*. DOI: 10.1080/01621459.1992.10475190.
- Hartig, F. (2021). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.1.
- Hudson, L. N. et al. (2014). The PREDICTS database: A global database of how local terrestrial biodiversity responds to human impacts. *Ecology and Evolution*. DOI: 10.1002/ece3.1303.
- (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. *Ecology and Evolution*. DOI: 10.1002/ece3.2579.
- 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.
- Johnson, T. F., Isaac, N. J., Paviolo, A., and González-Suárez, M. (2021). Handling missing values in trait data. *Global Ecology and Biogeography*. DOI: 10.1111/geb.13185.
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., and Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6.60, 3139. DOI: 10.21105/joss.03139.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*. DOI: 10.1038/44766.
- Penone, C., Davidson, A. D., Shoemaker, K. T., Di Marco, M., Rondinini, C., Brooks, T. M., Young, B. E., Graham, C. H., and Costa, G. C. (2014). Imputation of missing data in life-history trait datasets: Which approach performs the best? *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.12232.
- Rabosky, A. D., Cox, C., Rabosky, D., Title, P., Holmes, I., Feldman, A., and McGuire, J. (2016). Coral snakes predict the evolution of mimicry across New World snakes. *Nature Communications*, 7, 11484.
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3, 217–223.
- 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.
- Santos, T. (2018). Package ‘PVR’. Phylogenetic Eigenvectors Regression and Phylogenetic Signal-Representation Curve.

- Stekhoven, D. J. (2016). Nonparametric Missing Value Imputation using Random Forest. *R Package version 1.4*. DOI: 10.1093/bioinformatics/btr597.
- Stekhoven, D. J. and Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-type data. *Bioinformatics*. DOI: 10.1093/bioinformatics/btr597.
- 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.
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*. DOI: 10.1093/bioinformatics/17.6.520.