

# Supporting information for

Adrienne Etard

March 4, 2019

# Contents

<b>1</b>	<b>Trait data compilation</b>	<b>5</b>
1.1	Definition of diet categories . . . . .	5
1.2	Habitat affinities and broad specialisation . . . . .	5
1.3	Tackling taxonomic synonymy . . . . .	6
1.3.1	Dropping replicated phylogenetic tips . . . . .	6
1.4	Imputation robustness . . . . .	9
1.5	Trait distributions . . . . .	9

List of Tables

1	Weights used in the calculation of habitat breadth . . . . .	5
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 . . . . .	7

**List of Figures**

1	Distribution of habitat breadths across species for terrestrial vertebrates . . . . .	6
2	Procedure followed to drop replicated tips from phylogenies . . . . .	8
3	Phylogenetic trees before (A) and after (B) taxonomic correction, and after removing replicated tips (C) for species in the reptilian genus <i>Phrynocephalus</i> . . . . .	8

# 1 Trait data compilation

## 1.1 Definition of diet categories

For mammals and birds, diet was compiled from EltonTraits (ref). Primary diet was available in the avian dataset and declined in 5 categories: (1) plant or seed consumers; (2) fruit or nectar consumers; (3) For amphibians, diet information was extracted from AmphiBIO.

## 1.2 Habitat affinities and broad specialisation

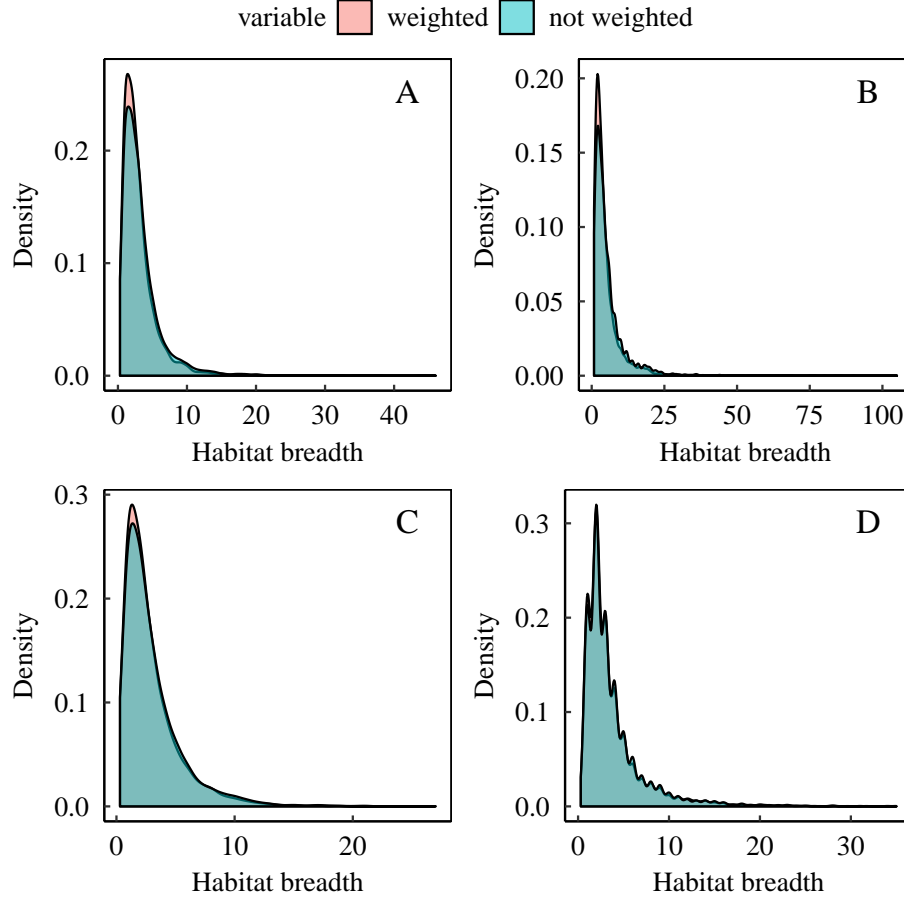
**Habitat preferences.** IUCN habitat data records habitat types in which species occur. Habitats are classified into 96 categories, which I pooled into 13 broader habitat variables: Forest, Savanna, Grassland, Shrubland, Wetland, Rocky areas, Caves and subterranean, Desert, Marine, Marine intertidal 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 IUCN database. Given that information regarding habitat suitability and habitat importance was also available in the IUCN data files, I used a weighted sum to calculate habitat breadth. Suitability was declined in three categories in the IUCN 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).

### 1.3 Tackling taxonomic synonymy

#### 1.3.1 Dropping replicated phylogenetic tips

2.6% of mammalian, 1.5% of avian, 1% of amphibian and 1.5% of reptilian species had 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, Figure 2.1 in 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.

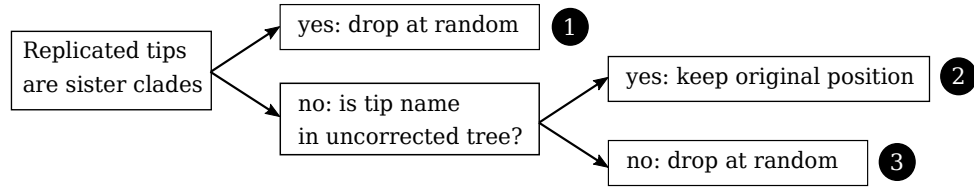
Figure 2 describes the procedure I used. If replicated tips were sister clades, the tip to conserve was chosen randomly among the replicates. Else, I chose to conserve the tree tip whose position was closest to the position of the same tip in the uncorrected tree, when present. In all other few cases, tips to drop were chosen randomly. See Figures 3, 4 and 5 for an example of each case.

**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.

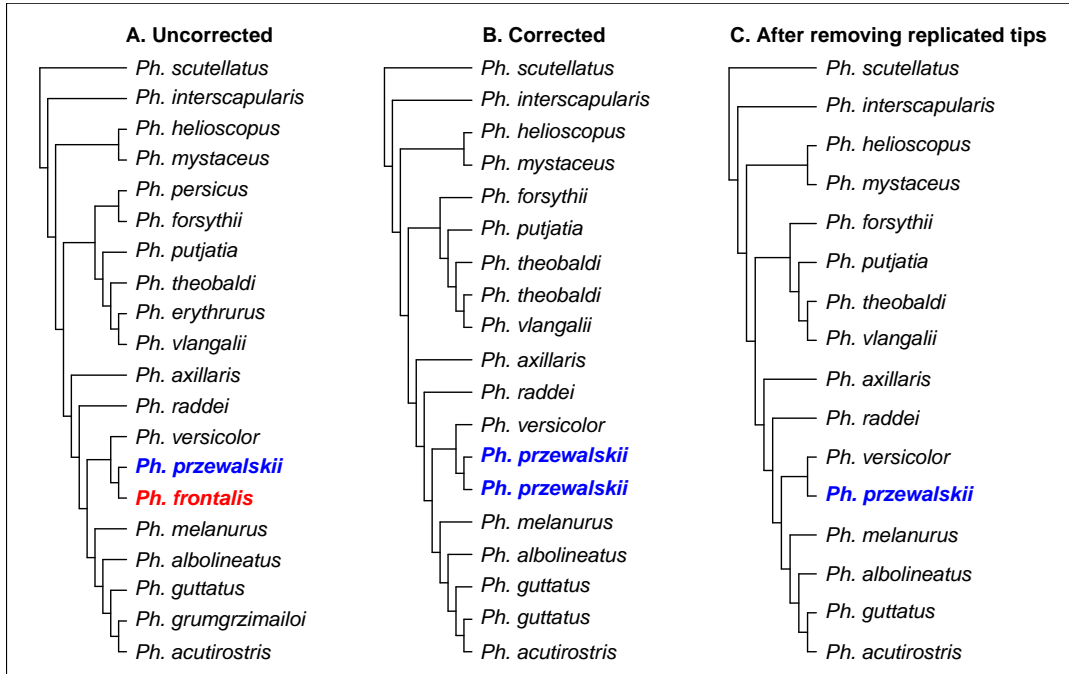
Class	Replicates			Sister clades	In uncorrected tree	‘Problematic’
	2	3	>3			
<b>Mammals</b>	129	7	2	28	131	<i>Heterogeomys cherriei</i> <i>Heterogeomys dariensis</i> <i>Hylopetes sagitta</i> <i>Marmosa paraguayana</i> <i>Neoromicia malagasyensis</i> <i>Plecturocebus discolor</i> <i>Proechimys trinitatis</i>
<b>Birds</b>	158	7	0	22	156	<i>Aerodramus vanikorensis</i> <i>Antrostomus arizonae</i> <i>Calendulauda erythrochlamys</i> <i>Eolophus roseicapilla</i> <i>Myiothlypis rivularis</i> <i>Poecile hudsonicus</i> <i>Spermestes bicolor</i>
<b>Reptiles</b>	68	2	0	17	68	<i>Liolaemus curicensis</i> <i>Salvator merianae</i>
<b>Amphibians</b>	41	4	2	8	43	<i>Lithobates berlandieri</i> <i>Uperodon taprobanicus</i>

## Case studies

**Case study 1 (Figure 3): replicated tips are sister clades.** In this example, 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 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, occurrences of that third case were rare.



**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: replicated tips are not sister clades, but the accepted name (tip label) figures in the original, uncorrected tree.**

**Case study 3: 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.



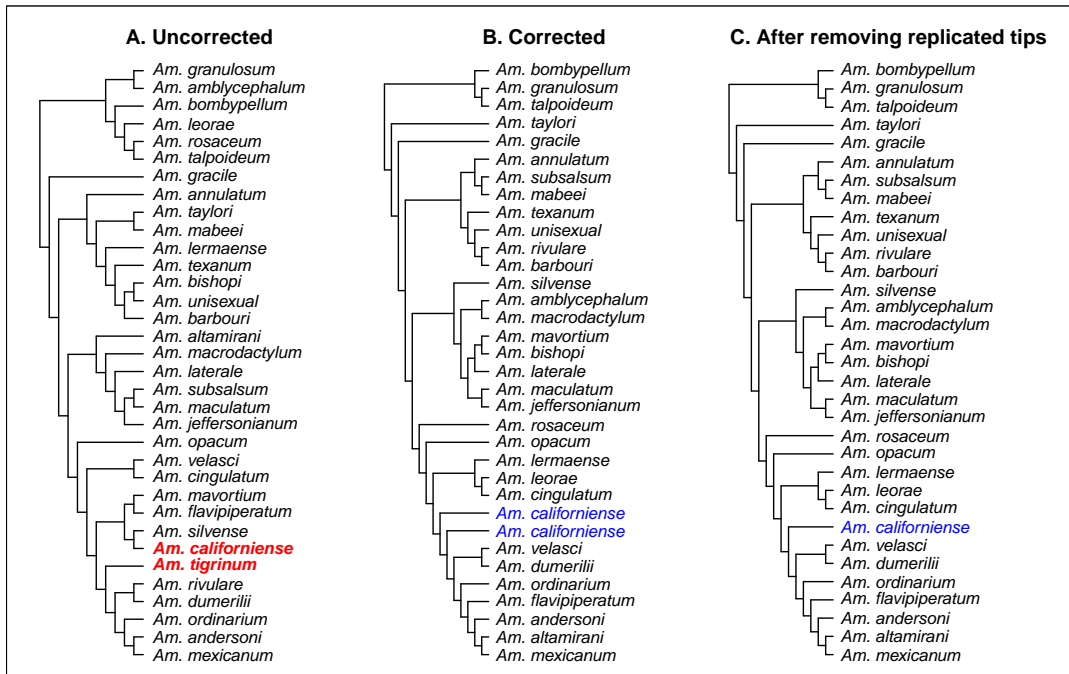


Figure 4:

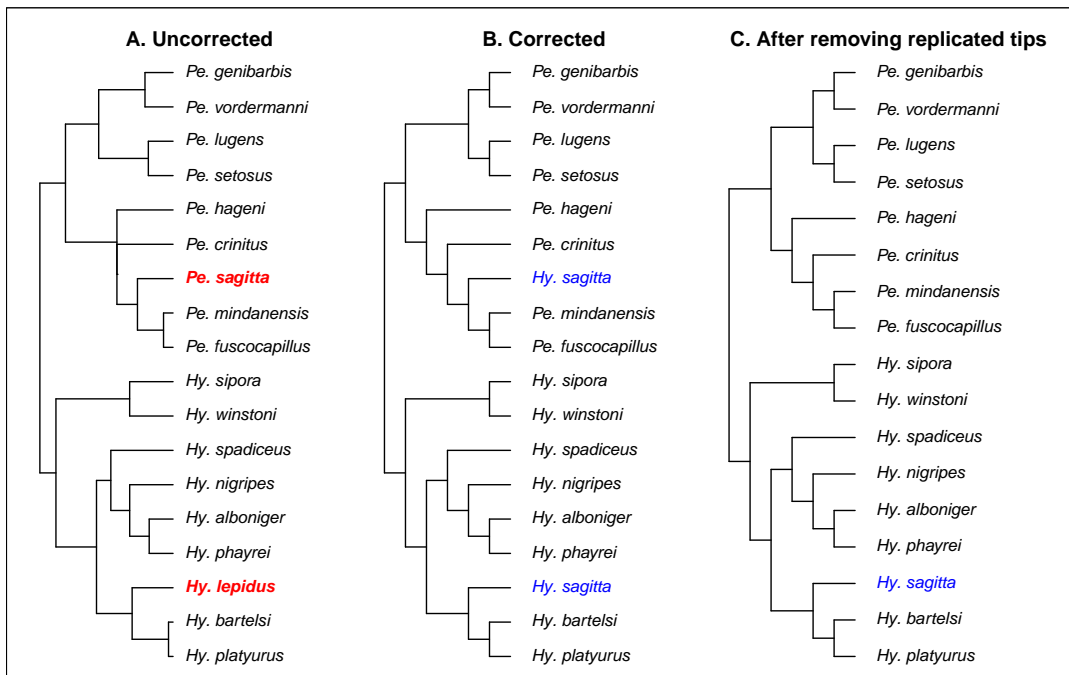


Figure 5:

## 1.4 Imputation robustness

## 1.5 Trait distributions