Supporting information for

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

Adrienne Etard

May 7, 2019

Contents

List of Tables

List of Figures

1 Trait data compilation

1.1 Habitat affinities and broad specialisation

Habitat preferences. The Red List habitat data records habitat types in which species occur. Habitats are classified into 96 categories, which I pooled into 13 broader habitat variables: Forest, Savanna, Grassland, Shrubland, Wetland, Rocky areas, Caves and subterranean, Desert, Marine, Marine intertidal or coastal/supratidal, Artificial, Introduced vegetation and Other/Unknown. Species habitat preferences were described using these variables as binary (taking 1 if a species was known to occur in the habitat and 0 otherwise).

Habitat breadth. Habitat breadth was calculated as the number of habitats recorded to be used by a species in the Red List database. Given that information regarding habitat suitability and habitat importance was also available in the Red List data files, I used a weighted sum to calculate habitat breadth. Suitability was declined in three categories in the Red List files: 'suitable', 'marginal' or 'unknown'. Habitats were recorded to be either of major importance, not of major importance or of unknown importance. I used the weights provided in Table 1 to produce weighted sums of the number of habitats used by each species. A comparison of the distribution of habitat breadths calculated with and without weights shows that weighting did not have a strong impact on the results (Figure 1).

Table 1: Weights used in the calculation of habitat breadth. Habitat breadth was calculated as the weighted sum of the number of habitats used by a species. Weights were assigned to each habitat given its importance and its suitability.

Suitability	Major importance		
	Yes	No	Unknown
Suitable	1	0.5	1
Marginal	0.3	0.3	0.3
Unknown	1	0.3	1

Degree of specialisation. A broad classification was adopted for species degree of specialisation. Using IUCN habitat files, I determined whether species were strictly natural habitat specialists or generalists. Generalists were species for which any habitat, suitable or of unknown suitability, was recorded to be artificial. Else, species were considered to be natural habitat specialists. When a habitat of an unknown type was considered suitable or was of unknown suitability, missing data was introduced.
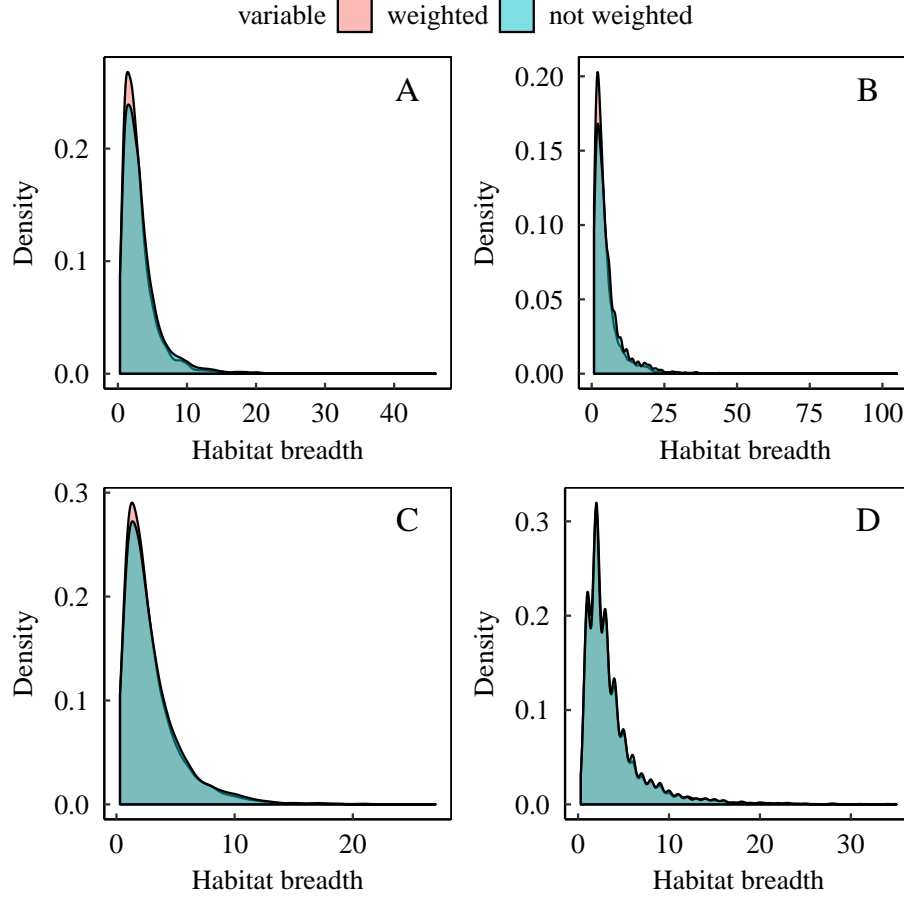
Figure 1: Distribution of habitat breadths across species for terrestrial vertebrates. The distribution is shown for calculations assigning weights to habitats according to their suitability and their importance (red surface area) or without weights (blue surface area). The distribution of habitat breadths slightly shifts to the left when a weighted sum is used (in **A**, **B** and **C**), as less importance is accorded to some habitats in the calculation (Table 1).

1.2 Tackling taxonomic synonymy

1.2.1 Dropping replicated phylogenetic tips

2.6% of mammalian, 1.5% of avian, 1% of amphibian and 1.5% of reptilian species had multiple multiple phylogenetic positions after taxonomic corrections. Most of these species had only two replicates (as most species with synonyms were found to have two identified synonym, c.f. main text). Table 2 provides the number of replicates across replicated species. Replicated tips corresponding to species described in the format *Genus cf.*, *Genus sp.*, *Genus spp.* or *Genus aff.* were all conserved in the phylogenies.

Table 2: Number of replicated tip labels in phylogenies, number of which are sister clades and identified species for which dropping tips at random was problematic. Most redundant tips appeared twice. When replicated tips were sister clades, tips to drop were chosen randomly; this did not affect the species phylogenetic position. When replicated tips were not sister clades, I verified whether the corrected tip name appeared in the original, uncorrected tree. If so, I kept the tip in the corrected tree whose position was closest to the position of the tip in the original tree. Note that I used tip order in that case, which is sensitive to branch permutation. Finally, for replicated tip labels that were neither sister clades nor figuring in the original tree, tips to drop were chosen randomly. I identified these as problematic cases, although occurrences were rare. See Figure for a visual representation of each case.

Class	Replicates			Sister clades	In uncorrected tree	'Problematic'
	2	**3**	**>3**			
Mammals	141	8	2	29	143	*Heterogeomys cherriei* *Heterogeomys dariensis* *Hylopetes sagitta* *Marmosa paraguayana* *Neoromicia brunnea* *Neoromicia malagasyensis* *Plecturocebus discolor* *Proechimys trinitatis*
Birds	158	7	0	21	160	*Antrostomus arizonae* *Calendulauda erythrochlamys* *Myiothlypis rivularis* *Spermestes bicolor*
Reptiles	68	2	0	17	69	*Salvator merianae*
Amphibians	41	4	2	8	44	*Lithobates berlandieri* *Uperodon taprobanicus*

Figure 2 describes the procedure I used. If replicated tips were sister clades, the tip to conserve was chosen randomly among the replicates. Else, I chose to conserve the tree tip whose position was closest to the position of the same tip in the uncorrected tree, when present. In all other few cases, tips to drop were chosen randomly. See Figures 3, 4 and 5 for an example of each case.
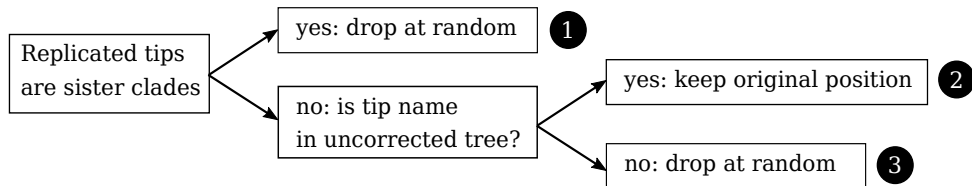
Figure 2: Procedure followed to drop replicated tips from phylogenies. Most redundant tip labels were replicated twice. **(1)** When replicated tips were sister clades, the tips to drop were chosen randomly, as it did not affect the 'true' phylogenetic position of the species; see case study 1 in Figure 3. **(2)** When replicated were not sister clades, I kept the tip whose position was closest to the position of the same tip in the uncorrected tree (Figure 4). **(3)** In a few cases, the corrected name did not appear in the original tree. Those were problematic cases, and the tips to drop were chosen randomly (Figure 5). Nevertheless, occurences of that third case were rare.

1.2.2 Case studies

Case study 1 (Figure 3): replicated tips are sister clades. In this example, the Przewalski's toadhead agama (*Phrynocephalus przewalskii*) figures in the original, uncorrected tree under two names identified as being synonyms (*P. przewalskii* and *P. frontalis*). As such, two replicated tips with the same name appear in the tree after the taxonomic correction. As they are sister clades, one tip is randomly dropped in the final corrected tree. In that case, the 'true' phylogenetic position of the species is not affected.

Figure 3: Phylogenetic trees before (A) and after (B) taxonomic correction, and after removing replicated tips (C) for species in the reptilian genus *Phrynocephalus*. The identified accepted name *P. przewalskii*, in blue (A), is identified as being a synonym for *P. frontalis*, which also figures in the uncorrected tree. As these clades are sister in the tree, one tip is dropped at random, and the final phylogenetic position of the species is not affected.

Case study 2 (Figure 4): replicated tips are not sister clades, but the accepted name (tip label) figures in the original, uncorrected tree. In that example, the amphibian species *Ambystoma californiense* (California tiger salamander) was also found under its identified synonym *A. tigrinum* in the original tree. In the final tree, the tip closest to the position of the same tip in the uncorrected tree was kept.
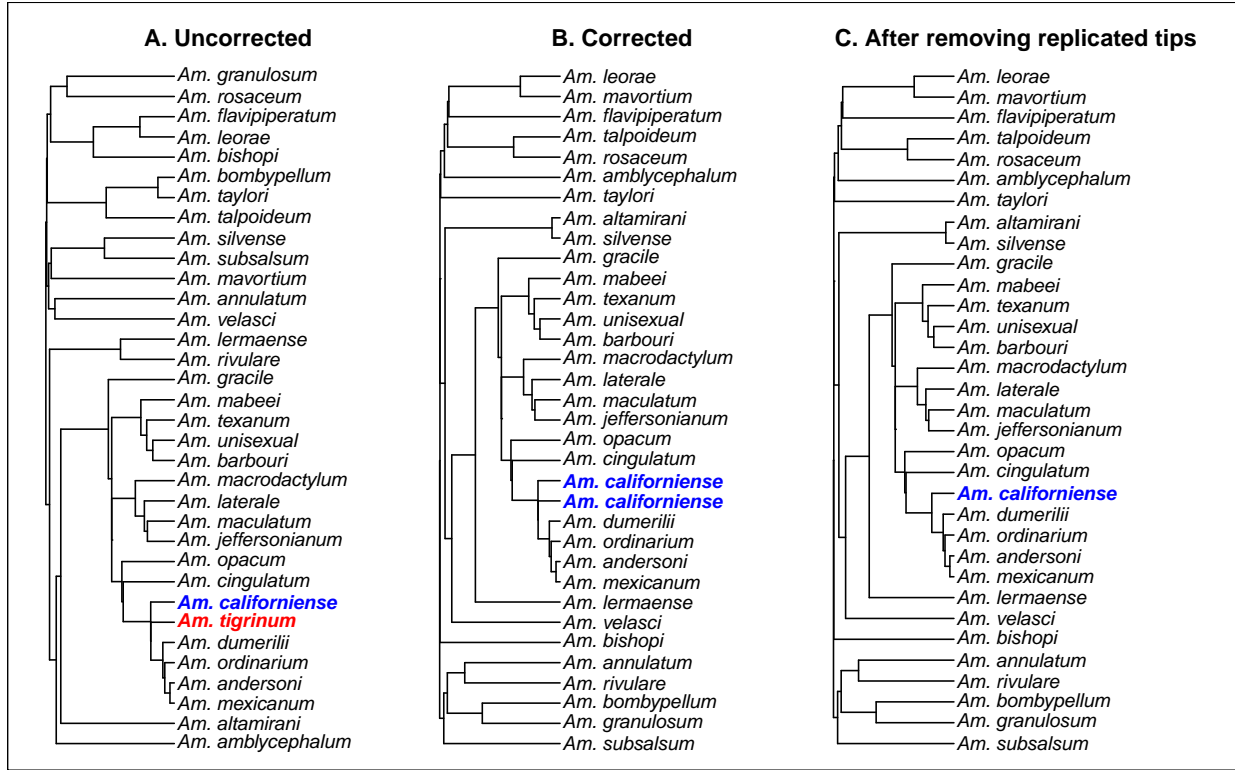
Figure 4: Phylogenetic trees before (A) and after (B) taxonomic correction, and after removing replicated tips (C) for species in the amphibian genus *Ambystoma*. The identified accepted name *A. californiense*, in blue (A), is identified as being a synonym for *A. tigrinum*, which also figures in the uncorrected tree. There are not sister clades; in the final tree, the tip closest to the position of the same tip in the uncorrected tree was kept.

9

Case study 3 (Figure 5): replicated tips are not sister clades, and the accepted name (tip label) does not figure in the original, uncorrected tree. These were the problematic species listed in Table 2. In this example, the mammalian species *Petinomys sagitta* (arrow flying squirrel) and *Hylopetes lepidus* (gray-cheeked flying squirrel) have been identified to be synonymic, but the accepted name does not figure in the original tree. They are not sister clades; hence, one tip is randomly chosen to be dropped.
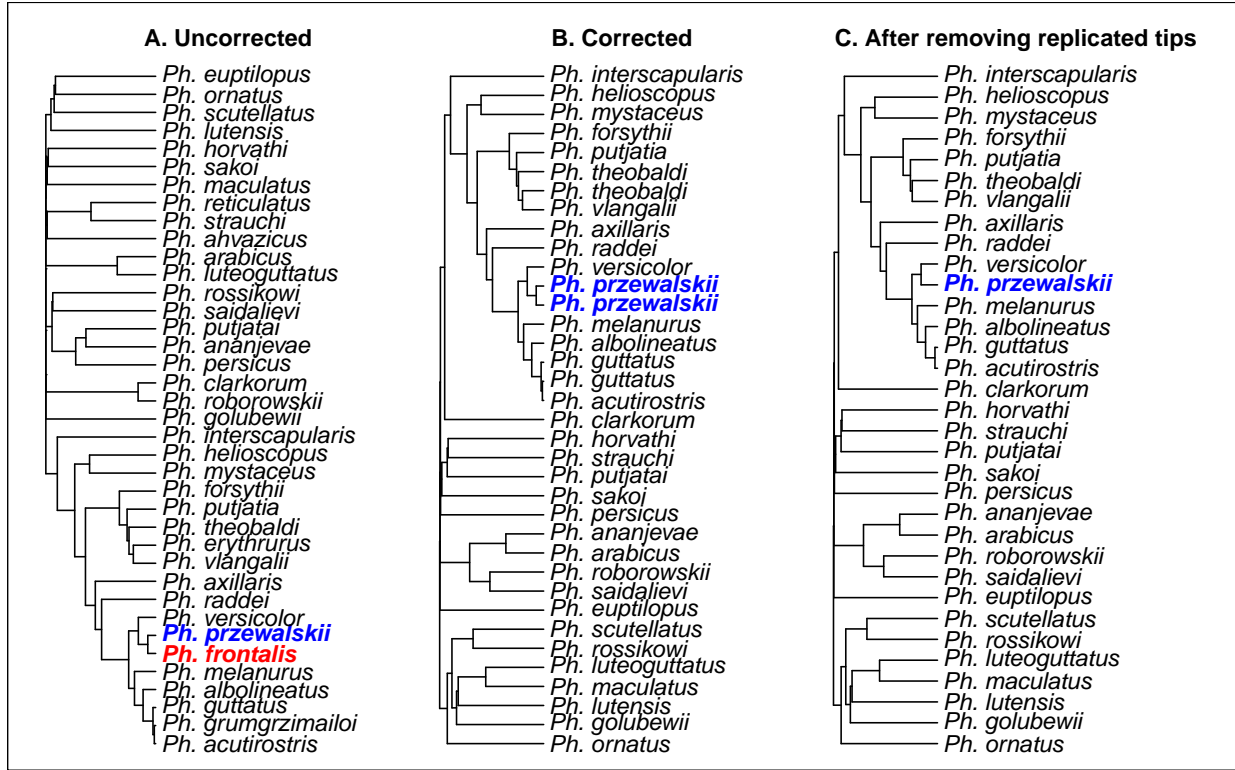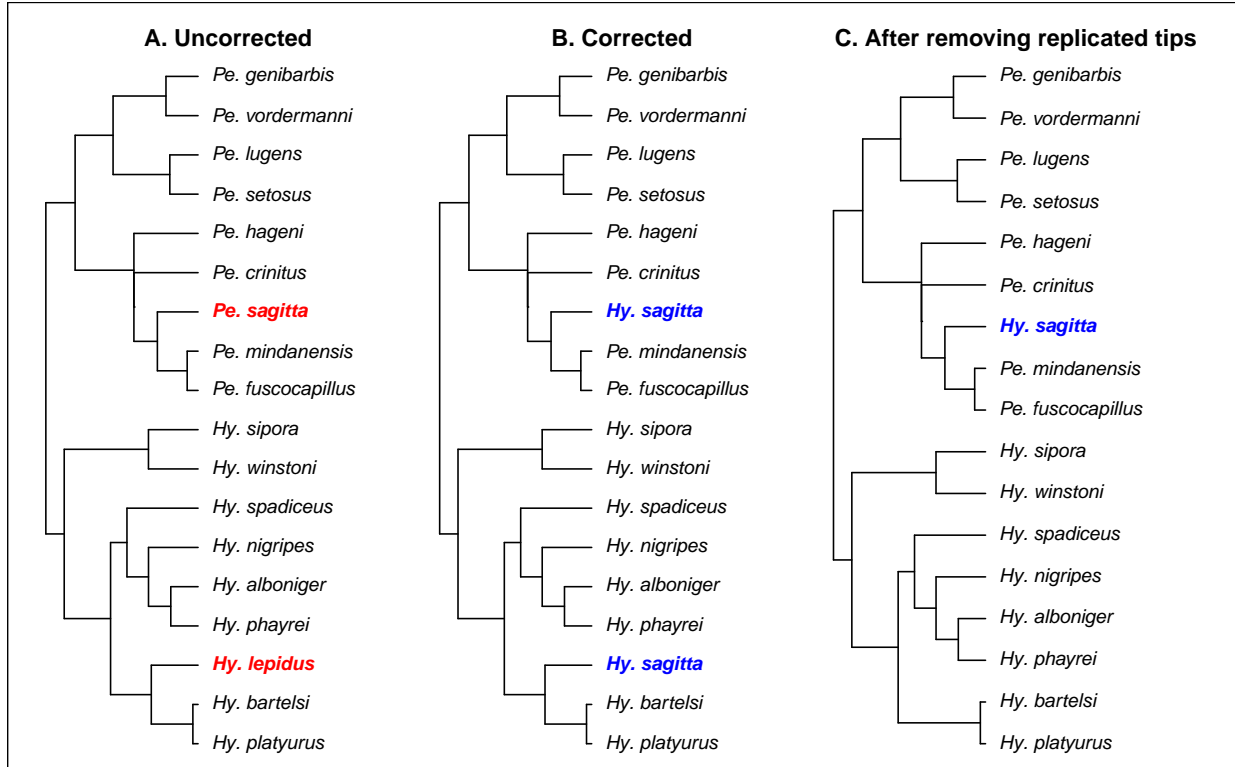

Figure 5: Phylogenetic trees before (A) and after (B) taxonomic correction, and after removing replicated tips (C) for species in the mammalian genera *Hylopetes* and *Petinomys*. The identified accepted name *H. sagitta*, in blue (B), does not figure in the uncorrected tree. It was identified to be the corresponding accepted name for *P. sagitta* and *H. lepidus*, which are not sister clades. One tip is randomly chosen to be dropped. Such cases were problematic but occurences were rare (Table 2).

2 Trait coverage and completeness for PREDICTS species
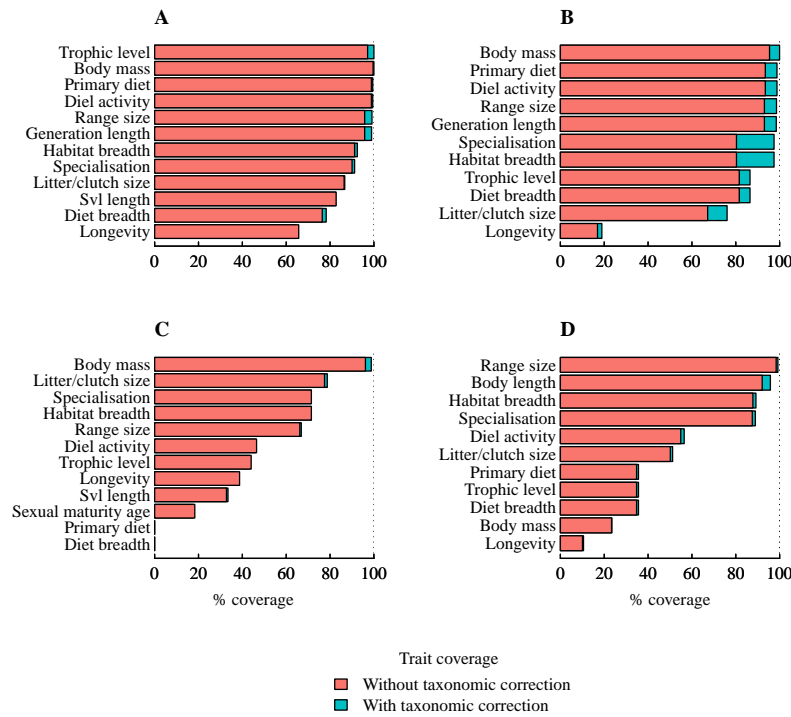
Figure 6: Trait coverage for species figuring in the PREDICTS database only.
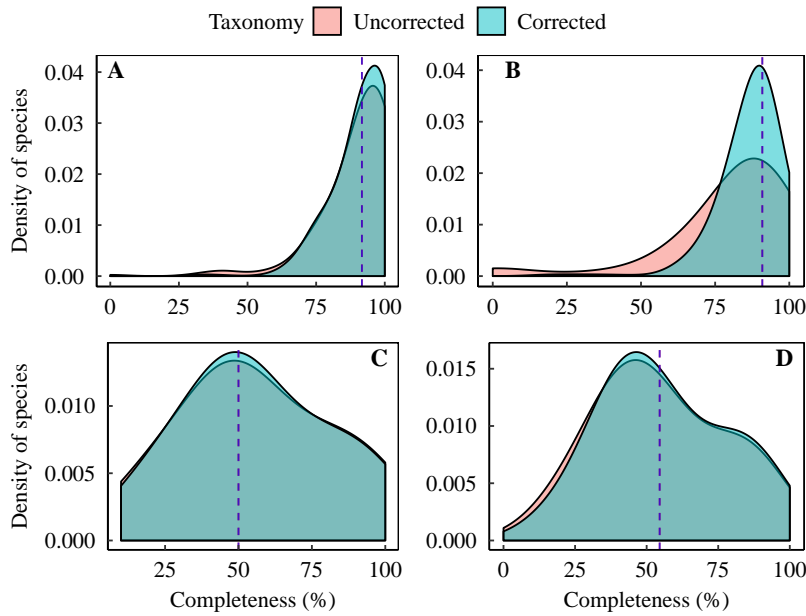
Figure 7: Distribution of completeness values across species figuring in the PREDICTS database only.

3 Missing values imputation

3.1 Phylogenetic signal

3.1.1 Phylogenetic signal computed with the original phylogenies

Table 3: Phylogenetic signal in continuous and categorical traits and in range size, calculated with original phylogenies. **BM**: body mass; **L**: longevity; **LCS**: litter/clutch size; **HB**: habitat breadth; **DB**: diet breadth; **GL**: generation length; **BL**: body length; **SM**: sexual maturity; **RS**: range size; **TL**: trophic level; **PD**: primary diet; **DA**: diel activity; **Sp**: specialisation. The phylogenetic signal in continuous traits was calculated with Pagel's λ. For categorical traits, the δ metric developed by Borges et al (2018) was used. A star indicates a significant signal (significant p-values scores for the log-likelihood ratio test in the case of λ; and significant difference from the simulated null distribution of δ for categorical traits, see SI for test results). 'na' are introduced for traits that were not considered in a class but may have been used in another as a predictor in missing values imputations. All traits showed significant phylogenetic signal, with signals for BM, L, LCS, and GL being particularly strong in mammals and birds (above 0.9). Here all calculations were conducted with the original phylogenies.

Class	Continuous target traits, additional predictors and range size: λ									Categorical traits: δ			
	BM	**L**	**LCS**	**HB**	**DB**	**GL**	**BL**	**SM**	**RS**	**TL**	**PD**	**DA**	**Sp**
Mammals	0.99*	0.92*	0.97*	0.71*	0.99*	0.98*	0.99*	na	0.75*	16*	46*	17*	1.4*
Birds	0.98*	0.85*	0.94*	0.57*	0.49*	0.97*	na	na	0.69*	10*	18*	$33{\cdot}10^3$*	1.3*
Reptiles	0.93*	0.92*	0.88*	0.47*	na	na	0.99*	0.95*	0.75*	4.4*	na	6.3*	1.8*
Amphibians	0.97*	0.76*	0.94*	0.84*	0.60*	na	0.96*	na	0.83*	2.6*	3.1*	1.7*	2.3*

Table 4: Wilcoxon rank sum test result for the significance of phylogenetic signal in categorical traits.
The results are shown here, for phylogenetic signal (δ) computed with both original and modified phylogenies.

Class	Trait	delta	p-value	Phylogeny
Birds	Specialisation	1.6	3.90E-10	Modified
Birds	Trophic level	9.98	2.91E-11	Modified
Birds	Diel activity	28469.32	3.90E-10	Modified
Birds	Primary diet	17.59	1.91E-06	Modified
Reptiles	Specialisation	1.54	3.90E-10	Modified
Reptiles	Trophic level	4.33	1.91E-06	Modified
Reptiles	Diel activity	7.08	3.90E-10	Modified
Mammals	Specialisation	1.38	3.90E-10	Modified
Mammals	Trophic level	16.96	1.14E-13	Modified
Mammals	Diel activity	18.72	3.90E-10	Modified
Mammals	Primary diet	49.61	1.86E-09	Modified
Amphibians	Specialisation	3.36	3.90E-10	Modified
Amphibians	Trophic level	18.17	9.31E-10	Modified
Amphibians	Diel activity	2.85	3.90E-10	Modified
Amphibians	Primary diet	3.68	9.54E-07	Modified
Birds	Specialisation	1.27	3.90E-10	Original
Birds	Trophic level	10.57	2.27E-13	Original
Birds	Diel activity	33504.1	3.90E-10	Original
Birds	Primary diet	18.49	0.00012	Original
Reptiles	Specialisation	1.77	3.90E-10	Original
Reptiles	Trophic level	4.4	5.82E-11	Original
Reptiles	Diel activity	6.35	3.90E-10	Original
Mammals	Specialisation	1.4	3.90E-10	Original
Mammals	Trophic level	16.3	5.68E-14	Original
Mammals	Diel activity	17.14	3.90E-10	Original
Mammals	Primary diet	45.64	0.00098	Original
Amphibians	Specialisation	2.27	3.90E-10	Original
Amphibians	Trophic level	2.62	1.46E-11	Original
Amphibians	Diel activity	1.73	3.90E-10	Original
Amphibians	Primary diet	3.14	1.16E-10	Original

Table 5: χ-squared test result for the significance of phylogenetic signal in continuous traits. The results are shown here, for phylogenetic signal (λ) computed with **original** phylogenies.

Class	Trait	lambda	p-value	Phylogeny
Birds	Body mass	9.8E-01	0.0E+00	Original
Birds	Longevity	8.5E-01	2.0E-222	Original
Birds	Litter/clutch size	9.4E-01	0.0E+00	Original
Birds	Diet breath	4.9E-01	2.7E-156	Original
Birds	Range size	6.9E-01	0.0E+00	Original
Birds	Habitat breadth	5.7E-01	0.0E+00	Original
Birds	Generation length	9.8E-01	0.0E+00	Original
Reptiles	Body mass	9.3E-01	0.0E+00	Original
Reptiles	Longevity	9.2E-01	2.2E-114	Original
Reptiles	Litter/clutch size	8.8E-01	0.0E+00	Original
Reptiles	Range size	7.6E-01	8.6E-148	Original
Reptiles	Habitat breadth	4.7E-01	1.9E-108	Original
Reptiles	Body length	9.9E-01	1.3E-316	Original
Reptiles	Maturity	9.5E-01	2.6E-59	Original
Mammals	Body mass	9.9E-01	0.0E+00	Original
Mammals	Longevity	9.2E-01	0.0E+00	Original
Mammals	Litter/clutch size	9.7E-01	0.0E+00	Original
Mammals	Diet breath	9.9E-01	0.0E+00	Original
Mammals	Range size	7.5E-01	1.0E-271	Original
Mammals	Habitat breadth	7.1E-01	2.9E-191	Original
Mammals	Generation length	9.8E-01	0.0E+00	Original
Mammals	Body length	9.9E-01	0.0E+00	Original
Amphibians	Body mass	9.7E-01	7.2E-49	Original
Amphibians	Longevity	7.6E-01	5.7E-23	Original
Amphibians	Litter/clutch size	9.4E-01	1.8E-204	Original
Amphibians	Diet breath	6.0E-01	1.5E-21	Original
Amphibians	Range size	8.3E-01	3.9E-175	Original
Amphibians	Habitat breadth	8.4E-01	1.3E-224	Original
Amphibians	Body length	9.6E-01	0.0E+00	Original

Table 6: χ-squared test result for the significance of phylogenetic signal in continuous traits. The results are shown here, for phylogenetic signal (λ) computed with **modified** phylogenies.

Class	Trait	lambda	p-value	Phylogeny
Birds	Body mass	9.8E-01	0.0E+00	Modified
Birds	Longevity	8.4E-01	2.6E-268	Modified
Birds	Litter/clutch size	9.4E-01	0.0E+00	Modified
Birds	Diet breath	4.9E-01	2.1E-236	Modified
Birds	Range size	7.0E-01	0.0E+00	Modified
Birds	Habitat breadth	6.0E-01	0.0E+00	Modified
Birds	Generation length	9.8E-01	0.0E+00	Modified
Reptiles	Body mass	8.8E-01	0.0E+00	Modified
Reptiles	Longevity	9.3E-01	5.2E-200	Modified
Reptiles	Litter/clutch size	8.6E-01	0.0E+00	Modified
Reptiles	Range size	6.9E-01	8.1E-257	Modified
Reptiles	Habitat breadth	4.5E-01	2.8E-174	Modified
Reptiles	Body length	9.6E-01	0.0E+00	Modified
Reptiles	Maturity	9.3E-01	1.1E-78	Modified
Mammals	Body mass	9.9E-01	0.0E+00	Modified
Mammals	Longevity	9.3E-01	0.0E+00	Modified
Mammals	Litter/clutch size	9.7E-01	0.0E+00	Modified
Mammals	Diet breath	9.9E-01	0.0E+00	Modified
Mammals	Range size	7.5E-01	1.3E-289	Modified
Mammals	Habitat breadth	7.1E-01	5.1E-196	Modified
Mammals	Generation length	9.8E-01	0.0E+00	Modified
Mammals	Body length	9.9E-01	0.0E+00	Modified
Amphibians	Body mass	9.7E-01	2.2E-107	Modified
Amphibians	Longevity	8.3E-01	3.6E-34	Modified
Amphibians	Litter/clutch size	9.4E-01	1.2E-303	Modified
Amphibians	Diet breath	7.8E-01	8.1E-62	Modified
Amphibians	Range size	8.2E-01	0.0E+00	Modified
Amphibians	Habitat breadth	8.1E-01	0.0E+00	Modified
Amphibians	Body length	9.5E-01	0.0E+00	Modified

3.2   Imputation congruence in categorical traits

Figures 8, 9, 10 and 11 show imputation congruence across the eight imputed datasets for each categorical trait. Here, I assessed the proportion of species for which imputed values were all similar across the eight imputed datasets. If at least one dataset imputed a different value, imputations were considered to not be congruent (contradicting). The results are presented for each class separately. A color code indicates whether the phylogenetic position of the species was initially known.

These figures show that:

- For all classes, imputation congruence was high for trophic level (Figure 8).

- For primary diet, congruence was very high for mammals and amphibians. More than half of avian species had similar values imputed. Nevertheless, in birds, there was a substantial amount of species for which at least one dataset imputed a different value (Figure 9).

- For both diel activity and degree of habitat specialisation, imputation congruence was high for mammals and birds, but less good in amphibians and reptiles (Figures 10 and 11).
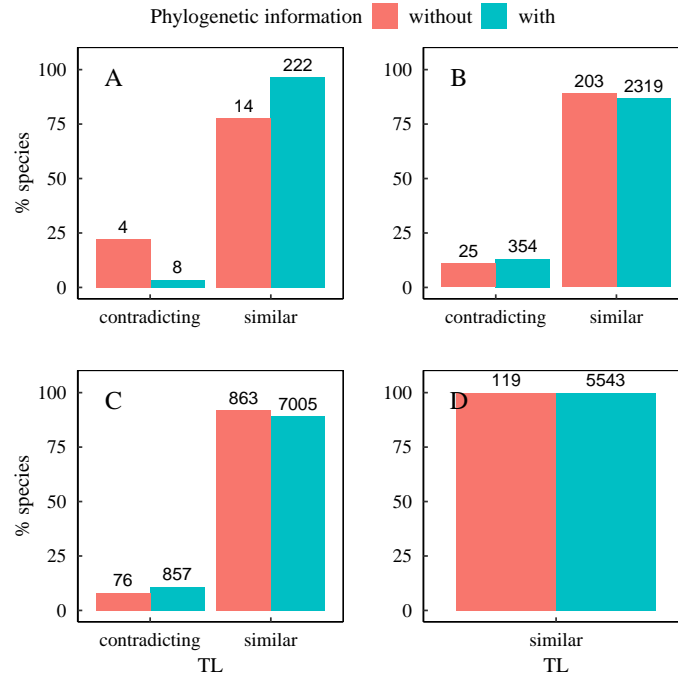
Figure 8: **Imputation congruence for trophic level across eight imputed datasets.** (**A**) Mammals, (**B**) birds, (**C**) reptiles and (**D**) amphibians.
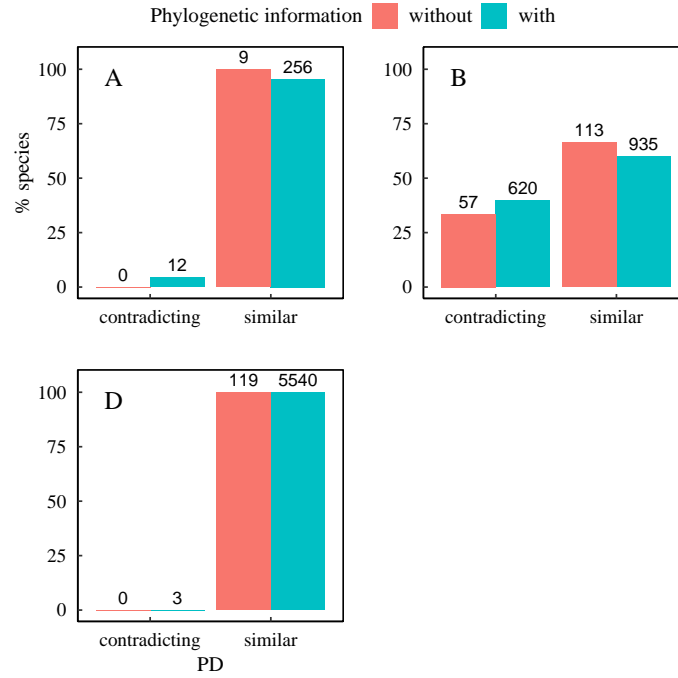
Figure 9: Imputation congruence for primary diet across eight imputed datasets. **(A)** Mammals, **(B)** birds and **(D)** amphibians.
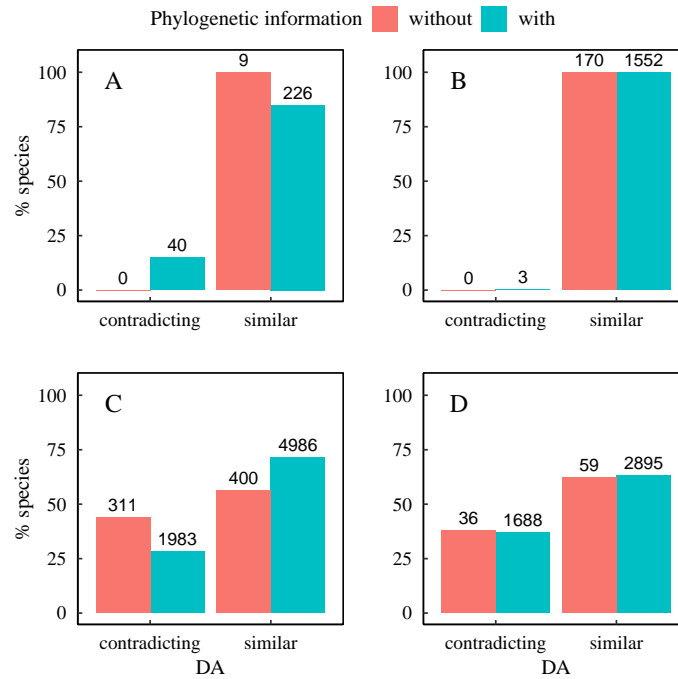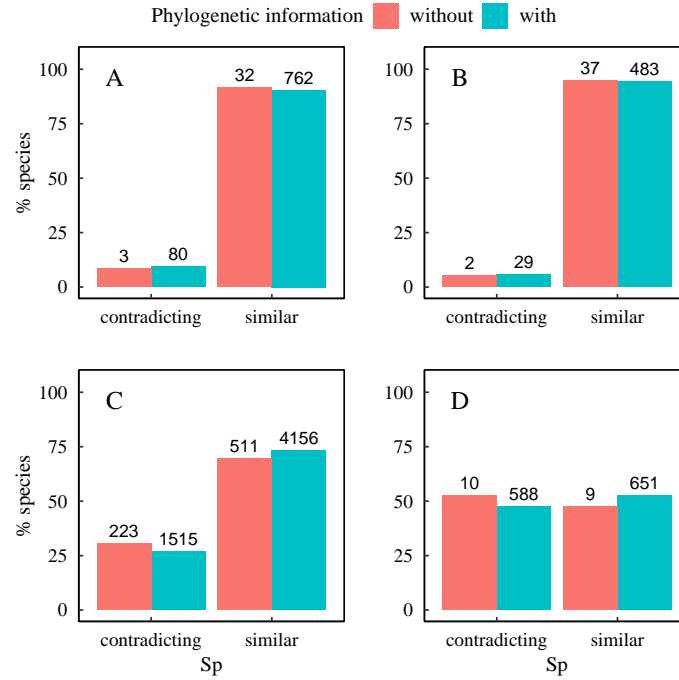

Figure 10: Imputation congruence for diel activity across eight imputed datasets. **(A)** Mammals, **(B)** birds, **(C)** reptiles and **(D)** amphibians.

Figure 11: Imputation congruence for habitat specialisation across eight imputed datasets. (**A**) Mammals, (**B**) birds, (**C**) reptiles and (**D**) amphibians.

3.3 Imputation results for continuous traits
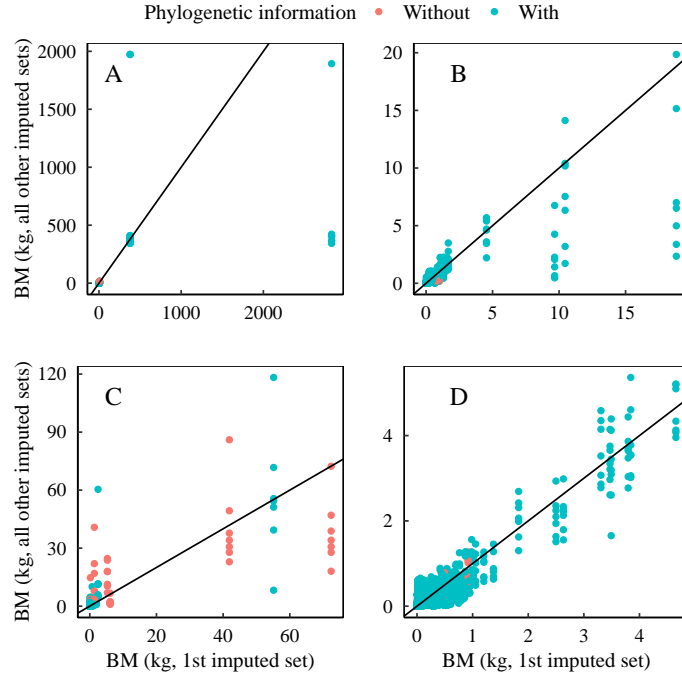
Figure 12: Imputation results for body mass across eight imputed datasets. **(A)** Mammals, **(B)** birds, **(C)** reptiles and **(D)** amphibians.
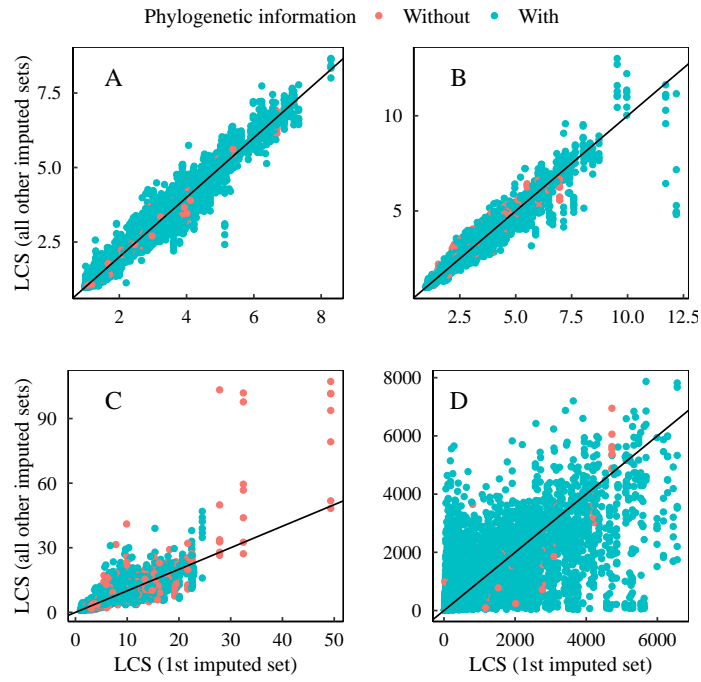
Figure 13: Imputation results for litter/clutch size across eight imputed datasets. **(A)** Mammals, **(B)** birds, **(C)** reptiles and **(D)** amphibians.
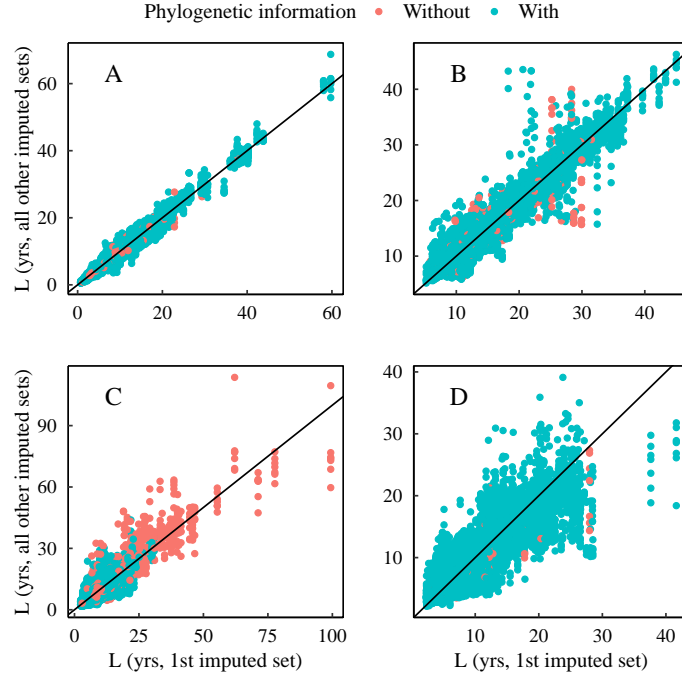
Figure 14: Imputation results for longevity across eight imputed datasets.(**A**) Mammals, (**B**) birds, (**C**) reptiles and (**D**) amphibians.
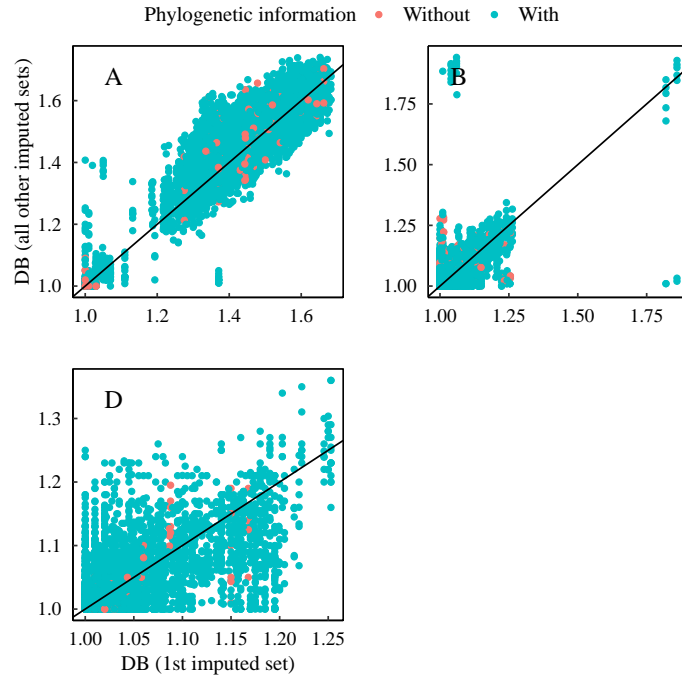

Figure 15: Imputation results for diet breadth across eight imputed datasets.(**A**) Mammals, (**B**) birds and (**D**) amphibians.
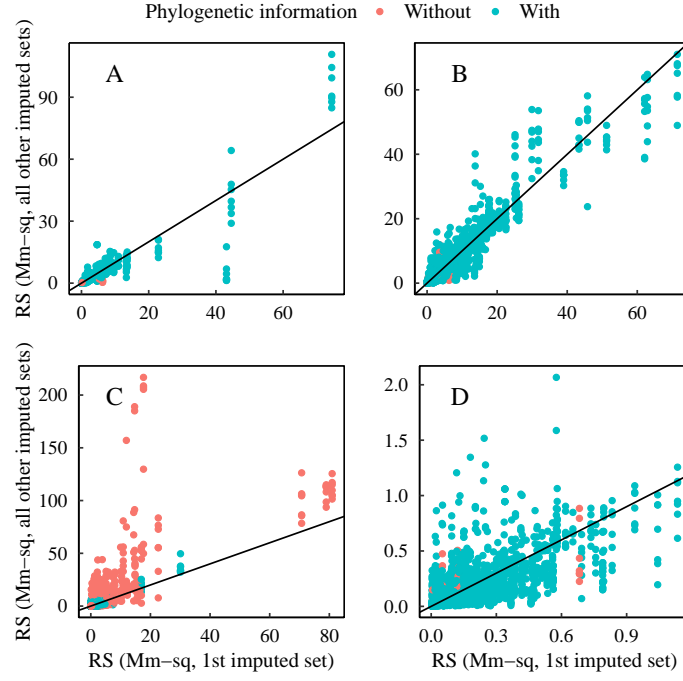
Figure 16: Imputation results for range size across eight imputed datasets. **(A)** Mammals, **(B)** birds, **(C)** reptiles and **(D)** amphibians.
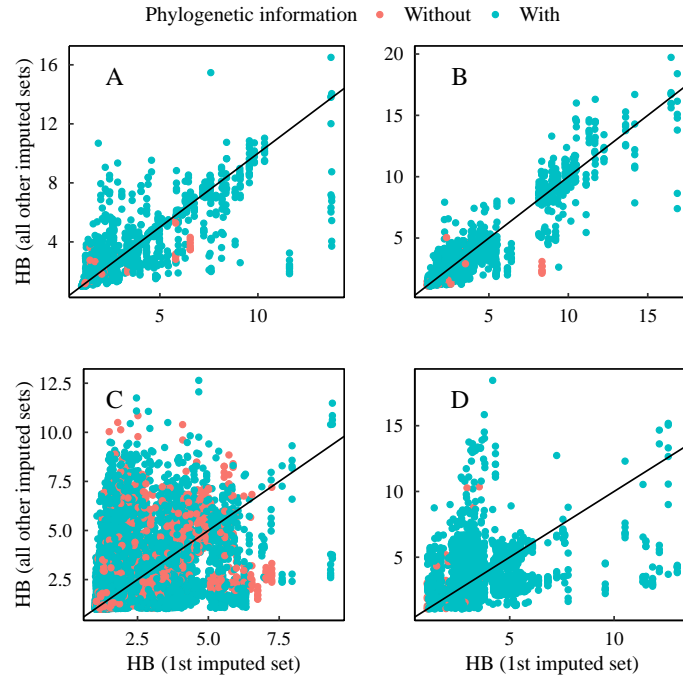

Figure 17: Imputation results for habitat breadth across eight imputed datasets. **(A)** Mammals, **(B)** birds, **(C)** reptiles and **(D)** amphibians.

3.4 Comparison with data from Cooke et al. (2019)

3.4.1 Comparison of imputed values in Cooke et al. (2019) against imputed values in this work

Cooke et al. (2019) released a comprehensive database of six avian and mammalian traits. The methods for data collection were very similar to the methods employed in this work. One notable difference in Cooke et al. (2019) was the preselection of traits was had more than 50% coverage. Our imputation methods also differed, as Cooke et al. employed multivariate chained equations (mice). Using species for which estimates were imputed in both Cooke et al. (2019) and in the present work, I compared the results of both methods for common imputed traits. Most species had different imputed results for primary diet (Figure 18). The results were overall qualitatively consistent for the other traits (diel activity, litter/clutch size and habitat breadth).
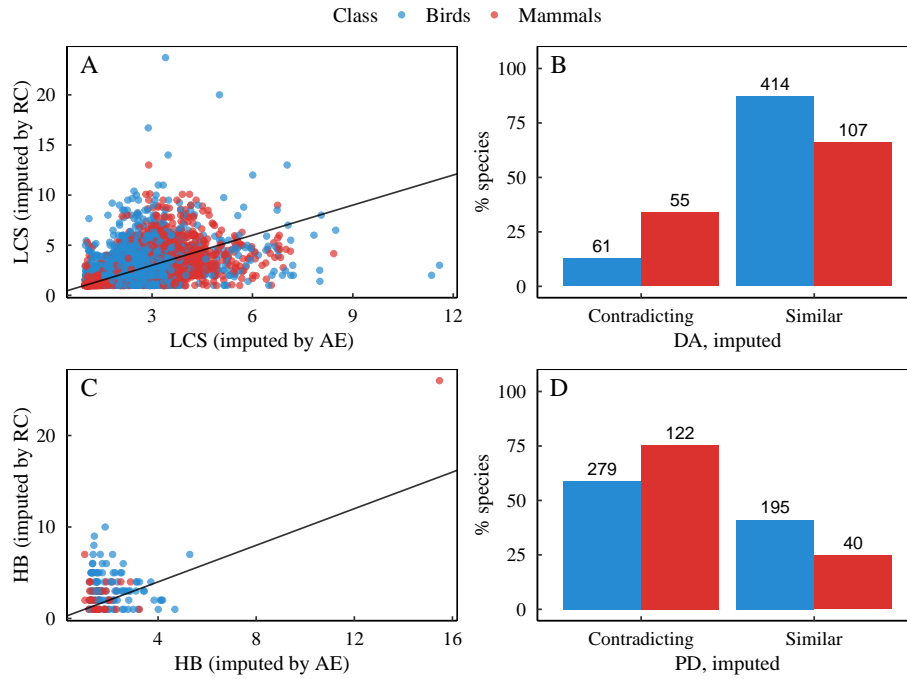

Figure 18: Values imputed by Cooke et al. against values imputed in this work.(A) Litter/clutch size, (B) diel activity, (C) habitat breadth and (D) primary diet.

3.4.2 Comparison of collected VS imputed values

Here, I compared values collected by Cooke et al. (2019) for a set of species for which I imputed the values, and vice-versa, when the collection design allowed for such comparisons. missForest seemed to have performed better than mice on continuous trait imputations (Figure 19 A, B, C, D): the correlation coefficients between imputed values and collected values was 0.47 for habitat breadth, imputed with missForest (B), and 0.33 imputed with mice (A). On litter/clutch size, the correlation coefficient was 0.89 for values imputed by missForest (C) and 0.41 for values imputed by mice (D). For categorical traits, both missForest and mice seemed to have performed well in predicting diel activity, but not well in predicting primary diet. Nevertheless, the sample sizes were low in each case, and the comparisons were conducted on different sets of species, so these results should not be extrapolated or interpreted in terms of imputation robustness.
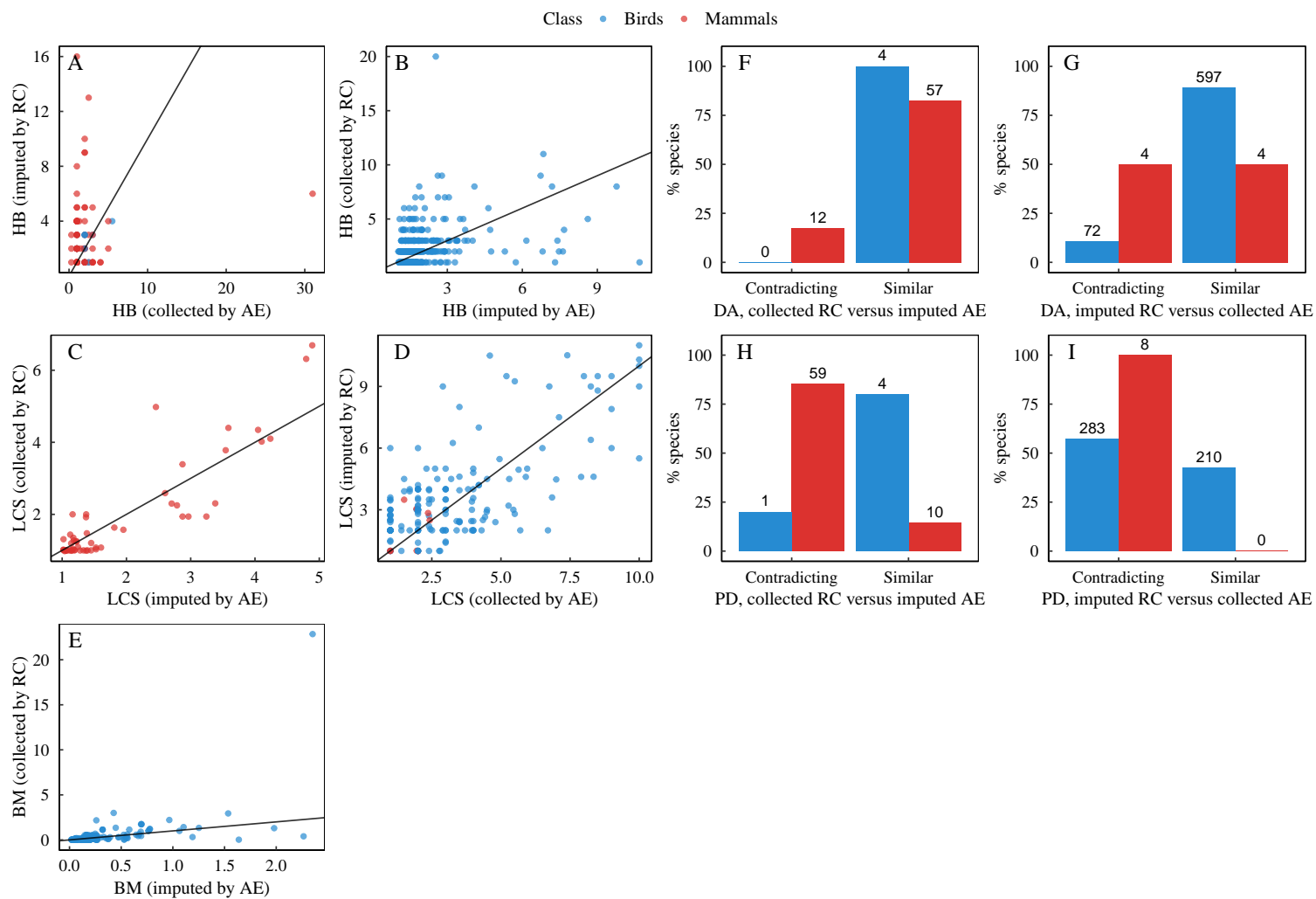
Figure 19: .(A) Mammals, (B) birds, (C) reptiles and (D) amphibians.

4 Patterns in missing values with regards to phylogenetic trees: plots with tip labels for within-family median completeness
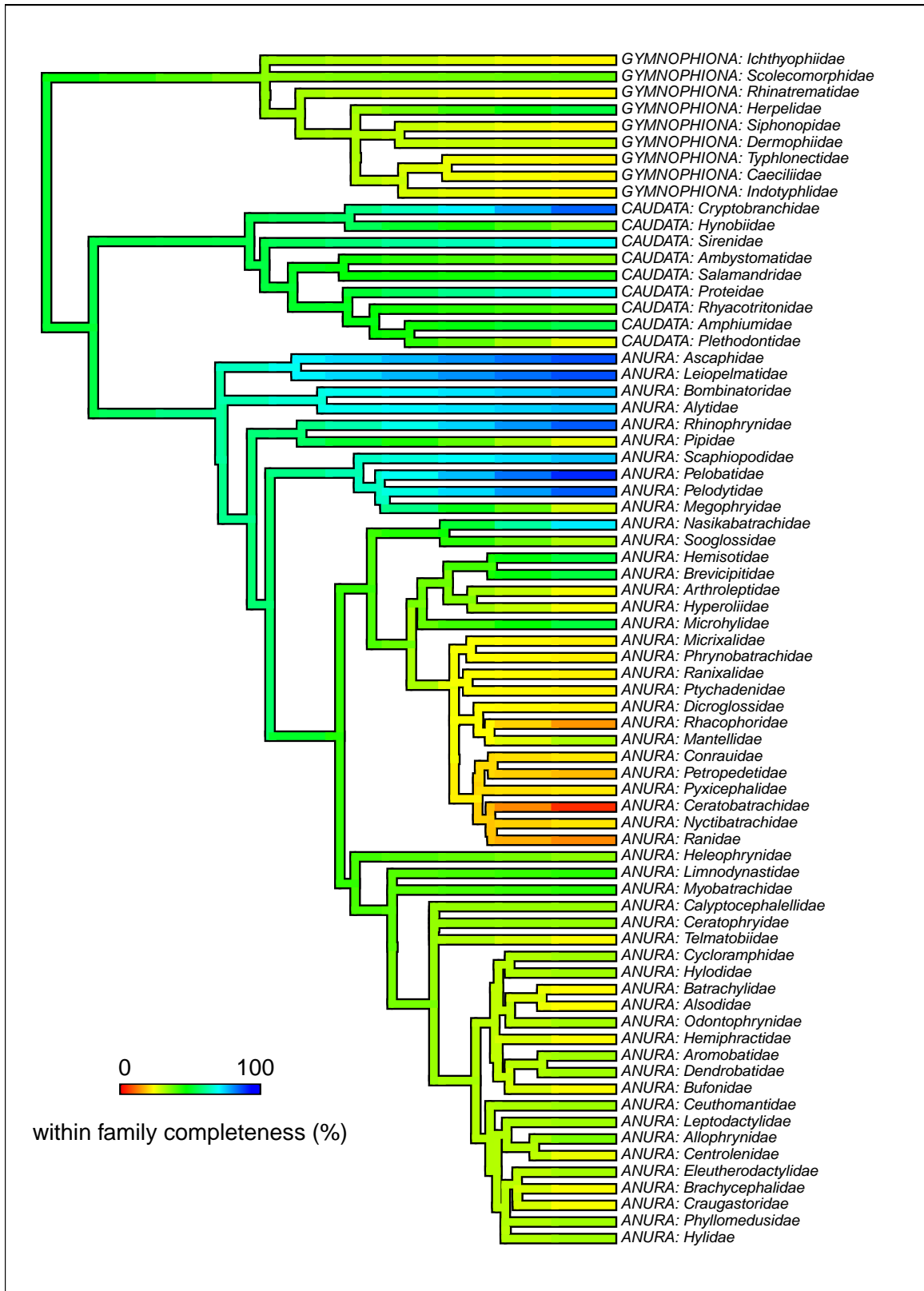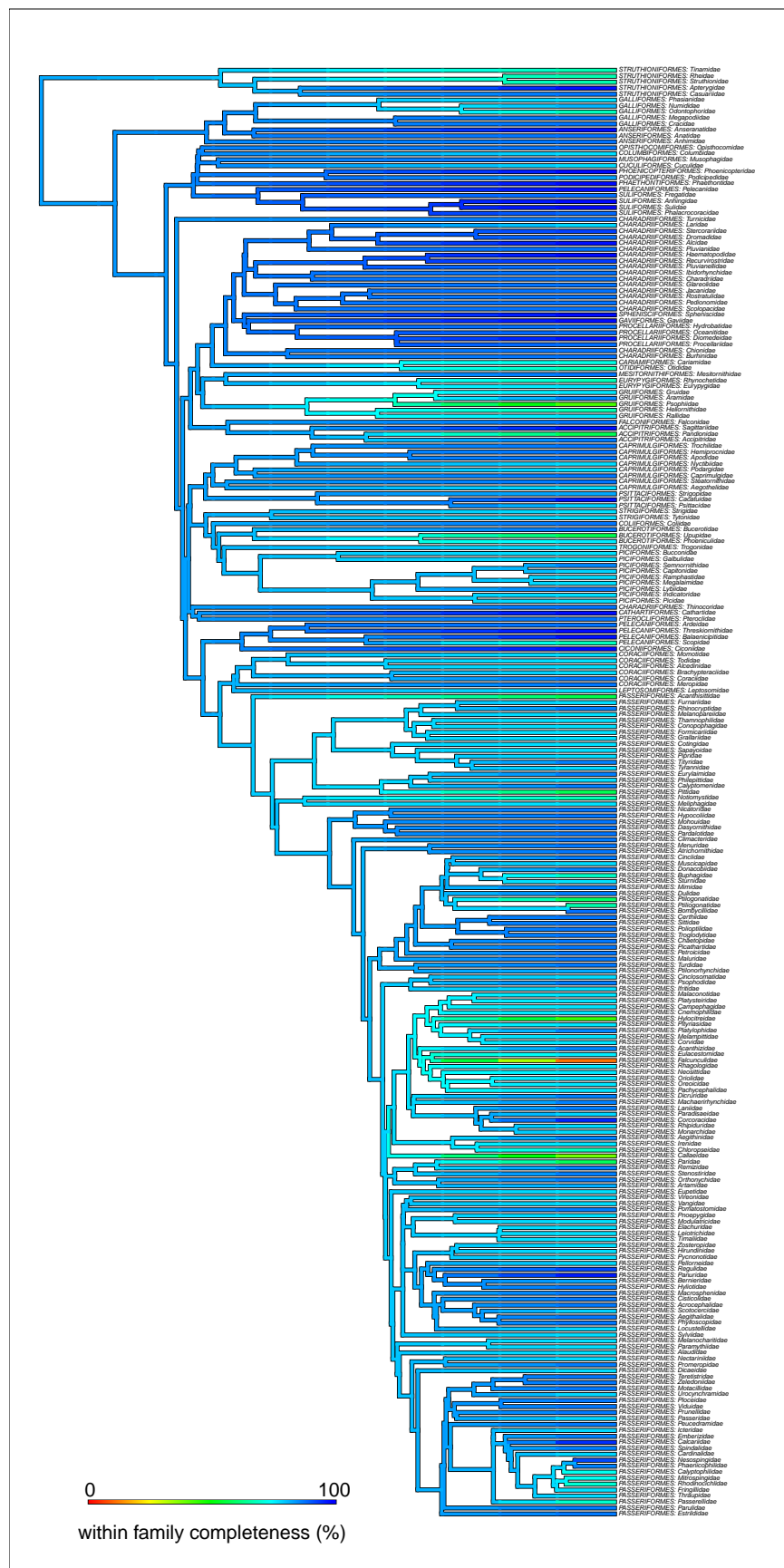
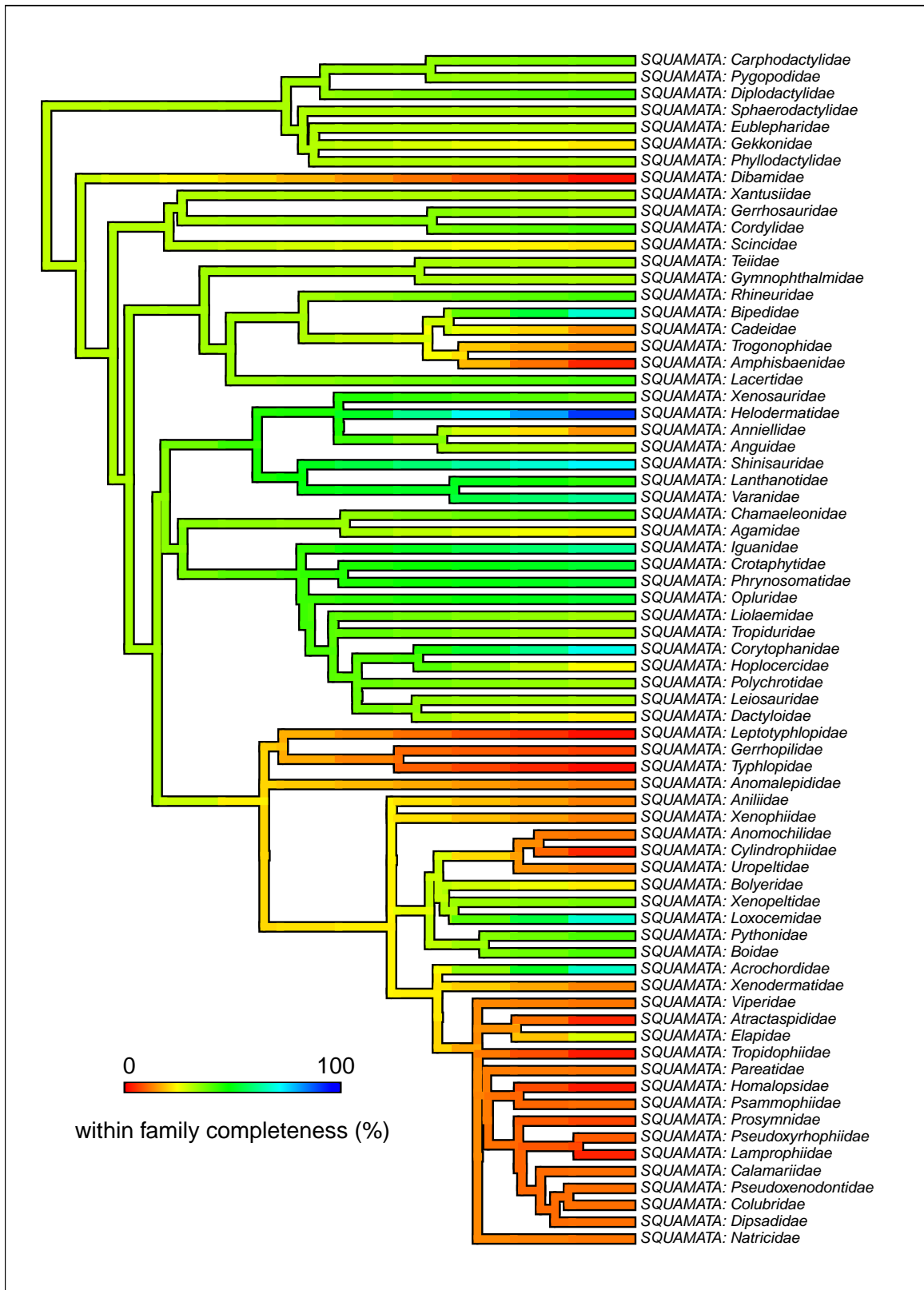Figure 20: Within-family median completeness for amphibians.

Figure 21: Within-family median completeness for mammals.

Figure 22: Within-family median completeness for birds.

Figure 23: Within-family median completeness for reptiles.
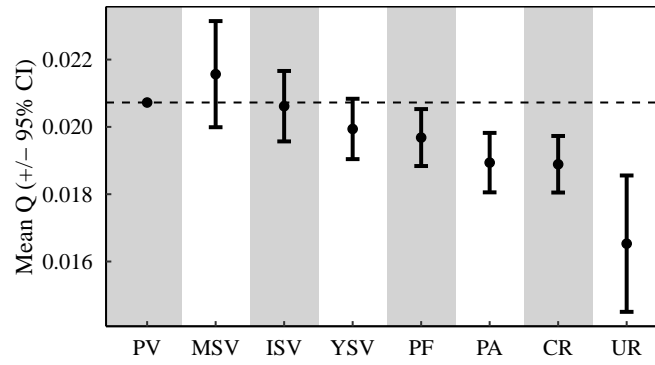
5 Mean effect of land-use on Rao's quadratic entropy

Figure 24: Mean effect of land-use on Rao's quadratic entropy in vertebrate communities. PV: primary vegetation; **MSV:** mature secondary vegetation; **ISV:** intermediate secondary vegetation; **YSV:** young secondary vegetation; **PF:** plantation forest; **CR:** cropland; **UR:** urban.