

Supporting information for

Intensive human land uses negatively affect vertebrate functional diversity

October 21, 2021

S1 Land-use categories in PREDICTS and sample sizes

Table S1: 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.

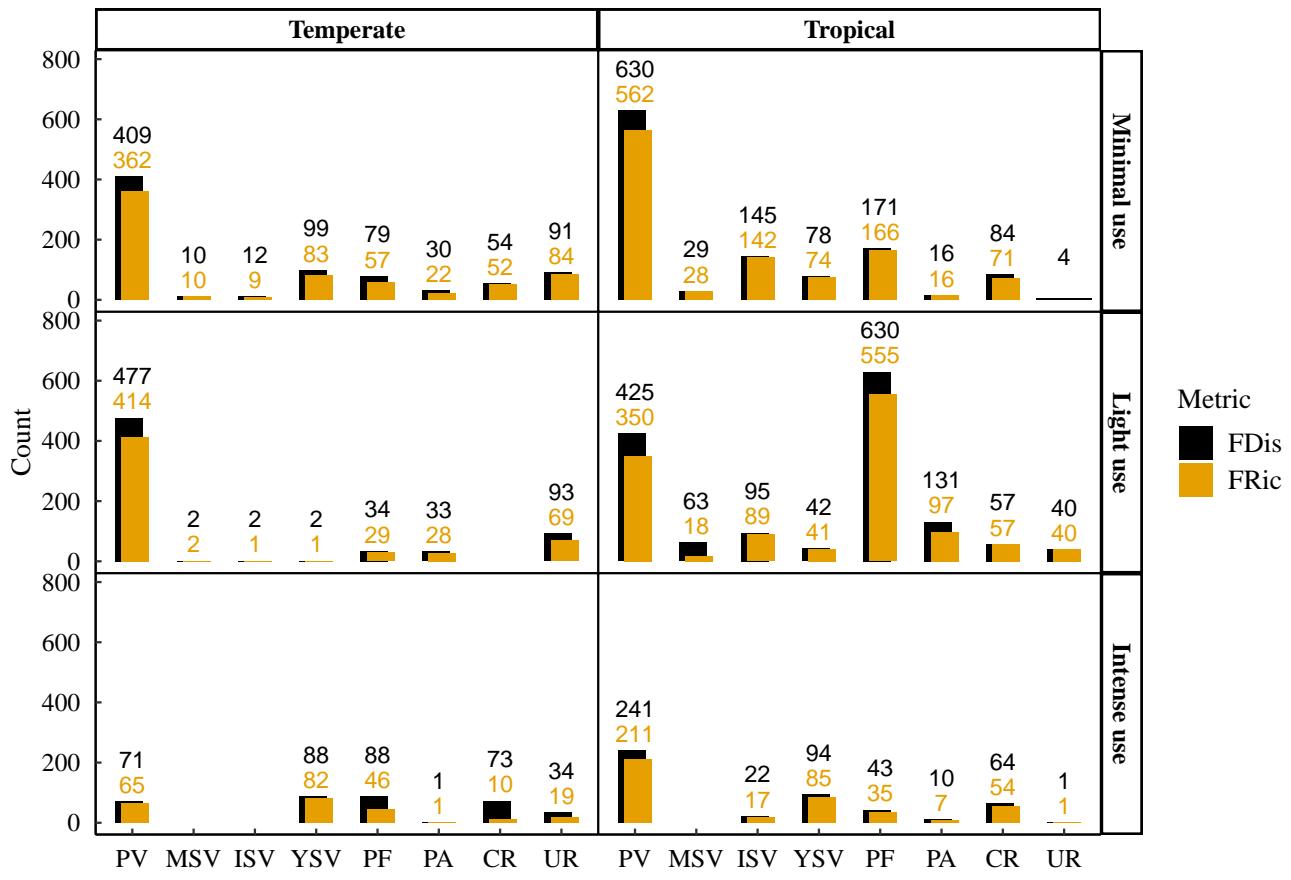


Figure S1: Number of sites in each land use and use intensity for which FRic and FDis were calculated, across all vertebrates.
The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).

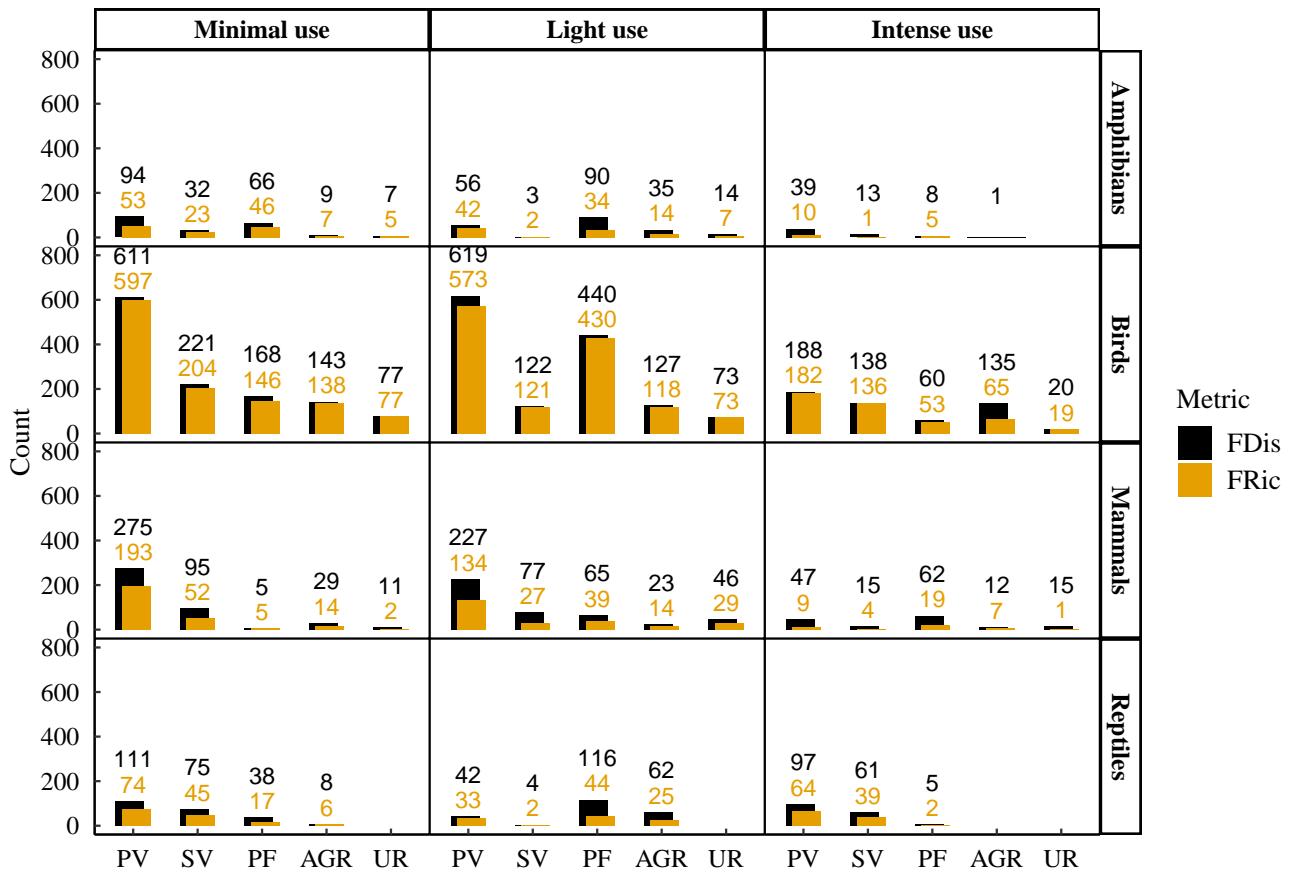


Figure S2: Number of sites in each land use, use intensity and each class for which FRic and FDis were calculated. The number of sites for FRic can be smaller than the number of sites for FDis because FRic could not necessarily be computed for all assemblages in which FDis was estimated (in assemblages where species richness was 2 or 3, FRic could not be computed).

S2 Trait data & imputation of missing trait values

We used a compilation of trait data across terrestrial vertebrates (Etard et al., 2020). We used 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, Fig. S3).

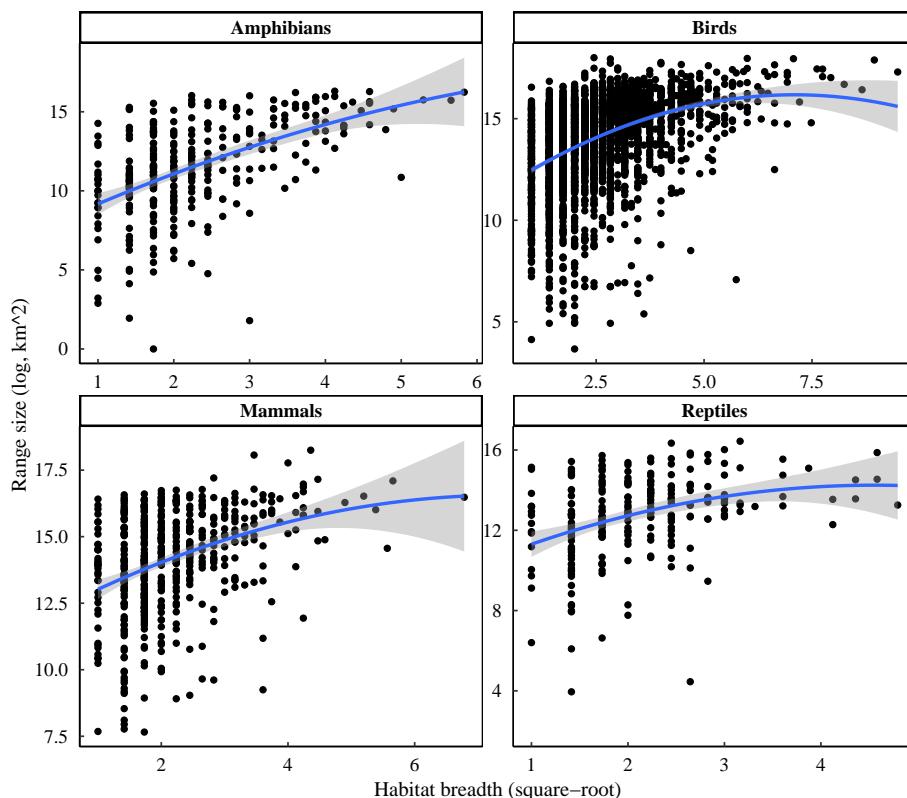


Figure S3: Relationship between habitat breadth and geographical range sizes. The derivation of geographical range sizes is described in Etard et al. (2020).

Trait-data coverage was highly variable among classes and traits, with important geographical and phylogenetic biases in trait data for reptiles and amphibians (Etard et al., 2020; Fig. S4, Fig. S5). To obtain complete species-trait datasets, we imputed missing trait values. Further, in order to assess the sensitivity of our models to variation in imputed values, we imputed the missing trait values eight independent times. This allowed us to assess the congruence of our model predictions when using the different imputed trait datasets in the analyses.

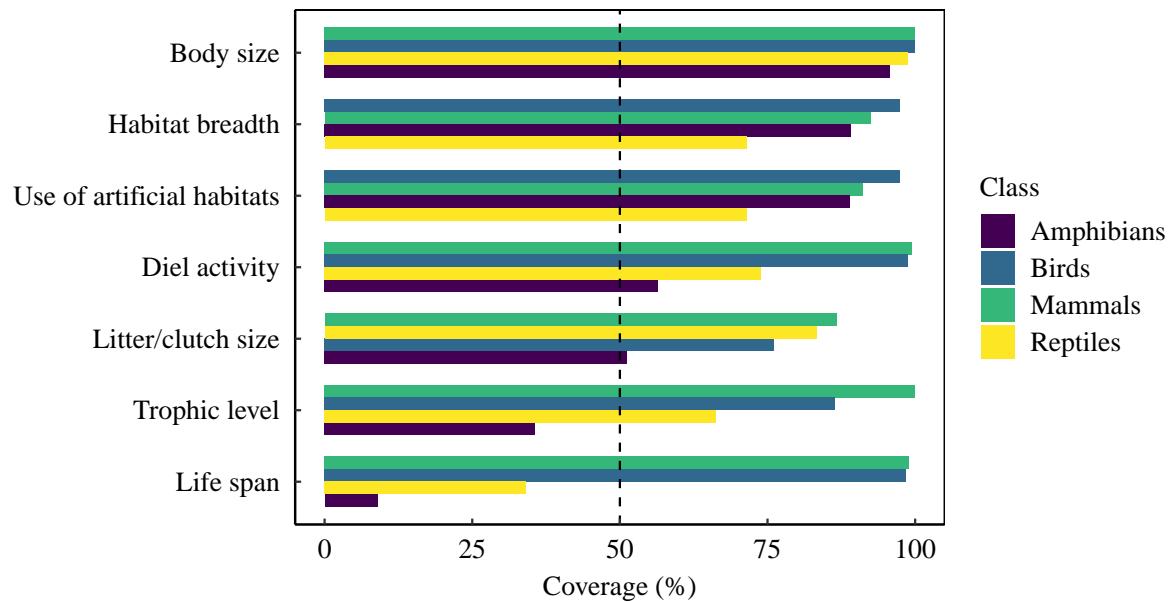


Figure S4: Trait coverage for the vertebrate species sampled in the PREDICTS database. For a given trait, coverage is calculated as the percentage of species for which an estimate was available.

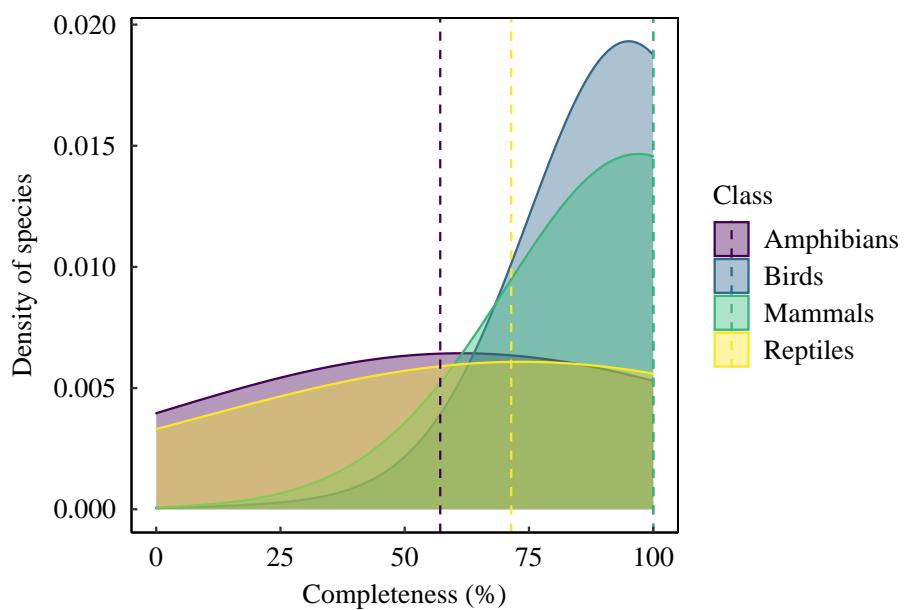


Figure S5: Distribution of trait completeness across the vertebrate species sampled in the PREDICTS database. For a given species, trait completeness is calculated as the proportion of traits for which an estimate was available. Dashed lines represent the median trait completeness.

14 **S2.1 Choice of imputation technique**

15 There exist several imputation techniques (Debastiani et al., 2021; Etard et al., 2020; Johnson et al., 2021;
16 Penone et al., 2014), such as K-nearest neighbour (Troyanskaya et al., 2001), multivariate imputation by
17 chained equations (van Buuren and Groothuis-Oudshoorn, 2011), random forest algorithms (implementable
18 in R with missForest; Stekhoven, 2016; Stekhoven and Bühlmann, 2012), and phylogenetic imputations (im-
19 plementable in R with PhyloPars, Bruggeman et al., 2009). Penone et al. (2014) assessed the performance
20 of these four techniques and showed that missForest and PhyloPars performed better when traits were phy-
21 logenetically conserved, and when the species phylogenetic position was included as a predictor of missing
22 trait values. PhyloPars can only handle continuous data, while missForest is compatible with mixed-type
23 (including categorical) data. When no phylogenetic information was included, mice was found to be the
24 best method, with fast imputations of mixed-type data (Penone et al., 2014). Therefore, to assess whether
25 missForest or mice was more appropriate here, we measured the phylogenetic signal in trait data. For con-
26 tinuous traits, we used Pagel's λ (Pagel, 1999), and for categorical traits we used Borges' δ (Borges et al.,
27 2018). Strong phylogenetic signal would indicate that traits are phylogenetically conserved, and hence miss-
28 Forest would be the most suited approach for imputing missing trait values, with the inclusion of species'
29 phylogenetic positions as a predictor.

30 **S2.2 Phylogenetic signal in traits**

31 Across all classes, similar traits were used for calculating functional diversity metrics: body mass, lit-
32 ter/clutch size, lifespan (using different proxies in different vertebrate classes: generation length for birds
33 and mammals, longevity for reptiles, and age at sexual maturity for amphibians), trophic level, diel activ-
34 ity, habitat breadth and use of artificial habitats. In addition, we included some class-specific traits for the
35 imputations, as certain class-specific traits could be useful predictors of other traits (such as body length
36 for instance in amphibians (Santini et al., 2018)). Table S2 details the traits that were included for the im-
37 putations in each class and the phylogenetic signal for each of these traits. Continuous traits were log-10
38 transformed before assessing Pagel's λ to improve normality. Pagel's λ was estimated using the phylosig
39 function of the phytools package (Revell, 2016), and Borges' δ was assessed using code provided by Borges
40 et al. (2018), available at : https://github.com/mrborges23/delta_statistic. To test for
41 the significance of δ , we generated null distributions of δ for each categorical trait by randomising trait vec-
42 tors 50 times, and calculating δ for each randomised vector – following the guidelines proposed by Borges
43 et al. (2018). We then tested whether the observed medians were greater than the null distributions using

44 one-sided Wilcoxon rank sum tests.

45 We used class-specific phylogenies to estimate phylogenetic signal, all downloaded on 13th April 2020.

46 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), we

47 downloaded trees from <https://data.vertlife.org/>. Trees were from Jetz et al. (2012) for birds,

48 from Jetz and Pyron (2018) for amphibians and from Tonini et al. (2016) for squamates. For each class, we

49 downloaded a distribution of 1,000 trees, from which we obtained consensus trees to estimate phylogenetic

50 signal (to that end, we used the TreeAnnotator programme of the BEAST software (Bouckaert et al., 2014)).

Table S2: 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 λ – 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 λ ; and a significant difference from the simulated null distribution of δ 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 λ								Borges' δ		
	BM	BL	GL	MA	ML	L	LCS	HB	TL	DA	UA
Amphibians	0.98*	0.94*	NA	0.85*	0.82*	NA	0.93*	0.99*	18*	3.4*	4.5*
Birds	0.99*	NA	0.97*	NA	NA	NA	0.95*	0.60*	13*	32e3*	1.8*
Mammals	0.99*	NA	0.97	NA	NA	NA	0.99	0.71	26*	17*	1.3*
Reptiles	1.0*	NA	NA	NA	0.94*	0.98*	1.0*	0.52*	6.3*	6.4*	1.4*

52 S2.3 Implementation of missForest imputations

53 As phylogenetic signals were strong in many categorical and continuous traits (Table S2), we 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.

60 Phylogenetic relationships were included as additional predictors in the form of phylogenetic eigen-

61 vectors (Diniz-Filho et al., 2012), extracted from the phylogenies using the PVR package (Santos, 2018).

62 Following Penone et al. (2014), we included the first ten phylogenetic eigenvectors as additional predictors

63 of missing trait values in each class, enough to minimise imputation error. As not all species were repre-

64 sented in the phylogenies, we also added taxonomic order as a predictor for all species. All traits in Table S2

65 were included in the imputations. Tuning parameters of missForest were set to ten maximum iterations and
66 to one hundred trees grown in each forest.

67 **S3 Degree of multicollinearity among traits**

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

Table S3: 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

⁸¹ **S4 Imputation performance**

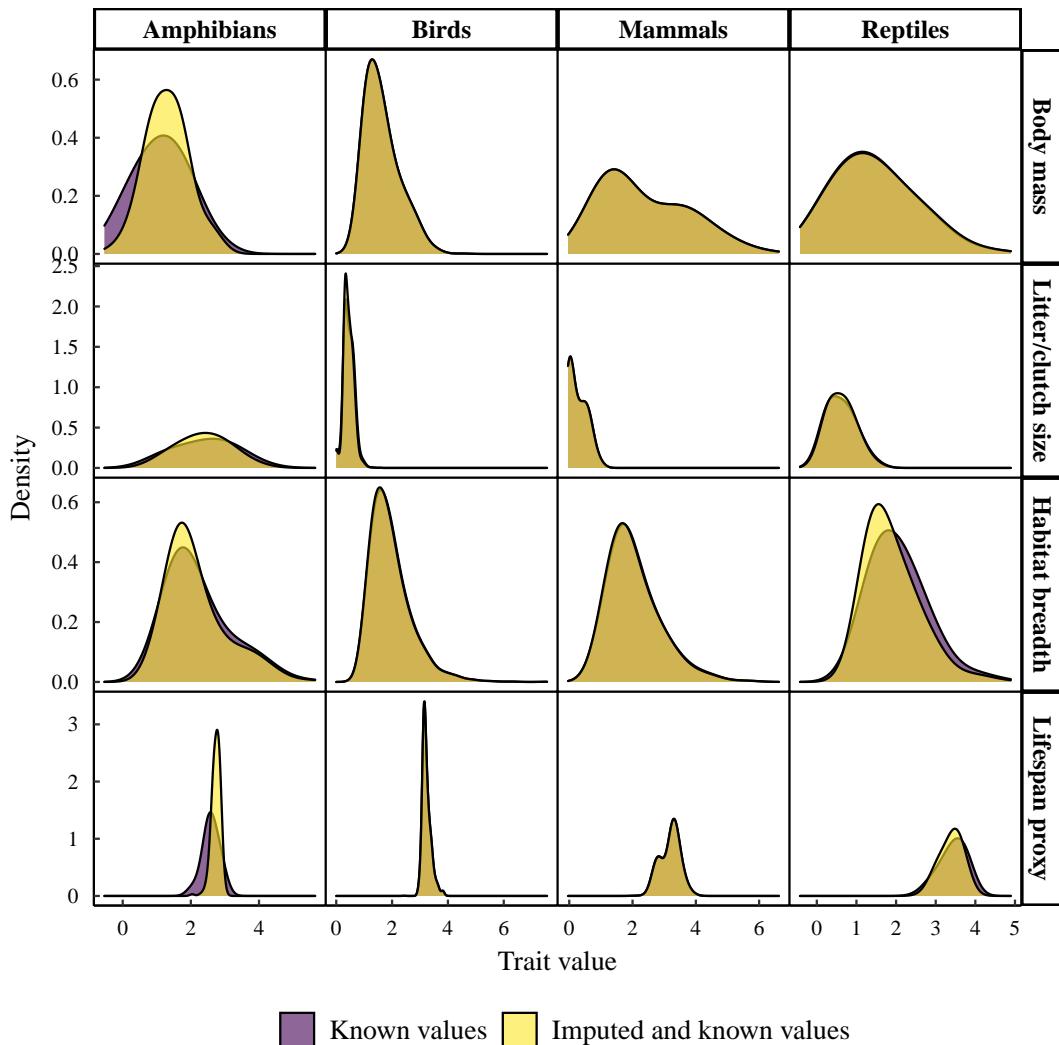


Figure S6: Distribution of continuous traits considered in the calculation of the functional diversity metrics (shown as density plots) before and after missing value imputations, in each class, for the species occurring in the PREDICTS database. All traits were log₁₀-transformed except Habitat breadth, which was square-rooted.

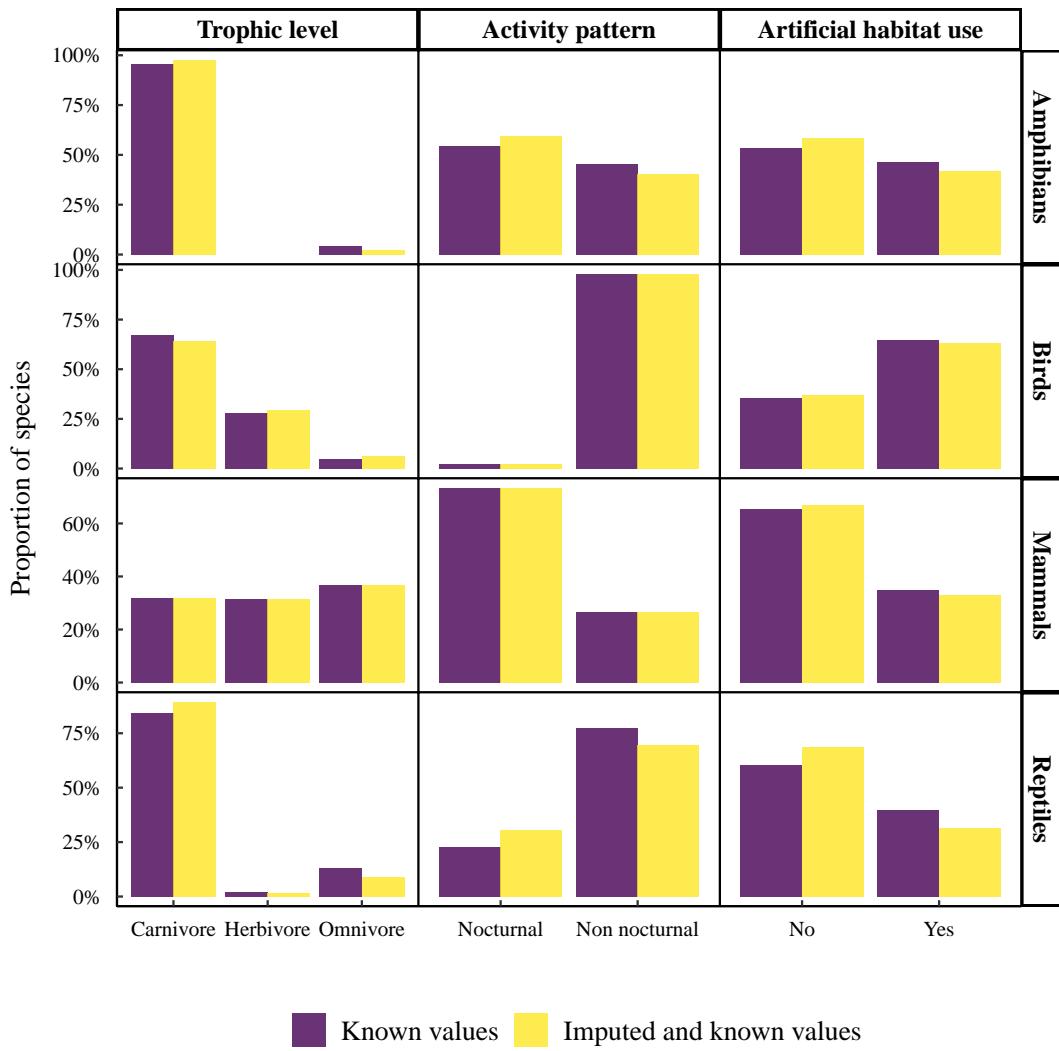


Figure S7: Frequency distribution for categorical traits considered in the calculation of the functional diversity metrics (shown as % of total species in each category) before and after missing value imputations, in each class, for the species occurring in the PREDICTS database.

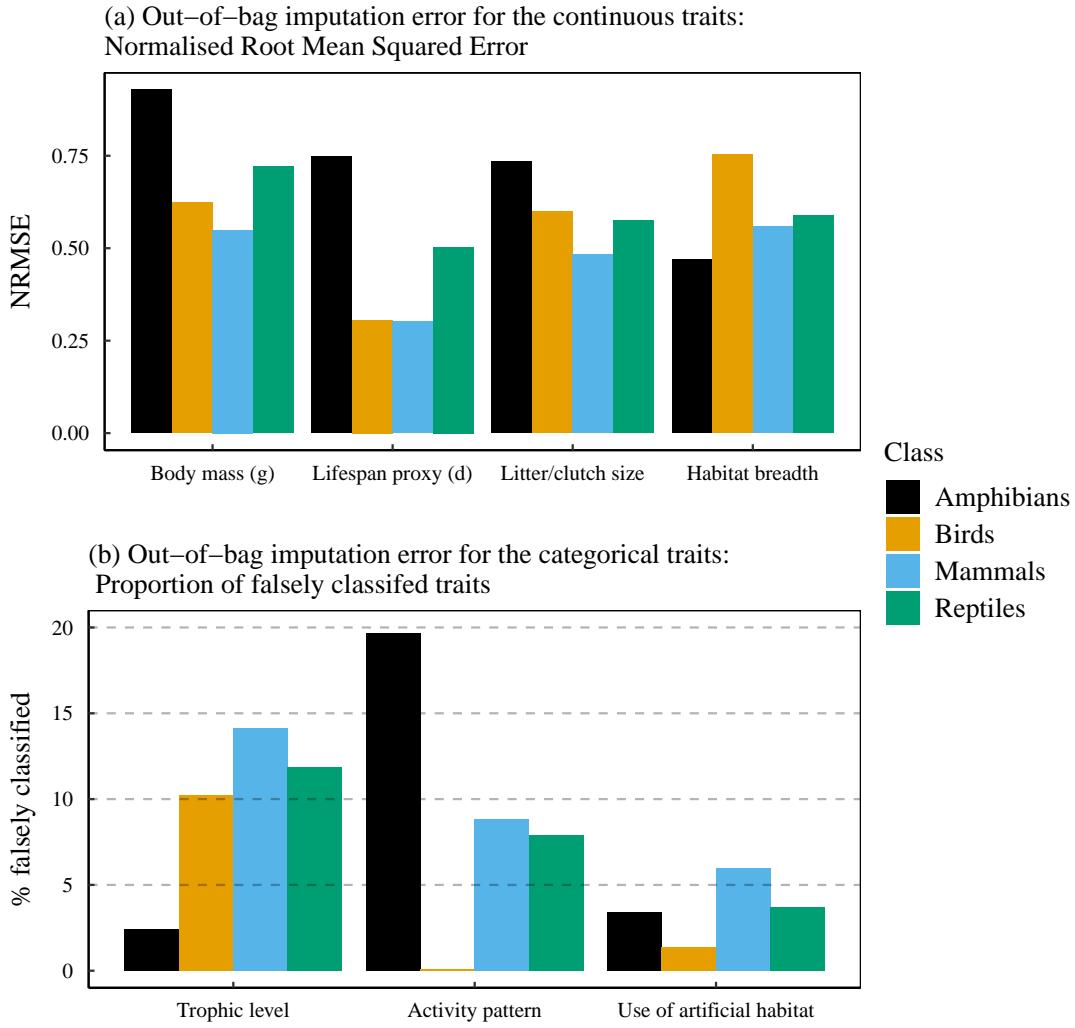


Figure S8: Out-of-bag imputation errors for the continuous traits (a) and categorical traits (b). For continuous traits, the Normalised Root Mean Squared Error (NRMSE) was obtained by dividing the Mean Squared Error (MSE) by the variance of the known trait distribution, then square-rooting the result. 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.

82 S5 Functional loss and functional gain

83 Across all vertebrates, we estimated functional loss and gain using 84 studies for the tropical subset and 39
84 studies for the temperate subset (51,514 and 30,470 pairwise comparisons between sites respectively, Table
85 S4). Because of this large number of pairwise comparisons, we did not develop a null modelling approach (if
86 we used 100 randomisations per pair of sites, we would need to compute functional loss and gain for more
87 than 8 million pairs, which would be very computationally demanding). We grouped mature, intermediate
88 and young secondary vegetation together in this analysis. We could not estimate the effects in all land uses
89 (for instance, sample sizes for tropical urban sites were too small).

90 Within classes, we used 18 tropical studies and 1 temperate study for amphibians; 38 and 21 for birds
91 (respectively); 28 and 9 for mammals; and 11 and 7 for reptiles. As sample sizes differed among pairs of land
92 uses and use intensities (Table S5), we were not able to estimates all effects, notably for the intensely-used
93 land uses.

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

Table S4: Sample sizes (number of pairs of sites) for the calculation of functional loss and functional gain across all vertebrates.

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 S5: 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	–

¹⁰⁴ **S6 Diagnostic plots**

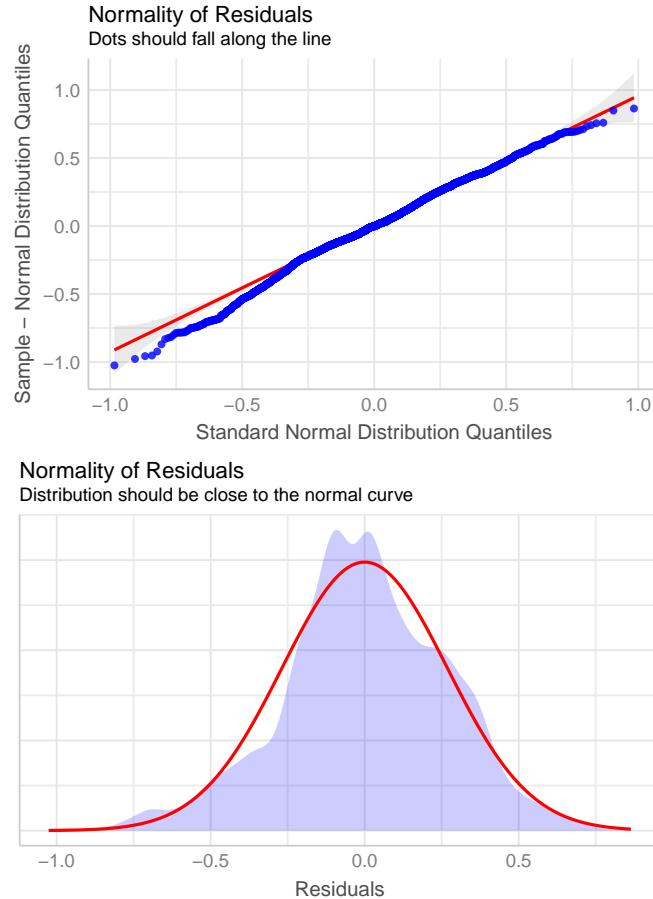


Figure S9: Diagnostic plots for Model 1a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

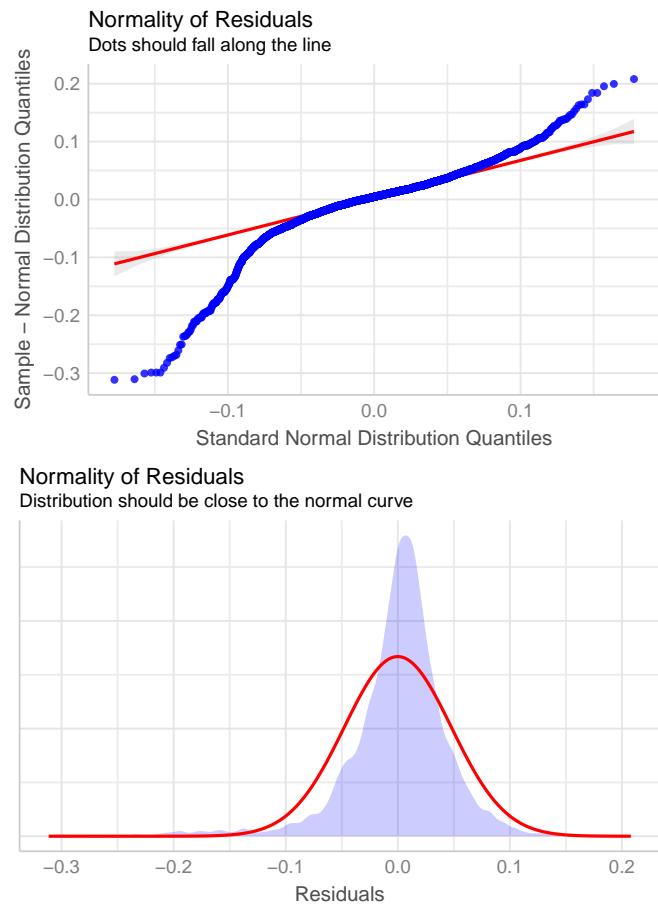


Figure S10: Diagnostic plots for Model 1b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

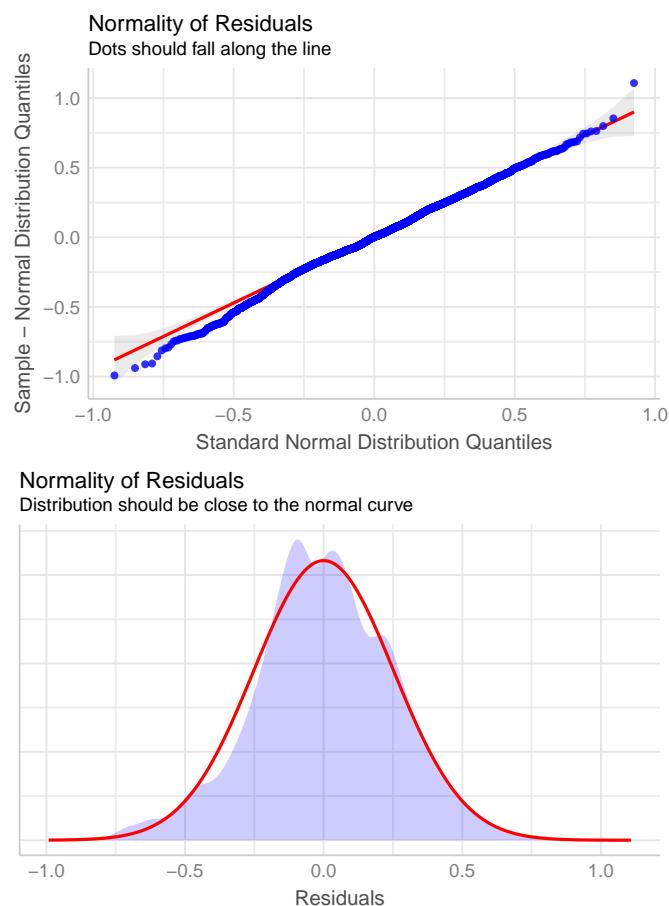


Figure S11: Diagnostic plots for Model 2a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

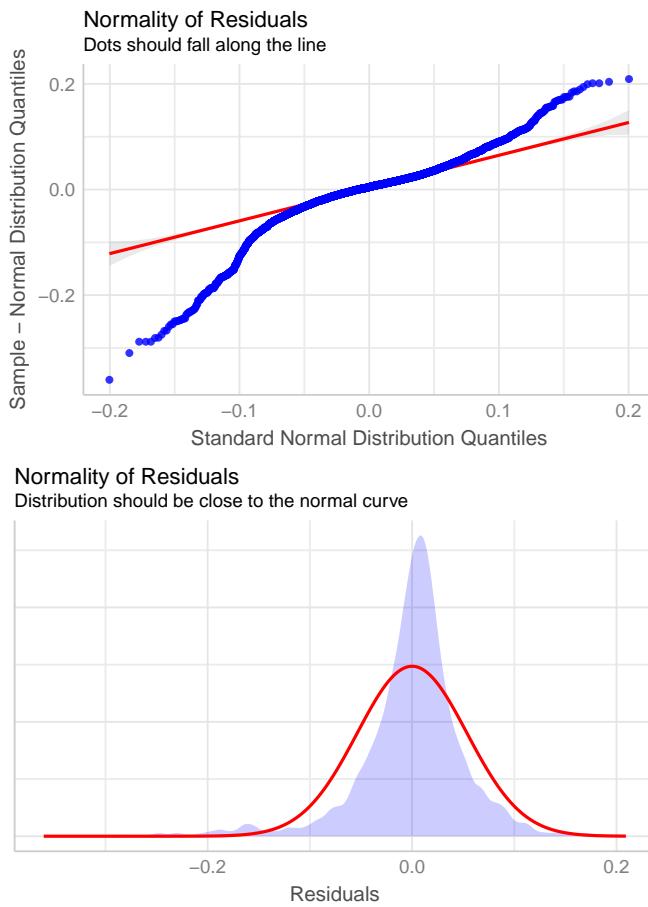


Figure S12: Diagnostic plots for Model 2b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

DHARMA residual diagnostics

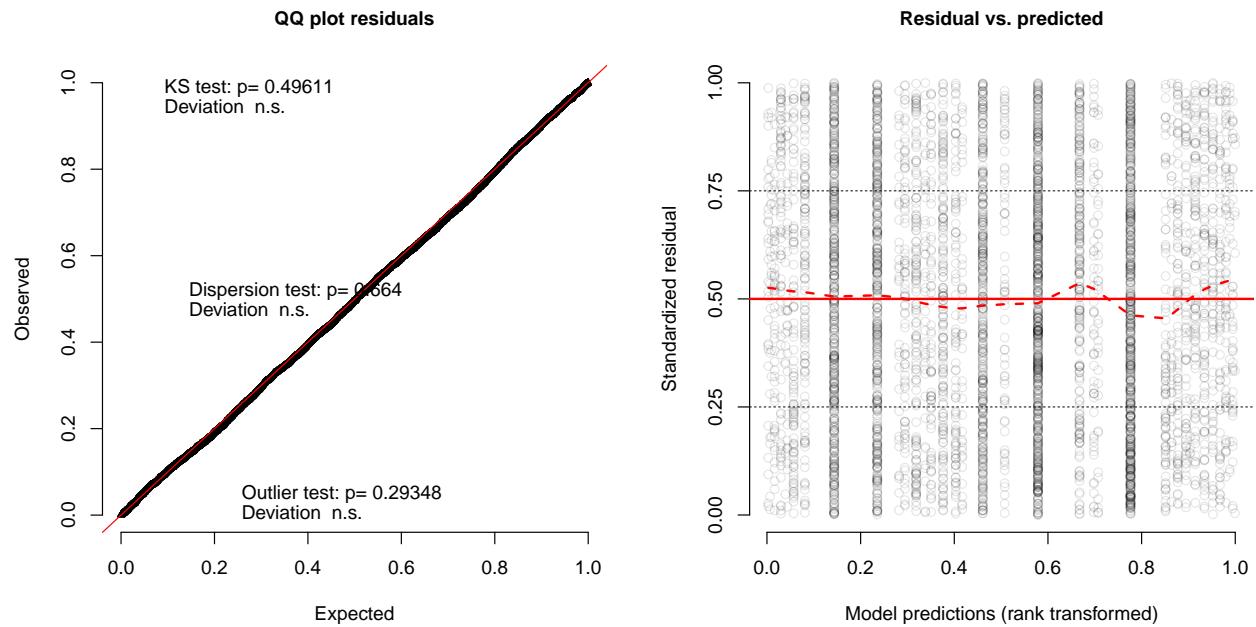


Figure S13: Diagnostic plots for Model 3, obtained using the ‘DHARMA’ R package (Hartig, 2021).

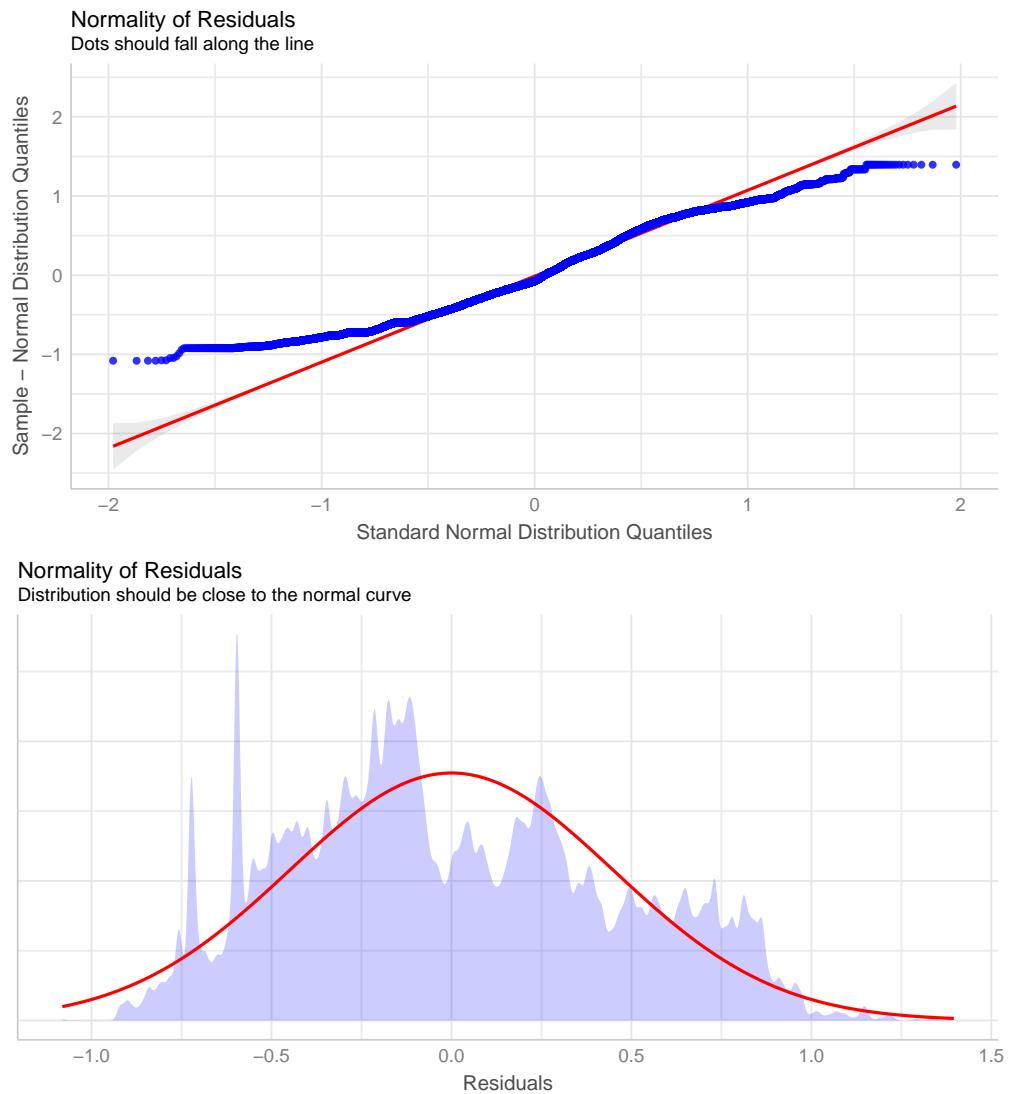


Figure S14: Diagnostic plots for Model 4a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

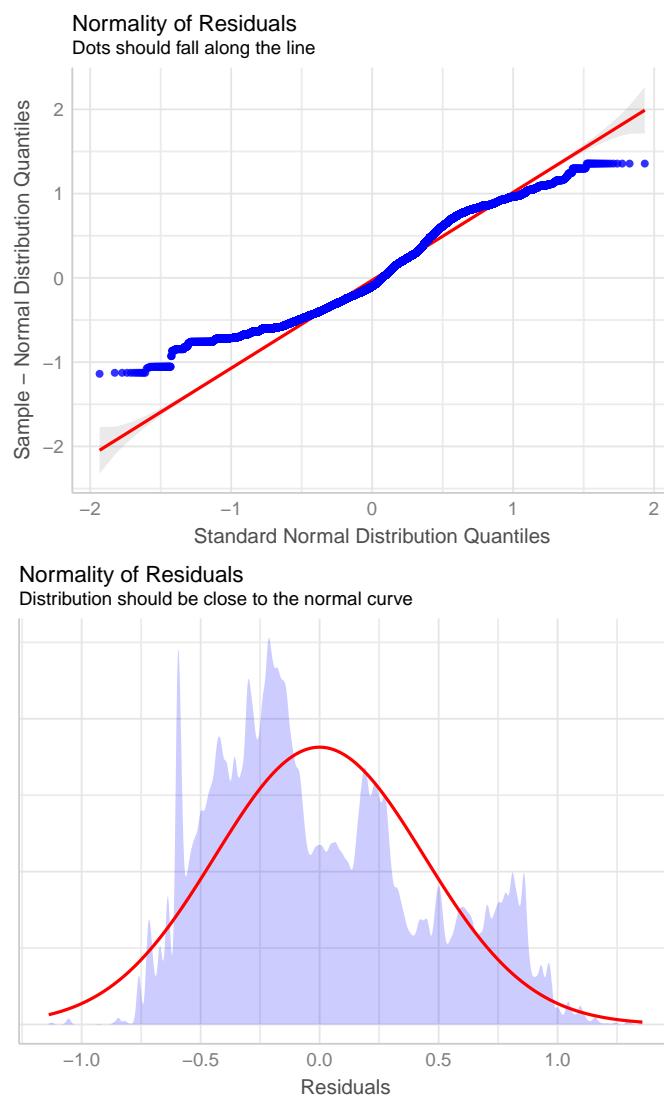


Figure S15: Diagnostic plots for Model 4b, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

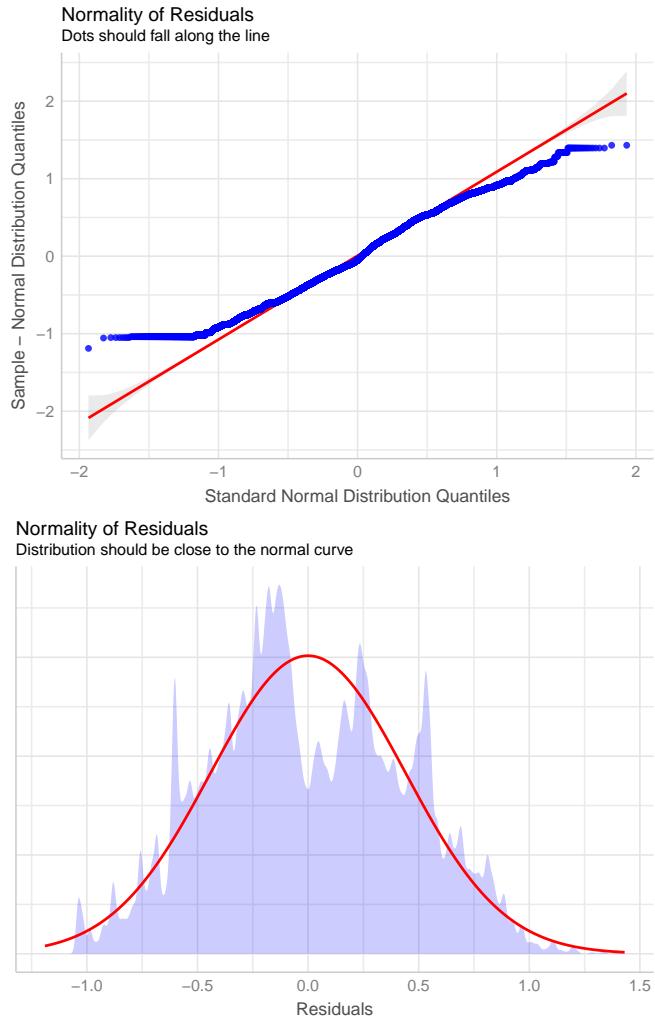


Figure S16: Diagnostic plots for Model 5a, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

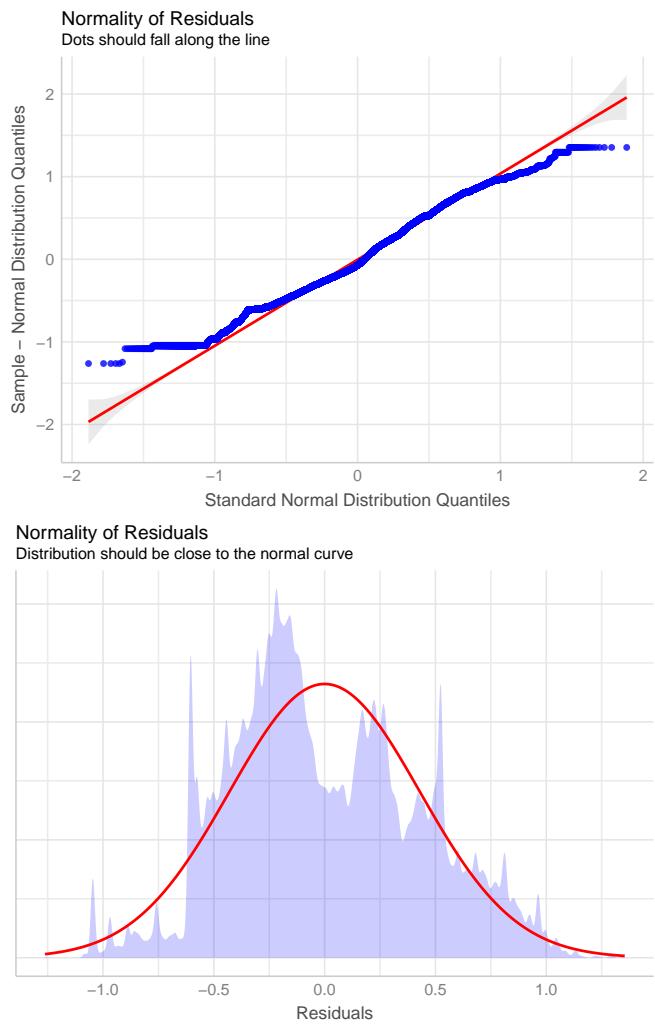


Figure S17: Diagnostic plots for **Model 5b**, obtained using the ‘performance’ R package (Lüdecke et al., 2021).

¹⁰⁵ **S7 Model robustness**

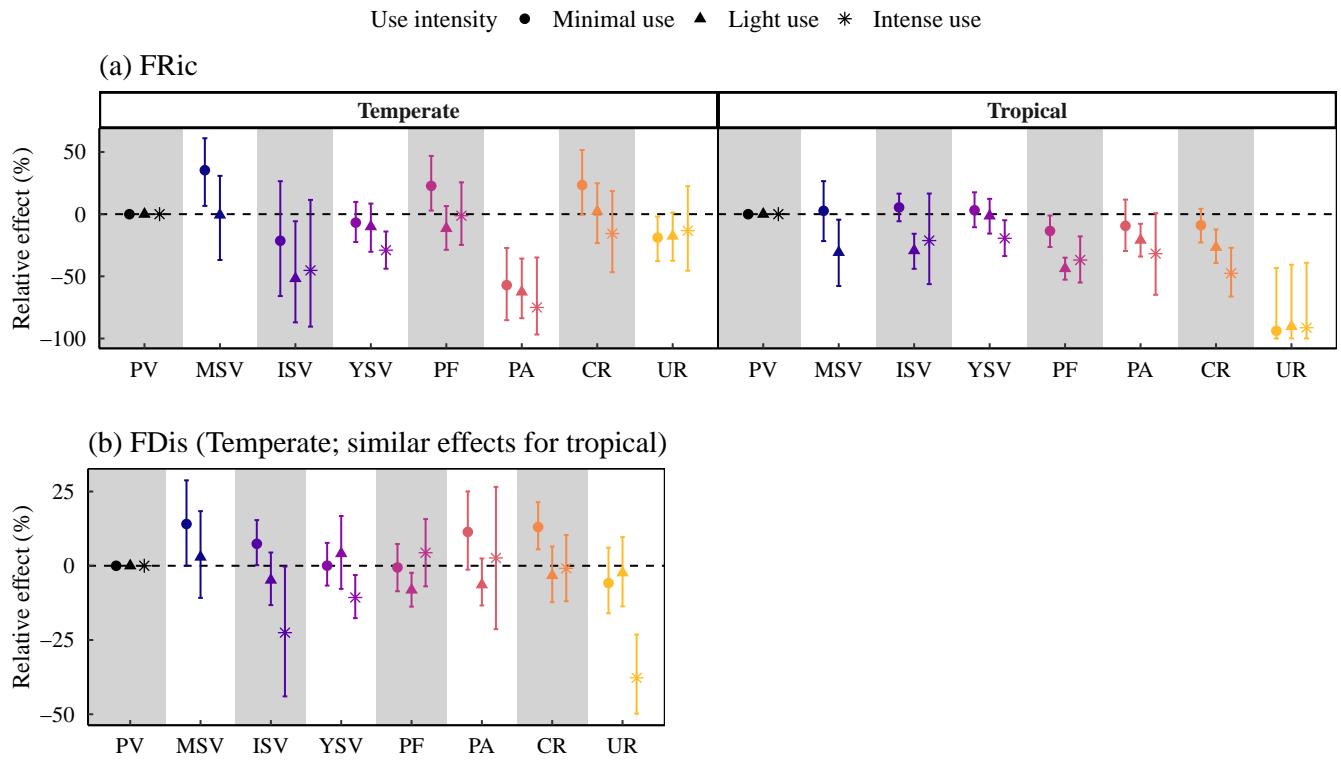


Figure S18: Effects of land use, use intensity and region on FRic (a) and FDis (b) across vertebrates, for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, we fitted Model 1a (see main text), which included the effects of land use, use intensity and region, as well as interactions between land use and use intensity and between land use and region. For FDis, we fitted Model 1b (see main text), which included effects of land use, use intensity and region, and interactions between land use and 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.

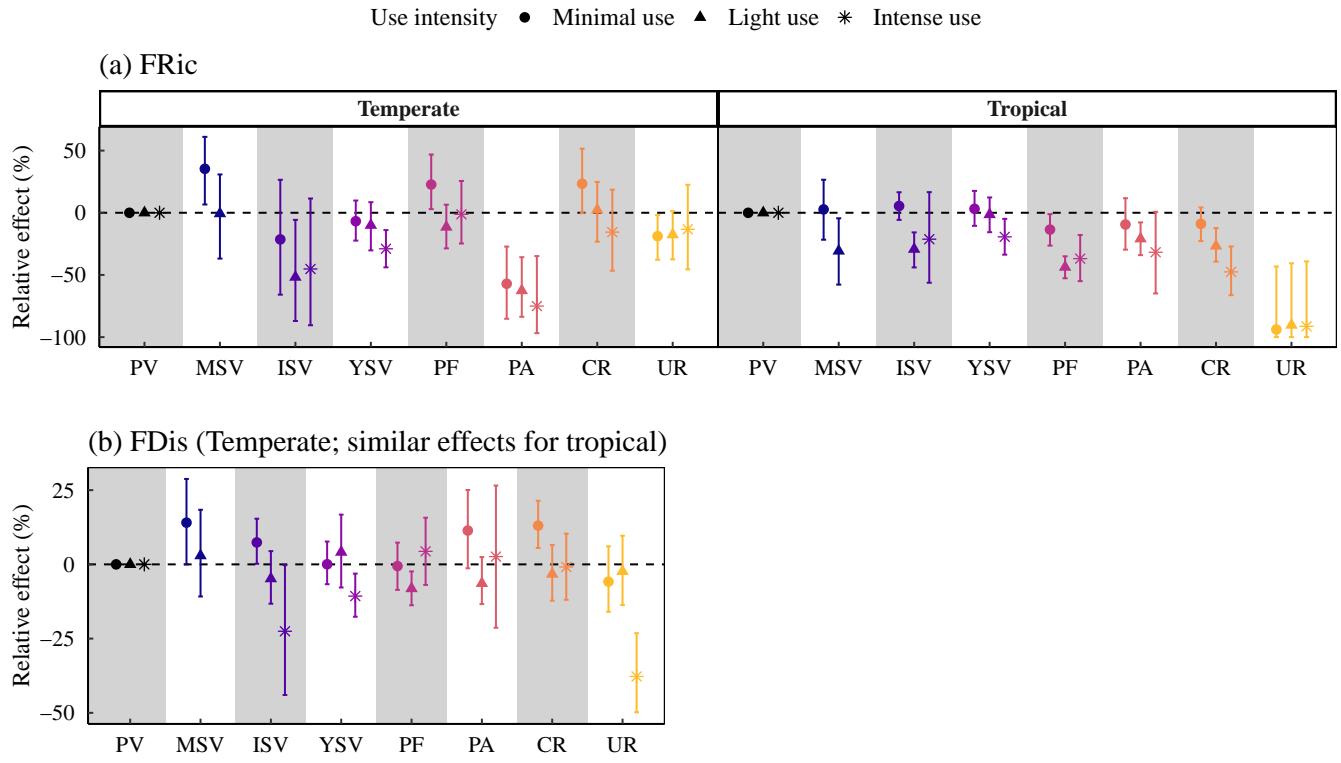


Figure S19: Effects of land use, use intensity and region on FRic (a) and FDis (b), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness), with geographical range size as an additional trait considered in the calculation of functional diversity metrics. Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, we fitted Model 1a (see main text), which included the effects of land use, use intensity and region, as well as interactions between land use and use intensity and between land use and region. For FDis, we fitted Model 1b (see main text), which included effects of land use, use intensity and region, and interactions between land use and 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.

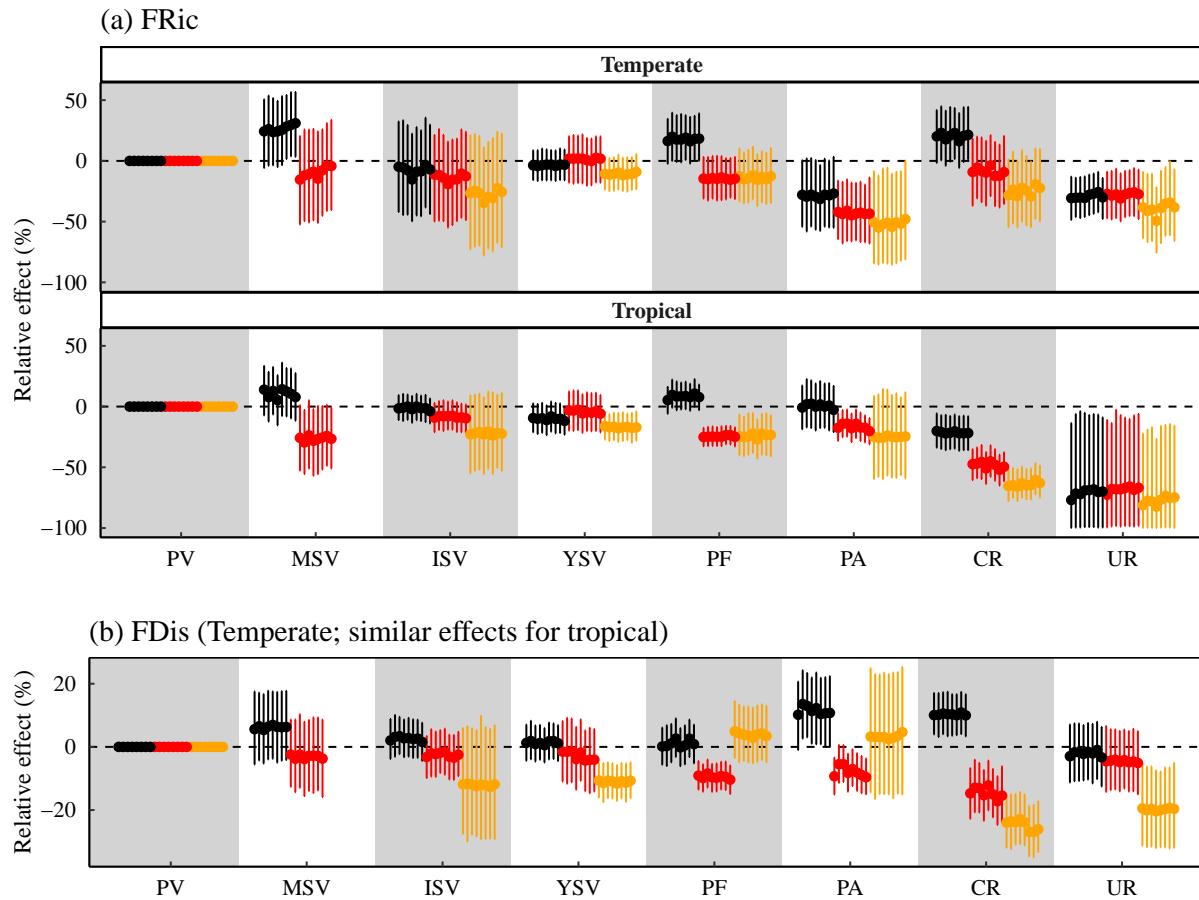


Figure S20: Effects of land use and use intensity on FRic (a) and FDis (b), obtained when calculating FRic and FDis with each set of imputed traits (eight in total). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, we fitted Model 1a (see main text), which included the effects of land use, use intensity and region, as well as interactions between land use and use intensity and between land use and region. For FDis, we fitted Model 1b (see main text), which included effects of land use, use intensity and region, and interactions between land use and 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.

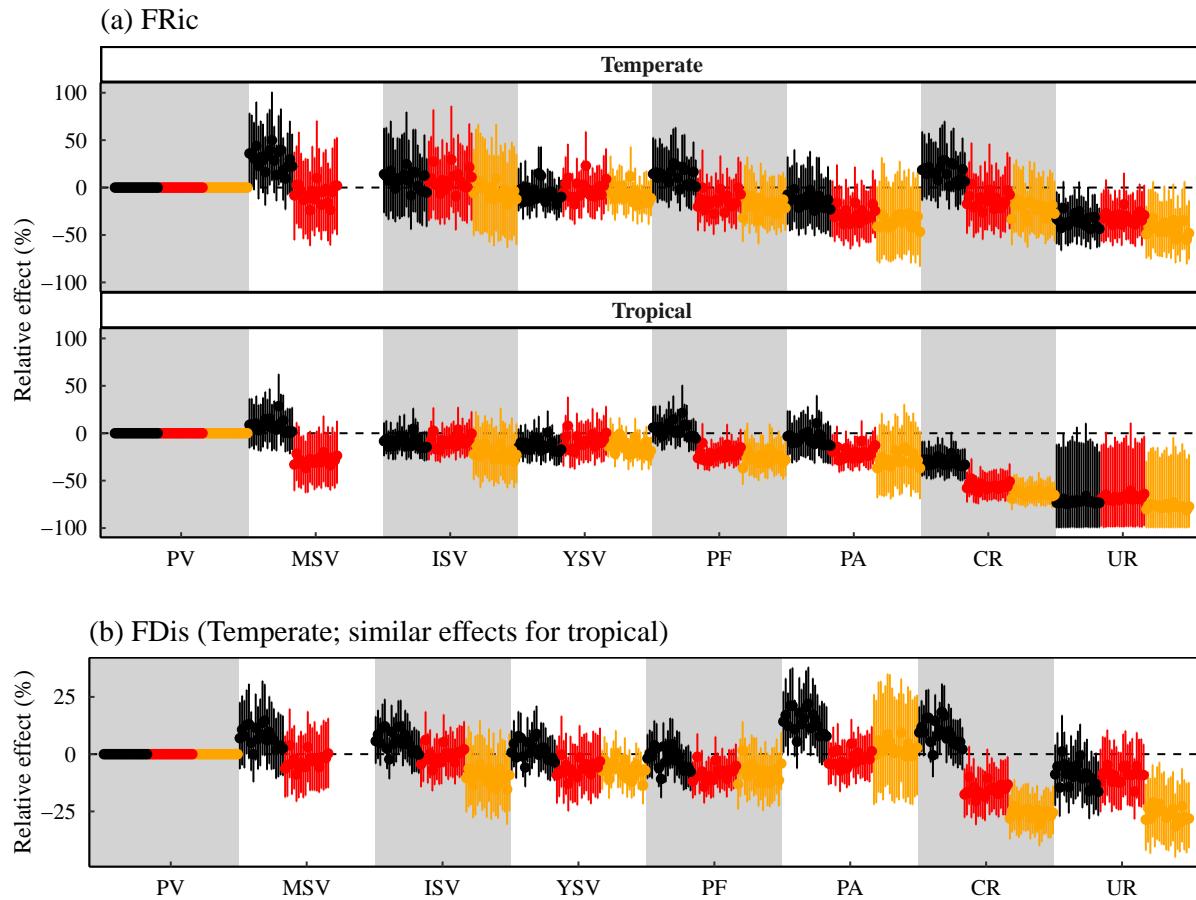


Figure S21: Effects of land use and use intensity on FRic (a) and FDis (b), obtained when re-sampling primary vegetation sites twenty independent times. We fixed the sample size for primary vegetation sites at 50 (see main text, Material and Methods, ‘Effects of land use and use intensity on FRic and FDis (hypothesis 1)’). For FRic, we fitted Model 1a (see main text), which included the effects of land use, use intensity and region, as well as interactions between land use and use intensity and between land use and region. For FDis, we fitted Model 1b (see main text), which included effects of land use, use intensity and region, and interactions between land use and 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.

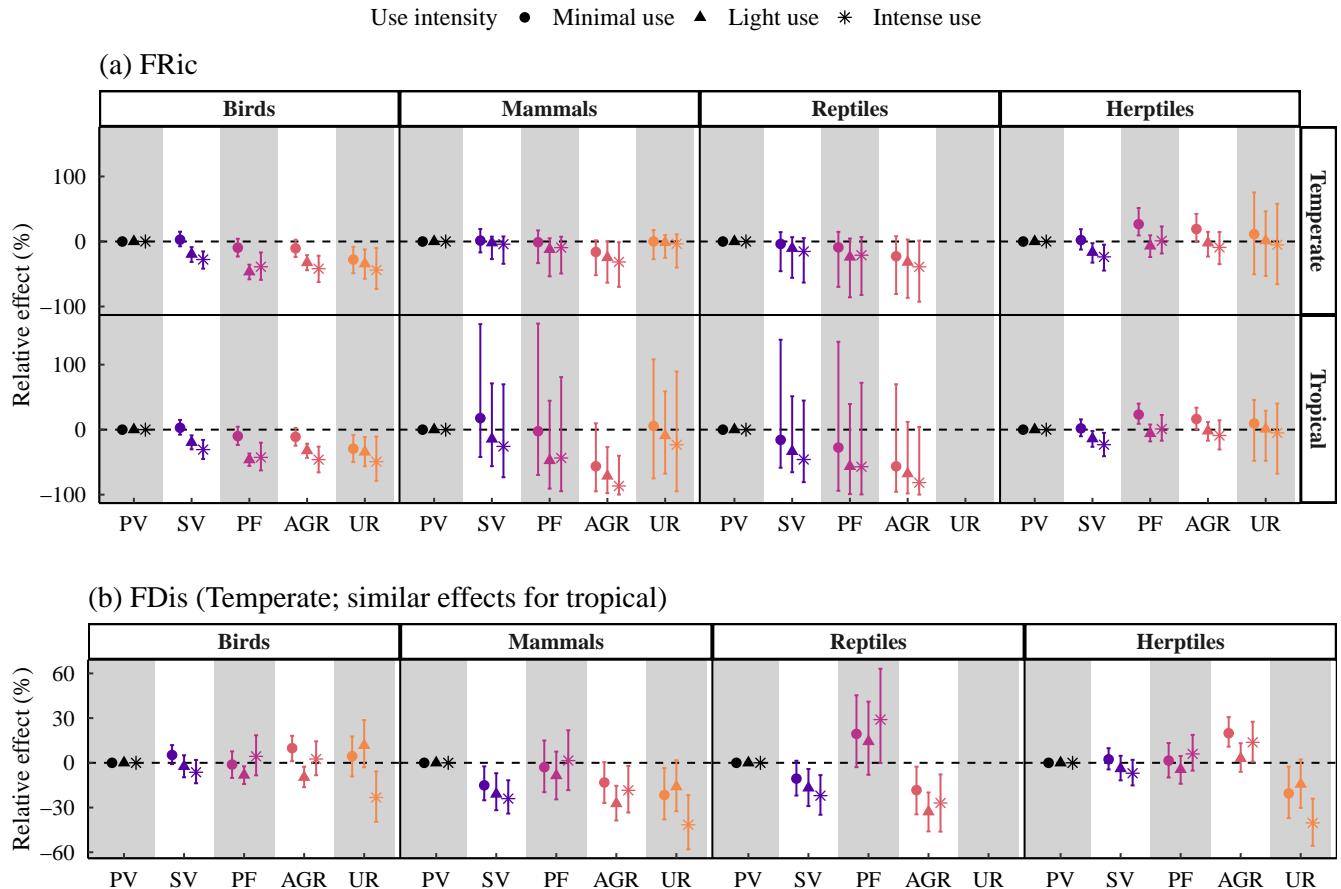
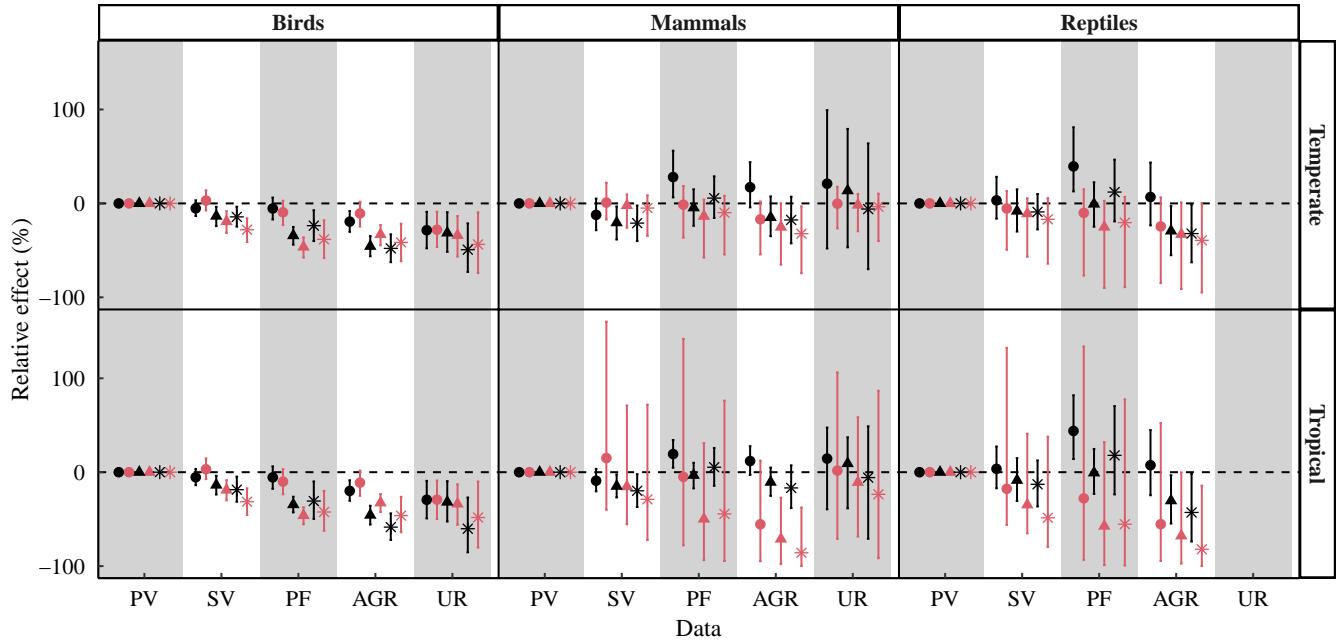


Figure S22: Effects of land use, use intensity and taxonomic class on FRic (a) and FDis (c), for the subset of species with complete trait data (i.e., excluding species with less than 100% trait completeness). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). We 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, use intensity and class, and interactions between land use and use intensity as well as land use and class. For FDis, the model included an additional interaction between 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

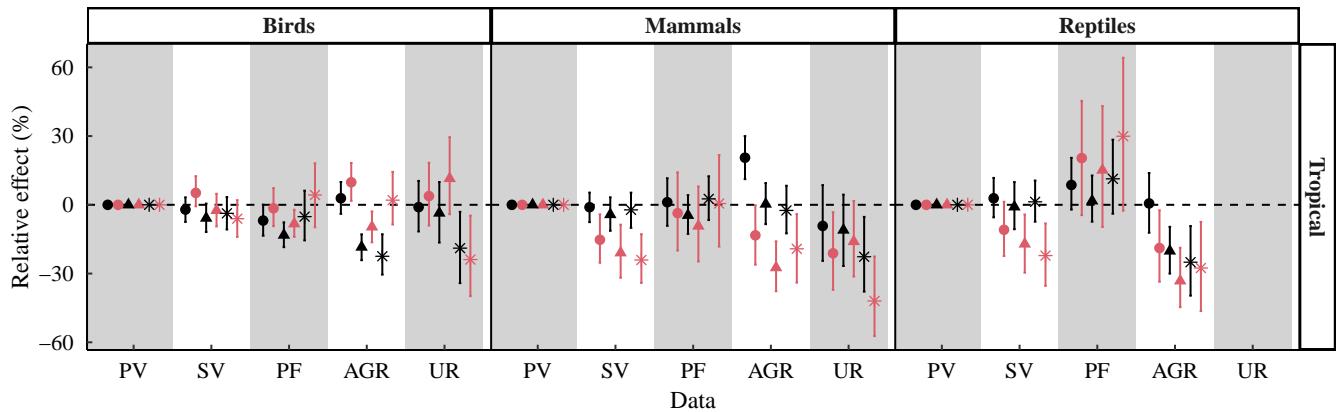
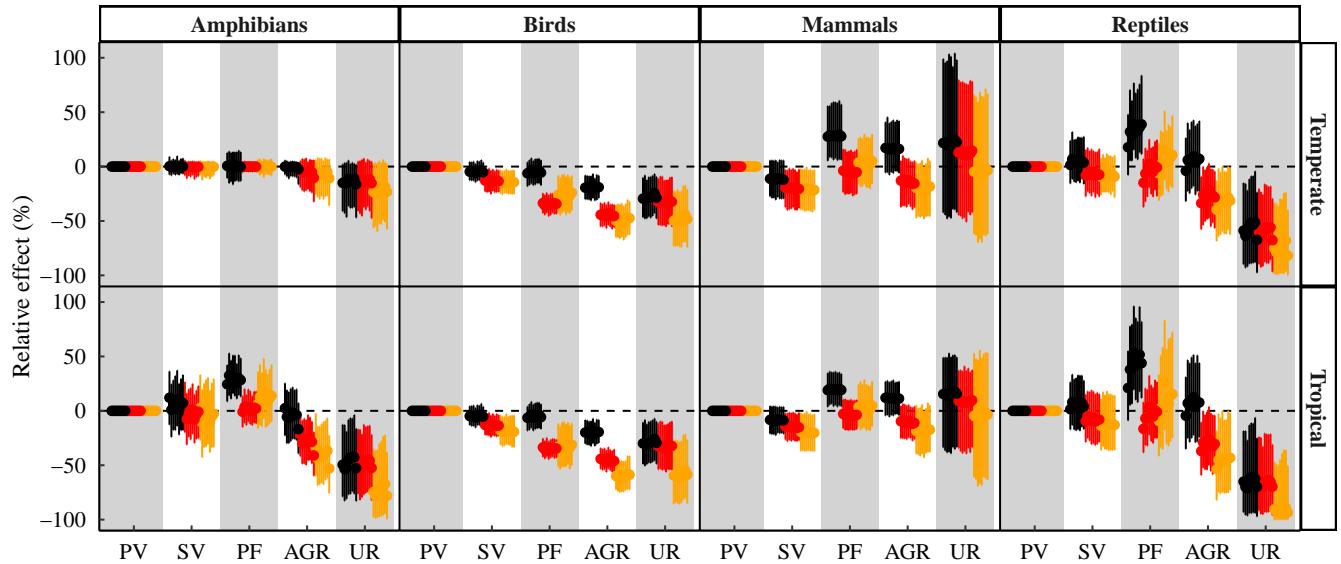


Figure S23: Effects of land use, region, use intensity and taxonomic class on FRic and FDis obtained with the imputed trait data (black points) or with the complete data subsets (red points). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, we fitted Model 2a (see main text), and we 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)

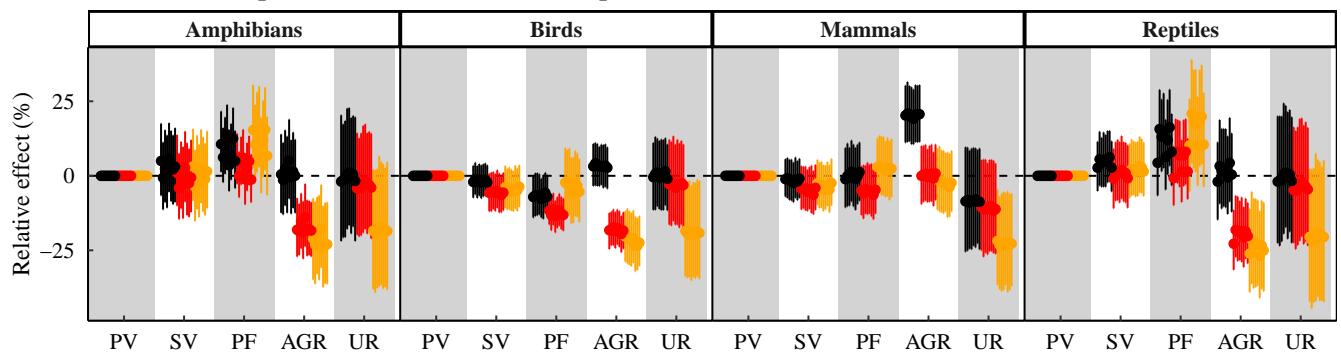


Figure S24: Effects of land use, region, use intensity and taxonomic class on FRic and FDis, obtained when calculating FRic and FDis with each set of imputed traits (eight in total). Effects are plotted as a % difference relative to the reference land use (primary vegetation, PV). For FRic, we fitted Model 2a (see main text), 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.

106 **S8 Model robustness – time since land-use conversion**

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

- 115 • For FRic, the best-fitting model includes the main effects of land use and time since land-use conver-
116 sion, but no interaction between these predictors. The model's summary (Table S6) show that time
117 since conversion has a significant negative effect on FRic, but the relationship between FRic and time
118 since land-use conversion is similar in different land uses (as there are no interactions between land
119 use and time since conversion, such that the slopes are similar in different land uses, and so the rate
120 at which FRic decreases with time is similar in different land uses). The intercept is only different for
121 urban land uses (significantly lower). Thus, based on this data subset, we expect time since land-use
122 conversion to have a similar effect in different land uses.

Table S6: Summary of the model explaining FRic by land use and time since land-use conversion, fitted on the subset of data for which we have 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

123 We then compare this model's predictions with a simpler model that doesn't account for time since
124 land-use conversion ($\text{FRic} \sim \text{Land use}$). The predictions (Fig. S25) show that including time since
125 land-use conversion doesn't bias our results, as we find a similar significant effect with both models in

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

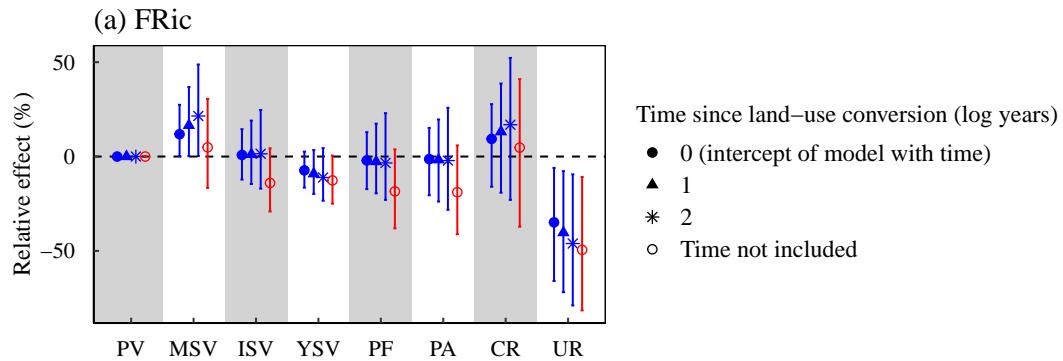


Figure S25: Effects of land use on FRic for the model that includes time since land-use conversion (blue points) versus the model that doesn't take time since land-use conversion into account (red points).

- 128 • For FDis, the best-fitting model includes the main effects of land use, time since land-use conversion
 129 as well as interactions between land use and time since land-use conversion (we didn't consider use
 130 intensity in the starting model because of sample size issues). Nevertheless, the main effect of time
 131 since land-use conversion is not significant (Table S7), and the relationship between time since land-
 132 use conversion and FDis is not significant in most land uses (except for plantation forest). Thus, we
 133 argue the available data don't allow us to properly investigate the relationship between time since
 134 land-use conversion and FDis.

Table S7: Summary of the model explaining FDis by land use and time since land-use conversion, fitted on the subset of data for which we 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

135 **Bibliography**

- 136 Borges, R., Machado, J. P., Gomes, C., Rocha, A. P., and Antunes, A. (2018). Measuring phylogenetic signal
137 between categorical traits and phylogenies. *Bioinformatics*. DOI: 10.1093/bioinformatics/bty800.
- 139 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., Suchard, M. A., Rambaut, A., and
140 Drummond, A. J. (2014). BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS
141 Computational Biology*. DOI: 10.1371/journal.pcbi.1003537.
- 142 Bruggeman, J., Heringa, J., and Brandt, B. W. (2009). PhyloPars: Estimation of missing parameter values
143 using phylogeny. *Nucleic Acids Research*. DOI: 10.1093/nar/gkp370.
- 144 Cadotte, M. W., Carscadden, K., and Mirochnick, N. (2011). Beyond species: Functional diversity and the
145 maintenance of ecological processes and services. *Journal of Applied Ecology*. DOI: 10.1111/j.1365-2664.2011.02048.x.
- 147 Cailliez, F. (1983). The analytical solution of the additive constant problem. *Psychometrika*. DOI: 10.
148 1007/BF02294026.
- 149 Debastiani, V. J., Bastazini, V. A. G., and Pillar, D (2021). Ecological Informatics Using phylogenetic infor-
150 mation to impute missing functional trait values in ecological databases. 63.April. DOI: 10.1016/j.
151 ecoinf.2021.101315.
- 152 Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., Morales-Castilla, I., Olalla-Tárraga, M. Á., Rodríguez,
153 M. Á., and Hawkins, B. A. (2012). On the selection of phylogenetic eigenvectors for ecological analyses.
154 *Ecography*. DOI: 10.1111/j.1600-0587.2011.06949.x.
- 155 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R., Gruber, B., Lafour-
156 cade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skid-
157 more, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and
158 a simulation study evaluating their performance. *Ecography*. DOI: 10.1111/j.1600-0587.2012.
159 07348.x.
- 160 Dray, S., Dufour, A.-B., and Chessel, D. (2007). The ade4 Package – II: Two-Table and K-Table Methods. *R
161 News*, 7.2, 47–52.
- 162 Etard, A., Morrill, S., and Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. *Global
163 Ecology and Biogeography*, November 2019, 1–16. DOI: 10.1111/geb.13184.
- 164 Faurby, S., Davis, M., Pedersen, R. Ø., Schowanek, S. D., Antonelli, A., and Svenning, J.-C. (2018).
165 PHYLACINE 1.2: The Phylogenetic Atlas of Mammal Macroecology. *Ecology*, 99.11, 2626. DOI: 10.
166 1002/ecy.2443.
- 167 Faurby, S., Pedersen, R. Ø., Davis, M., Schowanek, S. D., Jarvie, S., Antonelli, A., and Svenning, J.-C.
168 (2020). *MegaPast2Future/PHYLACINE_1.2: PHYLACINE Version 1.2.1*. DOI: 10.5281/zenodo.
169 3690867.
- 170 Fox, J. and Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical
171 Association*. DOI: 10.1080/01621459.1992.10475190.
- 172 Hartig, F. (2021). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Mod-
173 els*. R package version 0.4.1.

- 174 Hudson, L. N. et al. (2014). The PREDICTS database: A global database of how local terrestrial biodiversity
 175 responds to human impacts. *Ecology and Evolution*. DOI: 10.1002/ece3.1303.
- 176 Hudson, L. N. et al. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity
 177 In Changing Terrestrial Systems) project. *Ecology and Evolution*. DOI: 10.1002/ece3.2579.
- 178 Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., and Mooers, A. O. (2012). The global diversity of birds in
 179 space and time. *Nature*. DOI: 10.1038/nature11631.
- 180 Jetz, W. and Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with
 181 present imperilment across the amphibian tree of life. *Nature Ecology and Evolution*. DOI: 10.1038/
 182 s41559-018-0515-5.
- 183 Johnson, T. F., Isaac, N. J., Paviolo, A., and González-Suárez, M. (2021). Handling missing values in trait
 184 data. *Global Ecology and Biogeography*. DOI: 10.1111/geb.13185.
- 185 Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., and Makowski, D. (2021). performance: An R Pack-
 186 age for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*,
 187 6.60, 3139. DOI: 10.21105/joss.03139.
- 188 Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*. DOI: 10.1038/44766.
- 189 Penone, C., Davidson, A. D., Shoemaker, K. T., Di Marco, M., Rondinini, C., Brooks, T. M., Young, B. E.,
 190 Graham, C. H., and Costa, G. C. (2014). Imputation of missing data in life-history trait datasets: Which
 191 approach performs the best? *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.12232.
- 192 Revell, L. J. (2016). Package ‘phytools’. *R topics documented*.
- 193 Santini, L., Benítez-López, A., Ficetola, G. F., and Huijbregts, M. A. J. (2018). Length–mass allometries in
 194 amphibians. *Integrative Zoology*, 13.1, 36–45. DOI: 10.1111/1749-4877.12268.
- 195 Santos, T. (2018). Package ‘PVR’. *Phylogenetic Eigenvectors Regression and Phylogenetic Signal-Representation
 196 Curve*.
- 197 Stekhoven, D. J. (2016). Nonparametric Missing Value Imputation using Random Forest. *R Package version
 198 1.4*. DOI: 10.1093/bioinformatics/btr597.
- 199 Stekhoven, D. J. and Bühlmann, P. (2012). Missforest-Non-parametric missing value imputation for mixed-
 200 type data. *Bioinformatics*. DOI: 10.1093/bioinformatics/btr597.
- 201 Tonini, J. F. R., Beard, K. H., Ferreira, R. B., Jetz, W., and Pyron, R. A. (2016). Fully-sampled phylogenies
 202 of squamates reveal evolutionary patterns in threat status. *Biological Conservation*, 204, 23–31. DOI:
 203 <https://doi.org/10.1016/j.biocon.2016.03.039>.
- 204 Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., and Altman,
 205 R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*. DOI: 10.1093/
 206 bioinformatics/17.6.520.
- 207 van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations
 208 in R. *Journal of Statistical Software*, 45.3, 1–67.