















# Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls

José Alexandre F. Diniz-Filho<sup>1\*</sup> ,  
Lucas Jardim<sup>2</sup> , Jhonny J. M. Guedes<sup>3</sup> ; Leila Meyer<sup>4</sup> ,  
Juliana Stropp<sup>5</sup> , Livia Estéfane Fernandes Frateles<sup>3</sup> ,  
Rafael B. Pinto<sup>6</sup> , Lúcia G. Lohmann<sup>7</sup> , Geiziane Tessarolo<sup>3,8</sup> ,  
Claudio J. B. de Carvalho<sup>9</sup> , Richard J. Ladle<sup>10,11,12</sup>  and  
Joaquín Hortal<sup>5</sup> 

<sup>1</sup>Instituto de Ciências Biológicas, Departamento de Ecologia, Universidade Federal de Goiás, Goiânia, Brazil;

<sup>2</sup>Laboratório de Macroecologia, Universidade Federal de Goiás, GO, Brazil;

<sup>3</sup>Programa de Pós-Graduação em Ecologia e Evolução, Departamento de Ecologia, Campus Samambaia, Universidade Federal de Goiás, Goiânia, Brazil;

<sup>4</sup>Programa de Pós-Graduação em Biologia Vegetal, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil;

<sup>5</sup>Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales, Madrid, Spain;

<sup>6</sup>Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de Ciências Biológicas, Universidade Federal de Goiás, Goiânia, Brazil;

<sup>7</sup>Instituto de Biociências, Departamento de Botânica, Universidade de São Paulo, São Paulo, Brazil;

<sup>8</sup>Departamento de Biogeografia e Ecologia Aquática, Universidade Estadual de Goiás, Anápolis, GO, Brazil;

<sup>9</sup>Departamento de Zoologia, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, Brazil;

<sup>10</sup>Institute of Biological and Health Sciences, Federal University of Alagoas, Maceió, Brazil;

<sup>11</sup>CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661, Vairão, Portugal;

<sup>12</sup>BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal.

\*Correspondence: José Alexandre F. Diniz-Filho, diniz@ufg.br; jafdinizfilho@gmail.com

## Abstract

Species are the currency of most biodiversity studies. However, many shortfalls and biases remain in our biodiversity estimates, preventing a comprehensive understanding of the eco-evolutionary processes that have shaped the biodiversity currently available on Earth. Biased biodiversity estimates also jeopardize the effective implementation of data-driven conservation strategies, ultimately leading to biodiversity loss. Here, we delve into the concept of the Latitudinal Taxonomy Gradient (LTG) and show how this new idea provides an interesting conceptual link between the Linnean (i.e., our ignorance of how many species there are on Earth), Darwinian (i.e., our ignorance of species evolutionary relationships), and Wallacean (i.e., our ignorance on species distribution) shortfalls. More specifically, we contribute to an improved understanding of LTGs and establish the basis for the development of new methods that allow us to: (i) better account for the integration between different shortfalls and, (ii) estimate how these interactions may affect our understanding about the evolutionary components of richness gradients at macroecological scales.

## Highlights

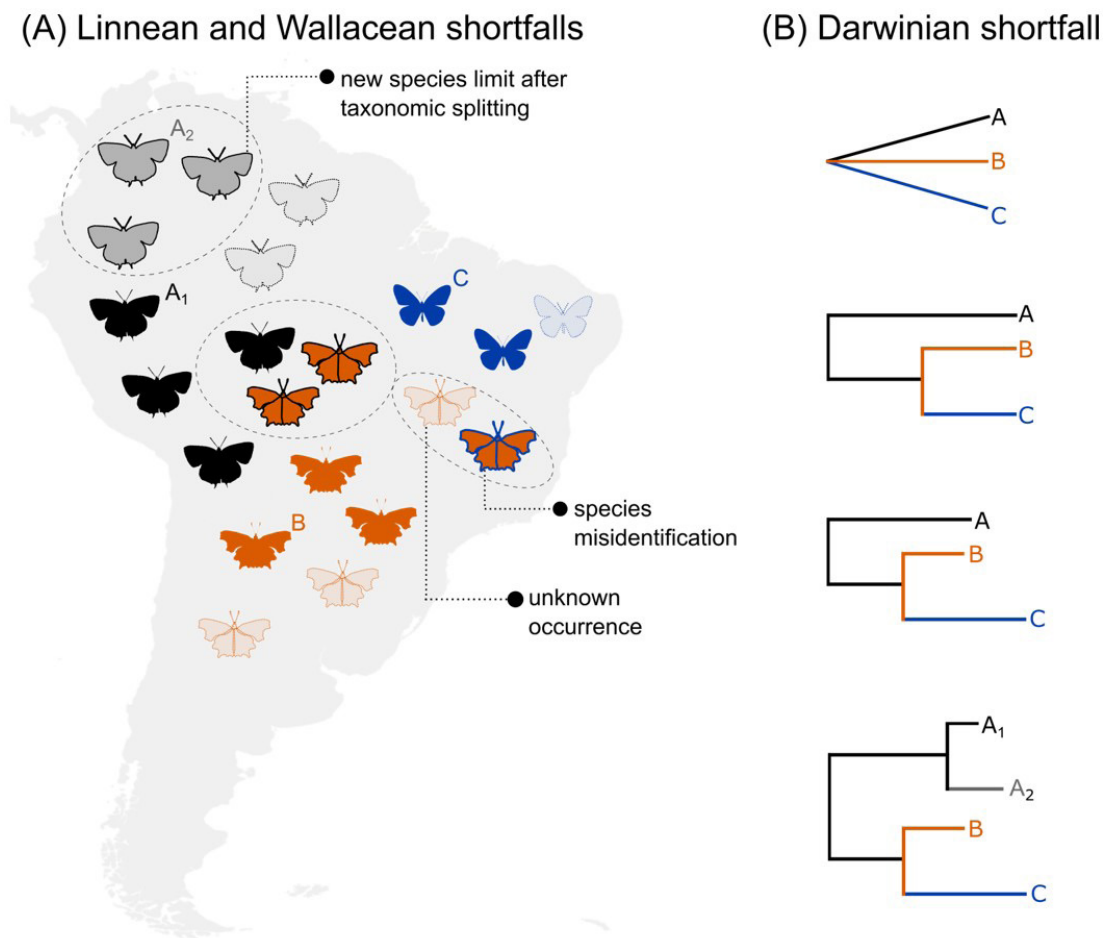
- Biodiversity shortfalls lead to biased biodiversity estimates, preventing a comprehensive understanding of biodiversity patterns, and jeopardizing the effective implementation of data-driven conservation strategies.
- When new species are described, species diversity patterns based on phylogenies are less affected than those based on richness.
- In this context, solving the Darwinian shortfall will lead to more robust evaluations of macroecological and biogeographic patterns.
- However, Latitudinal Taxonomic Gradients (LTGs) provide an interesting framework to explore how uneven knowledge about species taxonomy and geographical distribution link the Linnean, Wallacean, and Darwinian shortfalls and affect the evaluation of diversification patterns.

**Keywords:** biodiversity bias, Darwinian shortfall, diversity gradients, knowledge shortfalls, latitudinal taxonomy gradient, Linnean shortfall, macroecology, uncertainty, species diversity, Wallacean shortfall.

### Introduction

Our current biodiversity knowledge includes multiple shortfalls (see Hortal et al. 2015 for a review; Fig. 1; Table 1) that range from a lack of basic information about species taxonomy (Linnean shortfall) and their geographic ranges (Wallacean shortfalls) to more complex issues related to the lack of knowledge on population dynamics (Prestonian shortfall), functional traits (Raunkiaerian shortfall), species interactions (Eltonian shortfall), and various components of ecological niches (Hutchinsonian

shortfall). Because evolutionary dynamics impact biodiversity patterns at different scales, it is important to evaluate the origin and effect of biodiversity shortfalls through comparative methodological approaches that seek to investigate the evolutionary imprint on the patterns observed (Harvey and Pagel 1991; Mouquet et al. 2012; Nur et al. 2022). However, our knowledge about phylogenetic relationships, divergence times, and trait evolution is far from complete, hampering a comprehensive understanding of the evolutionary history of lineages (i.e., the



**Figure 1.** Illustration of the three biodiversity shortfalls discussed for a hypothetical group of butterflies. Within the left map, the known occurrence of species A, B and C is represented by solid drawings, whereas faded drawings indicate unknown occurrences (i.e., corresponding to the Wallacean shortfall). Notice that although there is a lack of knowledge on records for all species, only for species B the Wallacean shortfall implies a serious underestimate of geographic range position and extent, which would be expanded southward after knowledge increases. The dashed ellipses indicate different components of the Linnean shortfall, including the misidentification of species (the colors surrounding occurrences indicate incorrectly identified species) and the description of a new species (A2, in grey) that becomes recognized after a better evaluation of former species A based on more integrative taxonomic practices, which results in splitting its populations into A1 (corresponding to A) and A2 (new species). On the right, we show the different components of the Darwinian shortfall, starting with a first definition of the topology and branch lengths, followed by a more detailed evaluation of evolutionary models and, at the bottom, the addition of the newly recognized species A2 as a small branch within the original species A. Notice that adding this new species is related to the Darwinian shortfall, as it changes the topology and statistical distribution of branch lengths within clades (thus affecting the estimates of diversification rates, as shown in Figs. 2 and 3) and to the Wallacean shortfall, as it causes shifts in different properties of geographic ranges of species A.

**Table 1.** A brief description of the Linnean, Wallacean and Darwinian shortfalls (see Hortal et al. 2015), the three biodiversity shortfalls more directly involved in the effects of systematics on our ability to describe biogeographical patterns.

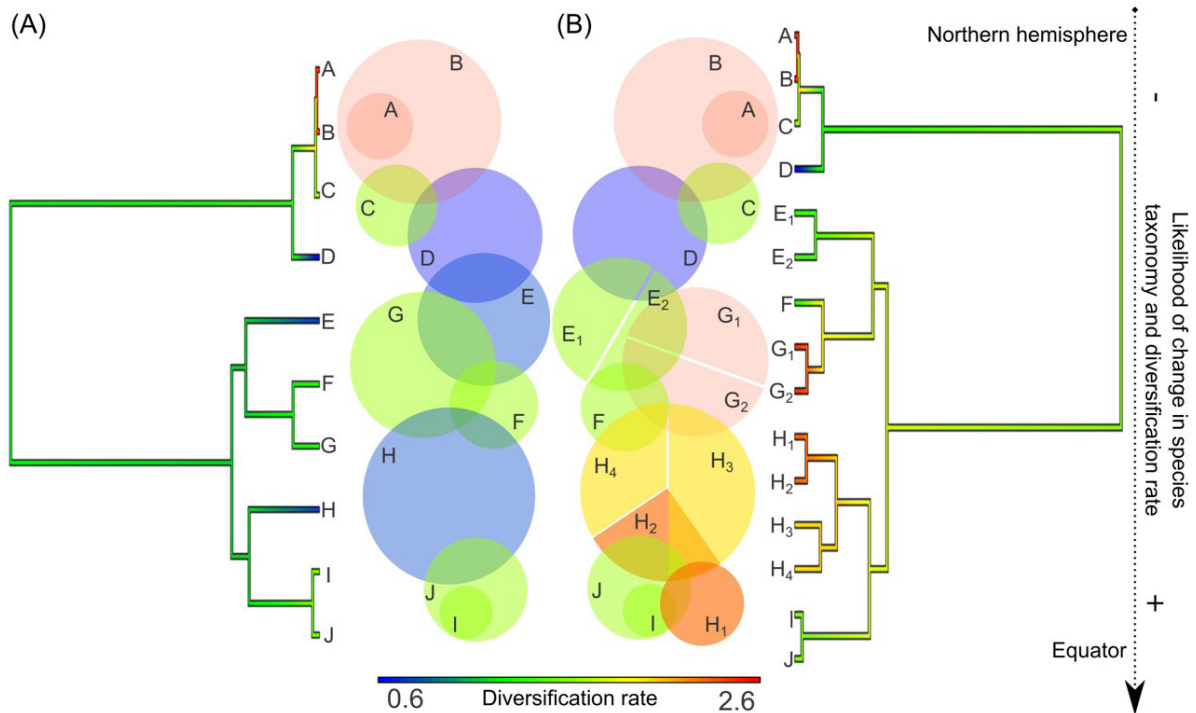
Shortfall	Description
Linnean	Most species, both extant and extinct, have not yet been identified and described, as a result of a number of reasons. Many unknown species may still be undiscovered in poorly sampled regions around the world. Simultaneously, even well-known regions may hide undescribed species because of poor taxonomic practices, or changes in the criteria used to delimitate biological units at a given evolutionary level (i.e., species). The Linnean shortfall is usually considered the most fundamental one, as it affects all other shortfalls.
Wallacean	There is typically a lack of knowledge of the whole extent of the geographic distribution ranges of most species, even for taxa with a reasonably good delimitation of most valid species. In many situations, the lack of knowledge on species distributions affects the estimates of species' extent of occurrence and geographic position. Further, lack of occurrence data within the polygons delimitating distribution ranges hampers estimating geographic occupancy that, in turn, relates to other shortfalls, mainly the Prestonian (lack of data on population abundance and dynamics) and Eltonian (lack of knowledge on biotic interactions).
Darwinian	There is a general lack of knowledge on the tempo and mode of evolution, in particular of the tree of life, i.e., the phylogenetic relationships among species. First, in many cases, basic knowledge of the topological relationships among species is lacking. Second, there are limited data on the distribution of branch lengths, which can be calibrated based on molecular distances and fossil data. Finally, we have a limited understanding on how functional traits and niche properties of the species evolve, and which one of the alternative evolutionary models they follow. This last component hampers establishing how the Darwinian shortfall is linked with other shortfalls such as the Raunkiaerian (lack of knowledge on species' traits and functions) and Hutchinsonian (lack of knowledge on species' responses to abiotic factors).

Darwinian shortfall; Diniz-Filho et al. 2013). These knowledge gaps inevitably affect our understanding of ecological patterns and their underlying processes, jeopardizing the establishment of efficient data-driven conservation strategies (Whittaker et al. 2005; Bini et al. 2006; Tucker et al. 2019).

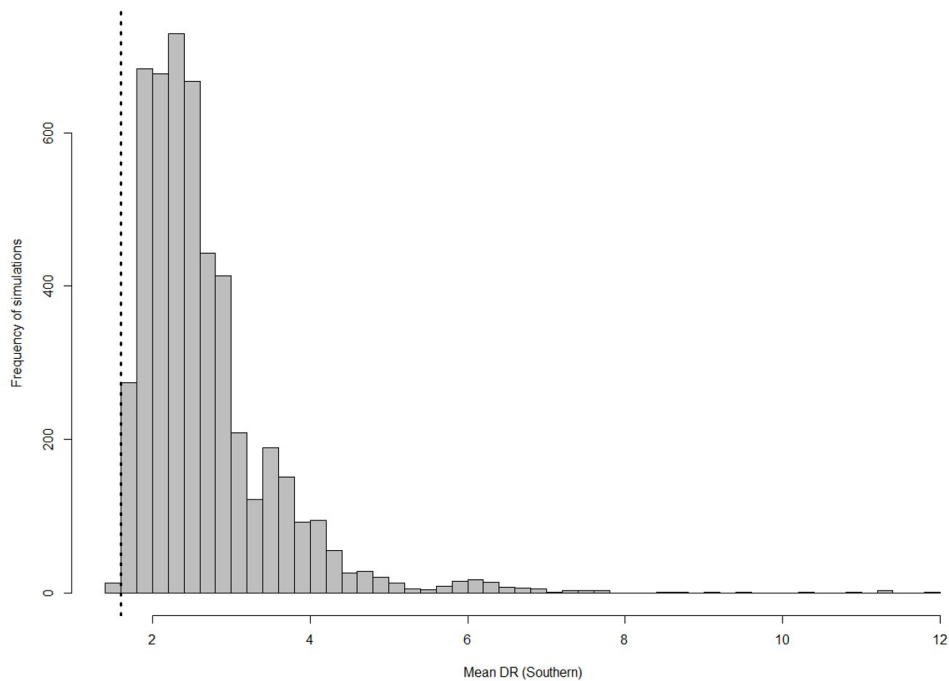
Even though the Linnean shortfall represents the most fundamental shortfall, impacting all other shortfalls in various ways (Hortal et al. 2015), the network of reciprocal dependencies among biodiversity knowledge shortfalls is still operationally and conceptually unclear. For example, describing new species within a clade (i.e., by solving the Linnean shortfall) increases the Phylogenetic Diversity (PD; Faith 1992), which is frequently estimated by summing all branch length paths connecting sets of species in a phylogeny. Nonetheless, the species richness-PD relationship tends to follow a power law function, eventually saturating PD values as more species are added to the pool (Fritz and Rahbek 2012; Tucker and Cadotte 2013; Voskamp et al. 2017). Once the backbone of the phylogeny is established, adding new species tend to bind terminal branches exclusively, not significantly impacting the overall phylogenetic knowledge (Blackburn et al. 2019). This process occurs, for instance, under “taxonomic

inflation” (Isaac et al. 2004; see also Lum et al. 2022), whereby new species primarily result from changes in taxonomic practice (e.g., changes in species concepts and/or species delimitation) rather than new discoveries.

Following the above reasoning, Diniz-Filho et al. (2013) proposed that solving our ignorance about the phylogenetic relationships among known species would provide the basis for consistent estimates of overall diversity patterns. With a robust knowledge of phylogenetic relationships, the resulting data would not be substantially impacted by the description of new species in the future, allowing for more robust and stable spatial conservation planning strategies. In other words, conservation applications are highly dependent on solving the first component of the Darwinian shortfall (i.e., lack of comprehensive available phylogenies). In fact, for practical applications, solving the Darwinian shortfall might be even more urgent than solving the Linnean shortfall, providing a “by-pass” for rapidly increasing our knowledge on overall biodiversity patterns. However, the presumed independence between the Darwinian and Linnean shortfalls at broad scales may not apply well to other macroecological and biogeographical patterns.



**Figure 2.** A schematic hypothetical example of the effects of the latitudinal taxonomic gradient (LTG) proposed by Freeman and Pennel (2021) in respect to diversification rates. On the left there is a phylogeny for 10 currently known species (left phylogeny). The northern species (A to D) are found in well-known regions and represent well-delimited species. Conversely, the species known from the southern region (E to J) are not well delimited. Based on surrogates of robustness of species delimitation (or after some empirical integrative taxonomic studies) researchers decide to split species E, G and H into several taxa, consequently subdividing their original geographic ranges. The color gradient in species’ geographic range and in the branches in each phylogeny range refer to the Diversification Rate (DR) and vary from 0.6 (blue) to 2.6 (red), mapped using contMap function of the R package *Phytools*.



**Figure 3.** The statistical distribution of mean DR in the southern region obtained by 5000 replications of the splitting of species illustrated in Fig. 2. In each replication, 5 cycles of splitting were simulated, randomly selecting any of the southern species and allocating the new derived species in a random position along the branch of the ancestral species. The dashed vertical line shows the mean DR in the northern species.

## Linking the Linnean and Darwinian shortfalls

Freeman and Pennell (2021) recently called attention to an important issue that leads to an unsuspected link between the Linnean and Darwinian shortfalls: taxonomic delimitation of species can be geographically structured, resulting in a “Latitudinal Taxonomic Gradient” (LTG). The LTG is characterized by a geographic trend in the quality of the information used for species delimitation. This uneven distribution of taxonomic knowledge can hamper estimates of global species richness (Stropp et al. 2022), affecting our ability to compare biodiversity data from different regions of the globe, from different biomes, or even data from different areas within a single biome. The methodological approaches and conceptual frameworks used to delimit species are beyond the scope of this article. Nonetheless, integrative analyses involving several biological dimensions (e.g., morphology, shape, behavior, molecular, and genetic variation; see Dayrat 2005, Will et al. 2005) are generally employed during species delimitation, leading to coherent and stable taxonomic units in terms of their past and future evolutionary responses.

More comprehensive and integrative approaches have been applied to species delimitation in the temperate regions (Freeman and Pennell 2021), leading to more stable taxa that are less prone to taxonomic changes over time in this part of the World. Conversely, the tropical regions house a larger number of species and clades that generally require substantial research efforts (Cayuela et al. 2011, Freestone and Inouye 2014). Tropical taxa are thus more prone to future taxonomic changes through splitting or lumping of species (Freeman and Pennell 2021; Stropp et al. 2022). This LTG is partially driven by differences in biological complexity in different parts of the world. Furthermore, social and economic challenges resulting from a lower number of taxonomists, lower investment in science, scientific colonialism, and political issues in the tropics (Balakrishnan 2005; Rodrigues et al. 2010; Costello et al. 2013), further aggravate the taxonomic disparities, contributing substantially for the establishment of the latitudinal taxonomic gradient.

The LTG is an interesting concept because it integrates biodiversity shortfalls and ignorance, with more effective taxonomic practices that consider an epistemological perspective on the social and economic processes guiding scientific practices worldwide. However, the LTG also reveals a clear link between the Linnean and Darwinian shortfalls, not previously discussed by Diniz-Filho et al. (2013). Although phylogenetic diversity metrics would be not widely affected by adding more species to the pool, these additions have a strong impact on estimates of speciation and extinction rates. Such a strong impact is largely due to the fact that the available diversification methods are sensitive to events occurring closer to the tips of phylogenies (see below). Indeed, when phylogenetic diversity metrics are applied in a geographical context (e.g., temperate vs. tropical comparisons), high uncertainty is introduced

into the analyses, biasing our interpretations of how diversification processes drive diversity in these regions (Rabosky et al. 2017; Schluter and Pennell 2017; Cássia-Silva et al. 2020; Tietje et al. 2022). In this context, the LTG (*sensu* Freeman and Pennell 2021) provides an interesting framework to explore how uneven knowledge about species taxonomy and geographical distribution may affect global diversification patterns.

## Overcoming and dealing with Latitudinal Taxonomic Gradients

Freeman and Pennell (2021) reviewed how LTGs can affect estimates of diversification rates and other macroecological patterns, including genetic diversity and population structure, responses to climate change, and extinction risks. According to their review, biases in estimates of diversification and other macroecological patterns can be mitigated by using standardized species concepts and improving the number and quality of taxonomic studies, especially in the tropics. While this represents a valid solution, it may be difficult to implement it, given the complicated social, political, and economic factors leading to unequal science practices globally. So, alternative solutions to deal with LTGs are welcome.

We argue that LTGs should also be viewed as a missing data problem, since there are many species ‘waiting’ to be described in the tropics (Joppa et al. 2011a; Moura and Jetz 2021). Missing data biases can be overcome by study design and statistical strategies (Little and Rubin 2019). Therefore, species may be missing completely at random along the phylogeny, causing an increase of uncertainty on the estimated parameters, but unbiased estimates. Also, species may have a missing likelihood correlated to certain known variables, which, when included in the models, can reduce estimate biases, as assumed here (e.g., Van Buuren 2018).

We suggest that, firstly, the probability of missing species needs to be defined. This probability may be decomposed into two components, related first to the species that have not yet been discovered, and second to the discovered species that are likely to undergo taxonomic changes, leading to either splitting a species into two or more taxa, or lumping several species into a single one. The first component of the Linnean shortfall consists of the number of unknown species, a topic that has been extensively discussed, with multiple approaches to estimate the total number of undescribed species available to date (e.g., species accumulation curves that account for taxonomic efforts and variation in lumping/clumping balances through time; see Joppa et al. 2011b). In turn, the second component of the Linnean shortfall consists of taxonomic changes affecting known species, either by splitting a species into two (or more) taxa, or by lumping several species into a single taxon. This latter component of the Linnean shortfall requires us to ask how certain taxonomists are about the validity of a given species or how well-established the taxonomy of a given species (or clade) is. The two components

of the Linnean shortfall are not independent and can occur sequentially.

There are several surrogates to measure the uncertainty of a species' validity, which can be divided in two main types: indicators of information completeness and estimates of taxonomic uncertainty. Indicators of information completeness can be estimated based on the amount of scientific (or taxonomic) research directed towards a given taxon, which can, in turn, be measured by the number of appearances of a given species in literature searches or searches conducted in scientific repositories (e.g., Google Scholar, Web of Science, Scopus). Estimates of this nature evaluate the effort devoted to studying the taxonomy of each taxon and are often complemented by measures of the rate of new species description or synonymizing over time (e.g., Medellín and Soberón 1999; Baselga et al. 2010). Rates close to zero indicate that the current alpha taxonomy is stable, with a low probability of taxonomic change. Although species additions may still happen due to lumping or splitting in subsequent revisionary work, these additions are relatively small when compared to the whole pool of described species. However, variations in the rate of species discovery (the first component of the Linnean shortfall) also depend on the distribution of taxonomic effort in geographic space, the taxonomic effort through the tree of life, and the taxonomic method employed. To account for all these issues, the amount of scientific effort can be coupled with macroecological knowledge about the geographic range size frequency distributions (RSFDs), species body sizes (with more recently described species showing smaller body sizes and restricted distribution), and the distance of individual species from research centers (e.g., Moura and Jetz 2021; Parsons et al. 2022).

To adequately address the methodological issues associated with the second component of the Linnean shortfall (i.e., changes in previously described species), it is important to keep in mind that estimates of taxonomic uncertainty go beyond the description of species accumulation curves through time. Instead, these estimates also seek to quantify the probability of future taxonomic change for individual taxa. In this context, a high probability of change reflects high taxonomic uncertainty and is associated with taxa that have an uncertain taxonomic delimitation. Changes in species delimitation generally result from lumping several species into a single taxon or splitting a "single" species into several taxa. Taxonomic lumping or splitting is not a zero-sum game where the total number of species remains relatively constant. Instead, certain groups and lineages can be subject to frequent splitting caused by the availability of new data showing that their populations pertain to different taxonomic entities (cryptic lineages) (e.g., Zachos et al. 2013; Delić et al. 2017; Damasco et al. 2021). In turn, other lineages can be reorganized by massive lumping based on detailed revisionary work that shows a continuum in morphological and genetic diversity (among others), indicating that taxonomically homogeneous entities have been independently

described more than once (e.g., Henderson 2020). Because cryptic lineages and taxonomic effort (and expertise) are often phylogenetically and geographically structured (Fontaneto et al. 2009; Rodrigues et al. 2010; Freitas et al. 2019; Blackburn et al. 2019), the distribution of taxonomic uncertainty throughout the phylogeny is uneven.

The uneven taxonomic uncertainty leads to unmeasured errors in diversification rates that can have a cascading impact on our understanding of how diversification rates relate to other aspects of biodiversity (e.g., species richness). To decrease these biases in diversification rates, it is important to consider the history of taxonomic progress for the taxa involved (Stropp et al. 2022) by evaluating the variation of valid species through time (Alroy 2002; Williams 2022). The probability of splitting or lumping a valid species in the future can be estimated by simulations (see below) that would allow the total number of species predicted under multiple simulations to match the overall number of species predicted by species accumulation curves (Diniz-Filho et al. 2005; Lobo and Martín-Piera 2002; Lobo et al. 2007) or the application of SAR/SAD (species–area relationships/species abundance distributions) analyses (ter Steege et al. 2020). Under this approach, higher mean split probabilities would lead to a higher number of 'breaks' in currently valid names and higher overall richness, allowing us to increase the mean split probabilities (i.e., a test of the relative effect of our ignorance of the coupled Linnean/Darwinian shortfall) and address questions such as: How many new species would need to be added to the phylogeny in order to turn around our understanding of macroecological patterns at different levels, i.e., the geographic patterns in diversification rates?

While approaches based on accumulation curves explicitly assume a correlation between the two components of the Linnean shortfall, other strategies could be developed for taxa with more stable accumulation curves. In this context, approaches that standardize the uncertainty in species validity and define split/lump probabilities are particularly relevant, especially when more pessimistic views on estimating the overall knowledge of species richness based on accumulation curves (Wilkinson et al. 2021) are considered. A simple example of the potential impact of the LTG can be obtained through an evaluation of how the high probability of splitting a poorly known species could change diversification rates (Fig. 2). In a hypothetical scheme, with four northern, well-known, and well-delimited species (A to D), and six southern, poorly-known, and poorly-delimited species (E to J), one might expect that the latter would have a higher probability of splitting. For example, the southern species E, G, and H could be rearranged into several new species (e.g., E1 and E2; G1 and G2; H1, H2, H3, and H4, respectively) after taxonomic revisions, whereas the southern species F, I, and J could remain valid and maintain their phylogenetic position. In this example, the splits in the southern clade would generate new hypothetical species that are closely related to the original ones, leading to an increase

from 10 to 15 valid species in the clade. Despite the increase of about 30% in the number of species, this increase would not drastically affect phylogenetic diversity (PD), which would only increase by about 11% (from 9.05 to 10.04), supporting earlier argument of robustness of PD estimates by Diniz-Filho et al. (2013; see also Lum et al. 2022).

However, a different situation emerges when we compare diversification rates between the northern vs. southern species using estimates based on Evolutionary Distinctiveness (DR; Jetz et al. 2012; Title and Rabosky 2019; see also Isaac et al. 2007). In this case, the northern region would show a much higher mean diversification rate ( $1.60 \pm 0.86$ ) than the southern one ( $1.14 \pm 0.32$ ). After accounting for the taxonomic uncertainty and splitting the southern species, the mean DR in the southern region would increase to  $1.73 \pm 0.39$ , surpassing the DR found in the north. Conversely, in a scenario where the northern species do not consist of independent evolutionary units (i.e., with any two species being lumped), DR would be reduced to  $0.98 \pm 0.33$  in the north, increasing the differences between the two regions. In other words, an inversion of the mean DR is observed after accounting for the LTG, with the higher species richness found in the south explained by higher DR.

Once the probabilities of splitting/lumping are defined, we can use simulation approaches to replicate this process, generating thousands of phylogenies, re-estimating the DRs for each species, and obtaining a distribution of the differences between the mean diversification rates (DR) in the two regions (Fig. 3). This approach could also be used to compare two or more biogeographic regions or to perform more detailed spatial analyses of DR through analyses that use other methods to estimate diversification rates, such as BAMM (Bayesian analysis of macroevolutionary mixtures; Rabosky 2014; Rabosky et al. 2017; Title and Rabosky 2019) or event-based methods such as GeoSSE or BioGeoBears (Goldberg et al. 2011; Matzke 2013).

It is important to highlight that the basic component of the Darwinian shortfall remains problematic for many groups of organisms, with many known species still lacking from the available phylogenetic frameworks. One way to handle this issue is by assigning Phylogenetically Unknown Taxa (PUTs) to the Most Derived Consensus Clades (MDCCs) through random approaches (Rangel et al. 2015). Alternatively, it is also possible to apply model-based approaches that take diversification rates into account (Thomas et al. 2013). From an operational perspective, simulations start by building new phylogenies after splitting some of the currently known species according to previously defined probabilities. This approach allows a hierarchical evaluation of the effects of the Darwinian shortfall (step 1) and the Linnean shortfall (step 2) while evaluating the uncertainties in diversification rates or any other macroecological patterns. Other more complex methodological approaches can be considered. For example, the fact that some of the currently described species are PUTs can be included as part of the definition of a species-splitting probability.

This approach can, however, generate interactions that can complicate hierarchical partitioning. Further developments and comparisons among different simulation approaches are necessary to define the best way to deal with the Linnean and Darwinian shortfalls simultaneously along LTGs.

## Geographical components of taxonomic uncertainty and the Wallacean shortfall

In our simple hypothetical example, the new species are found, in general, within the geographic ranges of the currently known species, and all of them are assigned to the same ancestral group with respect to the observed geographic distribution (northern or southern). However, for real applications, it would be necessary to evaluate how to assign newly derived species to one of the biogeographic regions (when using GeoSSE and related methods, for instance). This evaluation, in turn, involves a discussion about past speciation modes linking the new species. In general, this would be done by assuming the cohesion of the geographic distributions and, thus, assuming allopatric or parapatric speciation, but other scenarios are possible.

If a new sister species is described within the known occurrence of the “ancestral” species, in a single biogeographic region, it is straightforward to geographically assign it to the same region under allopatry or parapatry. However, if the currently known species is found in both regions, there are more possibilities to choose from. For instance, in a simple example of two derived species, both taxa can be found in both regions (their ranges overlap throughout both regions), or they can be symmetrically assigned to each of the regions (one to each region), or one can think that one of the species occurs only in one region as a small parapatric distribution, so that one of the regions can keep both species. These possibilities can also be randomly defined in the simulations, but it is also possible to use empirical patterns of RSFD, and their overlap among recently derived better-known species to define the most likely patterns of speciation.

Moreover, the discussion about the geographic position of the newly described sister species leads to another interesting connection among the Linnean, Darwinian, and Wallacean shortfalls. In practice, a first step is always to build species geographic extents of occurrence, which is now usually done based on local occurrence data (e.g., as available from GBIF). Dealing with such datasets, in turn, require procedures to evaluate data consistency and taxonomic resolution (e.g., Ribeiro et al. 2022), which may be challenging under the potential links between the three shortfalls. Additionally, we can use several methods to deal with this issue of occurrence data uncertainty and explore Wallacean shortfall, and a promising approach would be to use the ‘maps of biogeographical ignorance’ (MOBIs) of each currently known species (Tessarolo et al. 2021), or to use MOBIs created for entire groups altogether. MOBIs could be used to identify new sister species in poorly explored regions of the species’ range and

sometimes outside its currently known geographic extent. This association between the Linnean and Wallacean shortfall is well established and is illustrated by the derived species H1 in the hypothetical example mentioned above and would be a direct consequence of the interaction between taxonomic resolution and speciation dynamics under allopatry/parapatry (Fig. 2). Still, it is necessary to better evaluate at what rate expanding geographic extents of known species leads to the description of new species and how this relates to speciation dynamics (e.g., Skeel and Cardillo 2019).

Temperate versus tropical diversification patterns estimated using GeoSSE will likely not be very impacted by the Wallacean shortfall, given the usually highly right-skewed RSFDs (i.e., most species having very small geographic ranges; Gaston 1996; Castiglione et al. 2017). However, this impact would likely be more severe in analyses involving geographically more restricted areas such as ecoregions or biomes. Likewise, analyses evaluating diversification in a more explicit and continuous spatial context using methods that assign diversification rates to species (e.g., calculating DR for each species; see Fig. 2), would probably also be more severely impacted. In those cases, it would be necessary to assign a geographic centroid to the newly described species. Under these alternative speciation scenarios, it would be possible to simulate spatially explicit splitting and incorporate the uncertainty in the geographic coordinates of the centroids in the analyses hence, leading to a distribution of correlations between DR and latitude.

## Concluding Remarks

In this article, we have delved into the recent proposal of Latitudinal Taxonomy Gradients (Freeman and Pennell 2021) and highlight how this idea provides an interesting conceptual link between the Linnean, Darwinian, and Wallacean shortfalls. Apart from improving our understanding of LTGs “per se”, we also provide a basis for the development of new methods that better integrate different shortfalls. We further establish a basis for the development of methods aiming to estimate how the interactions between shortfalls may impact our understanding of the evolutionary component of richness gradients at macroecological scales.

## Author Contributions

JAFDF conceived the idea and wrote the first draft, with input from JH, LJ and JJMG. All authors contributed intellectually and participated in the final version of this manuscript.

## Acknowledgments

This manuscript is partially derived from a working group on “Biodiversity Shortfalls” held in November 2019 and sponsored by the National Institutes for Science and Technology (INCT) in Ecology, Evolution, and Biodiversity Conservation (CNPq proc. 465610/2014-5

and FAPEG proc. 201810267000023). JJMG and LEF are supported by Ph.D. and M.Sc. scholarships from CAPES, while LM and RBP are supported by postdoctoral fellowships from CAPES (PNPD). JS was funded by the funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Action (grant agreement #843234; project: TAXON-TIME) and by the Spanish Council for Scientific Research (IF\_ERC). GT and LJ are supported by a DTI fellowships from CNPq, while JAFD-F, LGL, and CJBC are supported by Productivity Grants from CNPq.

## References

- Alroy, J. (2002) How many named species are valid? *Proceedings of the National Academy of Sciences USA*, 99, 3706–3711. <https://doi.org/10.1073/pnas.06269109>
- Balakrishnan, R. (2005) Species concepts, species boundaries and species identification: a view from the tropics. *Systematic Biology*, 54, 689–693. <https://doi.org/10.1080/10635150590950308>
- Baselga, A., Lobo, J.M., Hortal, J., et al. (2010) Assessing alpha and beta taxonomy in eupelmid wasps: determinants of the probability of describing good species and synonyms. *Journal of Zoological Systematics and Evolutionary Research*, 48, 40–49. <https://doi.org/10.1111/j.1439-0469.2009.00523.x>
- Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F.L.V.B., et al. (2006) Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, 12, 475–482. <https://doi.org/10.1111/j.1366-9516.2006.00286.x>
- Blackburn, D.C., Giribet G., Soltis D.E. & Stanley E.L. (2019) Predicting the impact of describing new species on phylogenetic patterns. *Integrative Organismal Biology*, 1, 1–12. <https://doi.org/10.1093/iob/obz028>
- Van Buuren, S. (2018) Flexible imputation of missing data. CRC press. <https://doi.org/10.1201/9780429492259>
- Cássia-Silva, C., Freitas, C.G., Lemes, L.P., et al. (2020) Higher evolutionary rates in life-history traits in insular than in mainland palms. *Scientific Reports*, 10, 1–10. <https://doi.org/10.1038/s41598-020-78267-5>
- Castiglione, S., Mondanaro, A., Melchionna, M., et al. (2017) Diversification rates and the evolution of species range size frequency distribution.



- Frontiers in Ecology and Evolution, 5, 1–10. <https://doi.org/10.3389/fevo.2017.00147>
- Cayuela, L., De la Cruz, M. & Ruokolainen, K. (2011) A method to incorporate the effect of taxonomic uncertainty on multivariate analyses of ecological data. *Ecography*, 34, 94–102. <https://doi.org/10.1111/j.1600-0587.2009.05899.x>
- Costello, M.J., May, R.M. & Stork, N.E. (2013) Can we name Earth's species before they go extinct? *Science*, 339, 413–416. <https://doi.org/10.1126/science.1230318>
- Damasco, G., Baraloto, C., Vicentini, A., et al. (2021) Revisiting the hyperdominance of Neotropical tree species under a taxonomic, functional and evolutionary perspective. *Scientific Reports*, 11, 1–11. <https://doi.org/10.1038/s41598-021-88417-y>
- Dayrat, B. (2005) Towards integrative taxonomy. *Biological Journal of the Linnean Society*, 85, 407–415. <https://doi.org/10.1111/j.1095-8312.2005.00503.x>
- Delić, T., Trontelj, P., Rendoš, M. & Fišer C. (2017) The importance of naming cryptic species and the conservation of endemic subterranean amphipods. *Scientific Reports*, 7, 1–12. <https://doi.org/10.1038/s41598-017-02938-z>
- Diniz-Filho, J.A.F., Bastos, R.P., Rangel, T.F.L.V.B., et al. (2005) Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Global Ecology and Biogeography*, 14, 469–477. <https://doi.org/10.1111/j.1466-822X.2005.00165.x>
- Diniz-Filho, J.A.F., Loyola, R.D., Raia P., et al. (2013) Darwinian shortfalls in biodiversity conservation. *Trends in Ecology and Evolution*, 28, 689–695. <https://doi.org/10.1016/j.tree.2013.09.003>
- Faith, D.P. (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61, 1–10. [https://doi.org/10.1016/0006-3207\(92\)91201-3](https://doi.org/10.1016/0006-3207(92)91201-3)
- Fontaneto, D., Kaya, M., Herniou E.A., et al. (2009) Extreme levels of hidden diversity in microscopic animals (Rotifera) revealed by DNA taxonomy. *Molecular Phylogenetics and Evolution*, 53, 182–189. <https://doi.org/10.1016/j.ympev.2009.04.011>
- Freeman, B.G. & Pennell, M.W. (2021) The latitudinal taxonomy gradient. *Trends in Ecology and Evolution*, 36, 778–786. <https://doi.org/10.1016/j.tree.2021.05.003>
- Freestone, A.L. & Inouye, B.D. (2014) Non-random community assembly and high temporal turnover promote regional coexistence in tropics but not temperate zone. *Ecology*, 96, 264–273. <https://doi.org/10.1890/14-0145.1>
- Freitas, T. M. S., Stropp, J., Calegari, B. B., et al. (2021) Quantifying shortfalls in the knowledge on Neotropical Auchenipteridae fishes. *Fish and Fisheries*, 22, 87–104. <https://doi.org/10.1111/faf.12507>
- Fritz, S.A. & Rahbek, C. (2012) Global patterns of amphibian phylogenetic diversity. *Journal of Biogeography*, 39, 1373–1382. <https://doi.org/10.1111/j.1365-2699.2012.02757.x>
- Gaston, K.J. (1996) Species-range-size distributions: patterns, mechanisms and implications. *Trends in Ecology and Evolution*, 11, 197–201. [https://doi.org/10.1016/0169-5347\(96\)10027-6](https://doi.org/10.1016/0169-5347(96)10027-6)
- Goldberg, E. E., Lancaster, L. T. & Ree, R. H. (2011) Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. *Systematic Biology*, 60, 451–465. <https://doi.org/10.1093/sysbio/syr046>
- Harvey, P.H. & Pagel, M.D. (1991) *The comparative method in evolutionary biology*. Oxford University Press.
- Henderson, A. (2020) A revision of *Attalea* (Arecaceae, Arecoideae, Cocoseae, Attaleinae). *Phytotaxa*, 444, 1–76. <https://doi.org/10.11646/phytotaxa.444.1.1>
- Hortal, J., de Bello, F., Diniz-Filho, J.A.F., et al. (2015) Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46, 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Isaac, N.J.B., Mallet, J. & Mace, G.M. (2004) Taxonomic inflation: Its influence on macroecology and conservation. *Trends in Ecology and Evolution*, 19, 464–469. <https://doi.org/10.1016/j.tree.2004.06.004>
- Isaac, N.J.B., Turvey, S.T., Collen B., et al. (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE*, 2, e296. <https://doi.org/10.1371/journal.pone.0000296>

- Jetz, W., Thomas, G.H., Joy J.B., et al. (2012) The global diversity of birds in space and time. *Nature*, 491, 444–448. <https://doi.org/10.1038/nature11631>
- Joppa, L.N., Roberts, D.L., Myers, N. & Pimm S.L. (2011a). Biodiversity hotspots house most undiscovered plant species. *Proceedings of the National Academy of Sciences USA*, 108, 13171–13176. <https://doi.org/10.1073/pnas.1109389108>
- Joppa, L.N., Roberts, D.L. & Pimm, S.L. (2011b) How many species of flowering plants are there? *Proceedings of the Royal Society B*, 278, 554–559. <https://doi.org/10.1098/rspb.2010.1004>
- Little, R.J. & Rubin D.B. (2019) *Statistical analysis with missing data*. John Wiley & Sons. <https://doi.org/10.1002/9781119482260>
- Lobo, J.M. & Martín-Piera, F. (2002) Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. *Conservation Biology*, 16, 158–173. <https://doi.org/10.1046/j.1523-1739.2002.00211.x>
- Lobo, J.M., Baselga, A., Hortal, J., et al. (2007) How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Diversity and Distributions*, 13, 772–780. <https://doi.org/10.1111/j.1472-4642.2007.00383.x>
- Lum, D, Rheindt, F.E. & Chisholm, R.A. (2022) Tracking scientific discovery of avian phylogenetic diversity over 250 years. *Proceedings of the Royal Society of London B*, 289, 20220088. <https://doi.org/10.1098/rspb.2022.0088>
- Matzke, N. J. (2013) Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography*, 5, 242–248. <https://doi.org/10.21425/F5FBG19694>
- Medellín, R.A. & Soberón, J. (1999) Predictions of mammal diversity on four land masses. *Conservation Biology*, 13, 143–149. <https://doi.org/10.1046/j.1523-1739.1999.97315.x>
- Mouquet, N., Devictor, V., Meynard, C.N., et al. (2012) Ecophylogenetics: advances and perspectives. *Biological Reviews*, 87, 769–785. <https://doi.org/10.1111/j.1469-185X.2012.00224.x>
- Moura, M.R. & Jetz, W. (2021) Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nature Ecology and Evolution*, 5, 631–639. <https://doi.org/10.1038/s41559-021-01411-5>
- Nürk, N.M., Linder, H.P., Onstein, R.E., et al. (2020) Diversification in evolutionary arenas - assessment and synthesis. *Ecology and Evolution*, 10, 6163–6182. <https://doi.org/10.1002/ece3.6313>
- Parsons, D.J., Pelletier, T.A., Wieringa, J.G., et al. (2022) Analysis of biodiversity data suggests that mammal species are hidden in predictable places. *Proceedings of the National Academy of Sciences USA*, 119, e2103400119. <https://doi.org/10.1073/pnas.2103400119>
- Rabosky DL (2014) Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9, e89543. <https://doi.org/10.1371/journal.pone.0089543>
- Rabosky DL, Mitchell JS, Chang J (2017) Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. *Systematic Biology*, 66, 477–498. <https://doi.org/10.1093/sysbio/syx037>
- Rabosky, D.L., Chang, J., Title, P.O., et al. (2018) An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559, 392–395. <https://doi.org/10.1038/s41586-018-0273-1>
- Rangel, T.F., Colwell, R.K., Graves G.R., et al. (2015) Phylogenetic uncertainty revisited: implications for ecological analyses. *Evolution*, 69, 1301–1312. <https://doi.org/10.1111/evo.12644>
- Ribeiro BR, Velazco SJ, Guidoni-Martins K et al (2022) bdc: a toolkit for standardizing, integrating and cleaning biodiversity data. *Methods in Ecology and Evolution*, 13, 1421–1428. <https://doi.org/10.1111/2041-210X.13868>
- Rodrigues, A.S.L., Gray, C.L., Crowter, B.J., et al. (2010) A global assessment of amphibian taxonomic effort and expertise. *BioScience*, 60, 798–806. <https://doi.org/10.1525/bio.2010.60.10.6>
- Santos, A.M.C., Jones, O.R., Quicke, D.L.J. & Hortal, J. (2010) Assessing the reliability of biodiversity databases: identifying evenly inventoried island parasitoid faunas (Hymenoptera: Ichneumonidea) worldwide. *Insect*

- Conservation and Diversity, 3, 72-82. <https://doi.org/10.1111/j.1752-4598.2010.00079.x>
- Schluter, D. & Pennell, M.W. (2017) Speciation gradients and the distribution of biodiversity. *Nature*, 546, 48–55. <https://doi.org/10.1038/nature22897>
- Skeels, A. & Cardillo, M. (2019) Reconstructing the geography of speciation from contemporary biodiversity data. *American Naturalist*, 193, 240-255. <https://doi.org/10.1086/701125>
- ter Steege, H., Prado, P.I., Lima, R.A.F., et al. (2020) Biased-corrected richness estimates for the Amazonian tree flora. *Scientific Reports*, 10, 1–13. <https://doi.org/10.1038/s41598-020-66686-3>
- Stropp J., Ladle R.J., Emilio T., et al. (2022) Taxonomic uncertainty and the challenge of estimating global species richness. *Journal of Biogeography*, 49, 1654–1656. <https://doi.org/10.1111/jbi.14463>
- Tessarolo, G., Ladle, R.J., Lobo, J.M., et al. (2021) Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography*, 44, 1–13. <https://doi.org/10.1111/ecog.05793>
- Thomas, G.H., Hartmann, K., Jetz, W., et al. (2013) PASTIS: An R package to facilitate phylogenetic assembly with soft taxonomic inferences. *Methods in Ecology and Evolution*, 4, 1011–1017. <https://doi.org/10.1111/2041-210X.12117>
- Tietje, M., Antonelli, A., Baker, W.J., et al. (2022) Global variation in diversification rate and species richness are unlinked in plants. *Proceedings of the National Academy of Sciences USA*, 119, 1–9. <https://doi.org/10.1073/pnas.2120662119>
- Title, P.O. & Rabosky, D.L. (2019) Tip rates, phylogenies and diversification: what are we estimating, and how good are the estimates? *Methods in Ecology and Evolution*, 10, 821–834. <https://doi.org/10.1111/2041-210X.13153>
- Tucker, C.M. & Cadotte, M.W. (2013) Unifying measures of biodiversity: understanding when richness and phylogenetic diversity should be congruent. *Diversity and Distributions*, 19, 845–854. <https://doi.org/10.1111/ddi.12087>
- Tucker, C.M., Aze, T., Cadotte, M.W., et al. (2019) Assessing the utility of conserving evolutionary history. *Biological Reviews*, 94, 1740-1760. <https://doi.org/10.1111/brv.12526>
- Voskamp, A., Baker, D.J., Stephens P.A., et al. (2017) Global patterns in the divergence between phylogenetic diversity and species richness in terrestrial birds. *Journal of Biogeography*, 44, 709–721. <https://doi.org/10.1111/jbi.12916>
- Whittaker, R.J., Araújo, M.B., Jepson, P., et al. (2005) Conservation biogeography: assessment and prospect. *Diversity and Distributions*, 11, 3-23. <https://doi.org/10.1111/j.1366-9516.2005.00143.x>
- Wilkinson, B.H., Ivany, L.C. & Drummond, C.N. (2021) Estimating vertebrate biodiversity using the tempo of taxonomy - a view from Hubbert's peak. *Biological Journal of the Linnean Society*, 134, 402–422. <https://doi.org/10.1093/biolinnean/blab080>
- Will, K.W., Mishler, B.D & Wheeler, Q.D. (2005) The perils of DNA barcoding and the need for integrative taxonomy. *Systematic Biology*, 54, 844–851. <https://doi.org/10.1080/10635150500354878>
- Williams, P.H. (2022) Novel splitting/lumping index reflects the history of species concepts applied to bumblebees (Insecta: Apidae). *Zoological Journal of the Linnean Society*, 196, 704–719. <https://doi.org/10.1093/zoolinnean/zlab123>
- Zachos F.E., Apollonio M., Bärmann E. V., et al. (2013) Species inflation and taxonomic artefacts – a critical comment on recent trends in mammalian classification. *Mammalian Biology*, 78, 1–6. <https://doi.org/10.1016/j.mambio.2012.07.083>

Submitted: 9 December 2022

First decision: 17 January 2023

Accepted: 10 March 2023

Edited by Ole R. Vetaas and Robert J. Whittaker