Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity

6 authors, including:

Joaquín Hortal
Spanish National Research Council
169 PUBLICATIONS 5,068 CITATIONS

Jose Alexandre Felizola Diniz-Filho
Universidade Federal de Goiás
391 PUBLICATIONS 12,784 CITATIONS

Thomas M Lewinsohn
University of Campinas, Campinas, Brazil (UN…
128 PUBLICATIONS 2,757 CITATIONS

Jorge Lobo
Spanish National Research Council
318 PUBLICATIONS 10,276 CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Dung Removal Rates in Grasslands around the World View project

Project   Geographic patterns of diversification in the Neotropics View project

7

8 **Title page**

9 # Seven shortfalls that beset large-scale knowledge on
10 biodiversity

11 Joaquín Hortal[1,2,3*], Francesco de Bello[4,5], José Alexandre F. Diniz-Filho[2], Thomas M.
12 Lewinsohn[6], Jorge M. Lobo[1] and Richard J. Ladle[7,8,*]

13

14 1. Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales
15 (MNCN-CSIC), Madrid, Spain, email: jhortal@mncn.csic.es (JH), mcnj117@mncn.csic.es (JML)
16 2. Departamento de Ecologia, Instituto de Ciências Biologicas, Universidade Federal de Goiás,
17 Goiânia, Brazil, email: diniz@icb.ufg.br (JAFD-F)
18 3. cE3c, Centre for Ecology, Evolution and Environmental Changes –Faculdade de Ciências,
19 Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
20 4. Institute of Botany, Czech Academy of Sciences, Třeboň, Czech Republic, email:
21 fradebello@ctfc.es
22 5. Department of Botany, Faculty of Sciences, University of South Bohemia, Budějovice, Czech
23 Republic
24 6. Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas,
25 Brazil, email: thomasl@unicamp.br
26 7. ICBS, Universidade Federal de Alagoas, Maceió, Brazil
27 8. School of Geography and the Environment, University of Oxford, UK, email:
28 richard.ladle@ouce.ox.ac.uk

29 *Authors for correspondence:

30 Joaquín Hortal, Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias
31 Naturales (MNCN-CSIC), C/Jose Gutierrez Abascal 2, 28006 Madrid, Spain, email:
32 jhortal@mncn.csic.es

33 Richard J. Ladle, Institute of Biological and Health Sciences, Federal University of Alagoas,
34 Praça Afrânio Jorge, s/n, Prado, CEP: 57010-020, Maceió, AL, Brazil, email:
35 richard.ladle@ouce.ox.ac.uk

36

37

38 **Table of contents**

62

63

64

Keywords

Abstract

Ecologists and evolutionary biologists are increasingly using big-data approaches to tackle questions at large spatial, taxonomic and temporal scales. However, despite recent efforts to gather two centuries of biodiversity inventories into comprehensive databases, many crucial research questions remain unanswered. Here, we update the concept of knowledge shortfalls and review the tradeoffs between generality and uncertainty. We present seven key shortfalls of current biodiversity data. Four previously proposed shortfalls pinpoint knowledge gaps for species taxonomy (Linnean), distributions (Wallacean), abundance (Prestonian) and evolutionary patterns (Darwinian). We also redefine the Hutchinsonian shortfall for abiotic tolerances of species, and propose new shortfalls relating to limited knowledge of species traits (Raunkiaeran) and biotic interactions (Eltonian). We conclude with a general framework for the combined impacts and consequences of shortfalls of large-scale biodiversity knowledge for evolutionary and ecological research and consider ways of overcoming the seven shortfalls and dealing with the uncertainty they generate.

1. Introduction

Faced with the almost overwhelming complexity of the natural world, biologists have always sought to categorize and classify organisms in their search for patterns, underlying processes and organizing principles (Gleason 1926). Inevitably, such classifications reflect the goals and interests of the classifiers, so that they are abstractions that represent the 'real world' by surrogates within which 'scientific knowledge' is produced (Rosen 1996). Thus, knowledge (and ignorance) of nature is fundamentally influenced by the ways in which biological entities are classified and atomized into readily grasped units (e.g. communities, species, clades, traits, genes, etc.) for scientific usage. The ways in which biodiversity is measured should therefore be viewed as a limited subset of the myriad ways that the diversity of life *could* be classified. Moreover, within this narrow range of information, complete knowledge for any given characteristic of biodiversity is practically unachievable, due to the interaction between the complex temporal and spatial dynamics of nature and human capacity to survey it (Ladle and Hortal 2013). This unevenness in survey effort and research infrastructure results in high spatial and temporal variation in the quality and reliability of the data available for biodiversity research and conservation planning (Gaston and Rodrigues 2003, Mace 2004).

The fundamental and practical limits on biodiversity knowledge mean that scientists have to work with incomplete and often unrepresentative data on a limited number of organisms and their characteristics. The gaps, or *shortfalls*, in knowledge about the identity,

distribution, evolution and dynamics of global biodiversity need to be carefully recognized and quantified, since biased and unrepresentative knowledge compromises the capacity to describe existing biodiversity or make accurate predictions about how it might change in the future. Biased data can also lead to misidentification of ecological and evolutionary processes and inefficient use of limited conservation resources.

The objectives of this review are therefore to: i) identify key shortfalls in biodiversity knowledge; ii) review the origins, drivers and current explanations for these shortfalls; iii) assess the consequences of these shortfalls for ecological, evolutionary and conservation research, and; iv) propose strategies and tools by which these shortfalls may be overcome and, in parallel, how uncertainties and biases in biodiversity data can be most effectively factored into research and conservation practice.

SIDEBAR: THE IMPORTANCE OF IGNORANCE

"Thoroughly conscious ignorance is the prelude to every real advance in science." (James Clerk Maxwell, cited in Firestein 2012). In science ignorance refers to what we do not know. If classifying and understanding the known and discovering the unknown are the ultimate objectives of science, informed ignorance is a powerful research tool. A 'conscious ignorance' may help identify the right questions, avoid developing theories based on insufficient or misleading data, and ultimately, direct scientific research towards significant advancements. By expanding the known in a thoughtful, structured way, good science also increases perceived ignorance, i.e. the amount of the unknown that can be reached through further research.

2. The Shortfalls

Data scarceness, limited description of patterns and processes and gaps in theory are characteristic of all domains of ecology and evolution. Indeed, the existence of shortfalls in biodiversity knowledge is, arguably, a direct consequence of the complexity generated by an evolutionary system in which the rates of production of novel entities exceed the maximum rates at which we can describe them. In this sense, knowledge shortfalls can be defined as the gap between realized/extant knowledge and "complete knowledge" within a biological domain at a given moment of time (normally present day). We suggest that biodiversity data shortfalls can be broadly grouped into seven major categories (Table 1) corresponding to the knowledge domains of systematics, biogeography, population biology, evolution, functional (trait-based) ecology, abiotic tolerances and ecological interactions – five of these shortfalls having been proposed elsewhere (see Table 1). Being based on knowledge domains, the shortfalls are not exclusive to certain types of data. Rather, they may affect all or several of the different aspects studied from a source of information. A good example is the fossil record, to which virtually all shortfalls apply.

142   **Table 1.** Definitions (and original references) for the seven main shortfalls of biodiversity knowledge.

The seven main shortfalls of biodiversity knowledge
Linnean shortfall – Most of the species on Earth have not been described and catalogued (Brown & Lomolino 1998); this concept can be extended to extinct species (this review)
Wallacean shortfall – The knowledge on the geographic distribution of most species is incomplete, being most times inadequate at all scales (Lomolino 2004)
Prestonian shortfall – Lack of data on species abundances and their dynamics in space and time are often scarce (Cardoso et al. 2011)
Darwinian shortfall – Lack of knowledge about the tree of life and evolution of species and their traits (Diniz-Filho et al. 2013)
Raunkiæran shortfall – Lack of knowledge on species' traits and their ecological functions (this review)
Hutchinsonian shortfall – Lack of knowledge about the responses and tolerances of species to abiotic conditions (i.e., their scenopoetic niche) (this review, redefined from Cardoso et al. 2011)
Eltonian shortfall – Lack of enough knowledge on species' interactions and their effects on individual survival and fitness (this review)

143

2.1. Linnean shortfall

145   The Linnean shortfall is named after Karl von Linné, or 'Linnaeus' (1707 – 1778) and refers to
146   the discrepancy between formally described species and the number of species that actually
147   exist (Lomolino 2004). We propose that this shortfall should also include the knowledge gap on
148   extinct species. The Linnean shortfall is increasingly severe for organisms that are smaller in
149   size, complexity, niche width, distributional range and which are less phenotypically
150   conspicuous, with this pattern holding both between and within taxonomic groups (Riddle et
151   al. 2011). The magnitude of the Linnean shortfall is unknown for two reasons. First, the
152   number of formally described species is constantly changing due to new descriptions, revisions
153   and unresolved synonyms (Chapman 2009; May 2010; Baselga 2010) as well as difficulties in
154   establishing a unified species concept or agreement on operational tools to delimit different
155   taxa (Dayrat 2005; Hebert & Gregory 2005). The most comprehensive and authoritative global
156   index of species is the Catalogue of Life (www.catalogueoflife.org), which currently (January
157   2015) has records for more than 1.6 million species (Roskov et al. 2014). Second, the predicted
158   number of species is highly sensitive to the estimation method adopted and to parameter
159   values: estimates range from 2 to up 100 million eukaryotic species (May 2010) with more
160   recent global species richness estimates converging on a narrower band of 2–10 million
161   species (Mora et al. 2011; Costello et al. 2012; Caley et al. 2014).

162   The Linnean shortfall incorporates two distinct categories of unknown species: those
163   yet to be sampled, and collected species that have not yet been described. Species in the
164   former category are most frequent in the large unsurveyed regions of the world. These include
165   remote regions such as the forests of southwest Amazonia (Bush & Lovejoy 2007; Hopkins
166   2007), but also poorly studied ecosystems such as the deep-sea (Rex & Etter 2010) or the
167   upper canopies of rainforests (Ellwood & Foster 2004). Collected, but as yet undescribed,
168   species may run into the hundreds of thousands and are largely a consequence of the lack of
169   funding and capacity in global taxonomy (see sidebar). Including these undocumented species
170   in the Catalogue of Life may be problematic since new entries should be validated by an expert
171   after scrutinizing descriptions and specimens – in many cases, even though the taxonomic
172   groups are still extant, their corresponding experts no longer are (Hopkins & Freckleton 2002).

174   *"To know that we know what we know, and that we do not know what we do not know, this is true*
175   *knowledge." (Henry David Thoreau, Walden, 1854). Getting an overview of knowledge (and lack of it) for*
176   *a given topic involves determining what we do know, what we are aware that we do not know, and*
177   *recognizing that there are facts far beyond our current knowledge. Or, according to Jackson's (2012)*
178   *bestiary of ignorance, determining the 'known knowns', the 'known unknowns', and the 'unknown*
179   *unknowns'. Jackson's classification tacitly includes a fourth category, the 'unknown knowns': facts that*
180   *we have recorded, but which are not easily accessible, or are so basic that we are unaware that we know*
181   *them. Despite not being a shortfall, this latter category is particularly important for biodiversity research*
182   *since it includes the information stored in Natural History collections and "grey literature", which could*
183   *be made accessible with new bioinformatics tools.*

184

2.2. Wallacean Shortfall

186   The Wallacean shortfall is named after Alfred Russel Wallace (1823–1913) and refers to the
187   lack of knowledge about the geographical distribution of species (Lomolino 2004). This
188   shortfall stems from geographic biases in the information on species distributions (Figure 1a)
189   causing many maps of observed biodiversity to closely emulate maps of survey effort (Hortal
190   et al. 2007). Such coincidence also depends on the grain at which species' distributions are
191   recorded and analyzed (see section 3.2). Specifically, knowledge of species distributions is
192   intimately connected with temporal and spatial variation in surveying effort (Hortal et al. 2008,
193   Boakes et al. 2010). That some regions are better sampled than others is inevitable given the
194   stark differences in scientific capacity and accessibility between countries and regions
195   (Rodrigues et al. 2010). For example, the Wallacean shortfall is particularly acute in remote
196   and inaccessible regions, such as the forests of southwest Amazonia and the Congo basin.
197   Approximately 40% of Amazonia has never been surveyed and we do not have an accurate
198   geographic distribution for any of the plant species that occur in this region (Bush and Lovejoy
199   2007). It should also be noted that distribution data typically varies in relation to political
200   rather than ecological units, and may therefore be heavily biased in relation to historical
201   patterns of collecting, analyzing and collating biogeographical data (Rodrigues et al. 2010,
202   Stropp et al. 2015, Meyer et al. 2015).

203

204

205

206

207
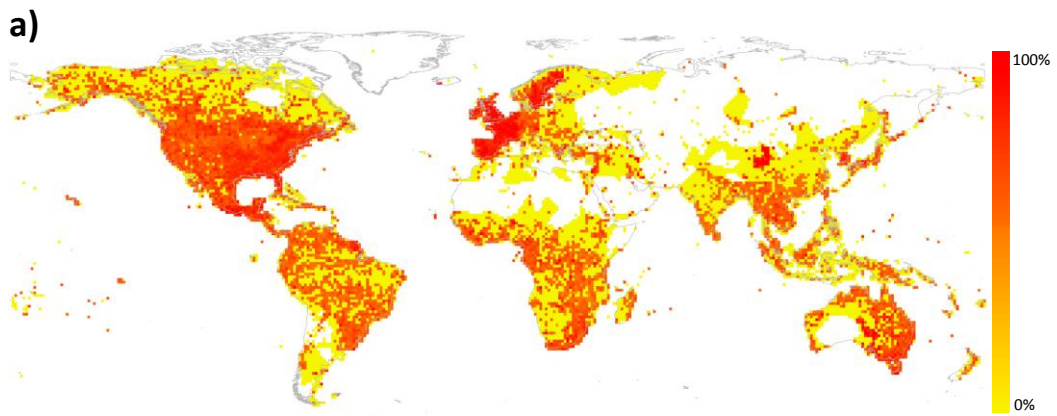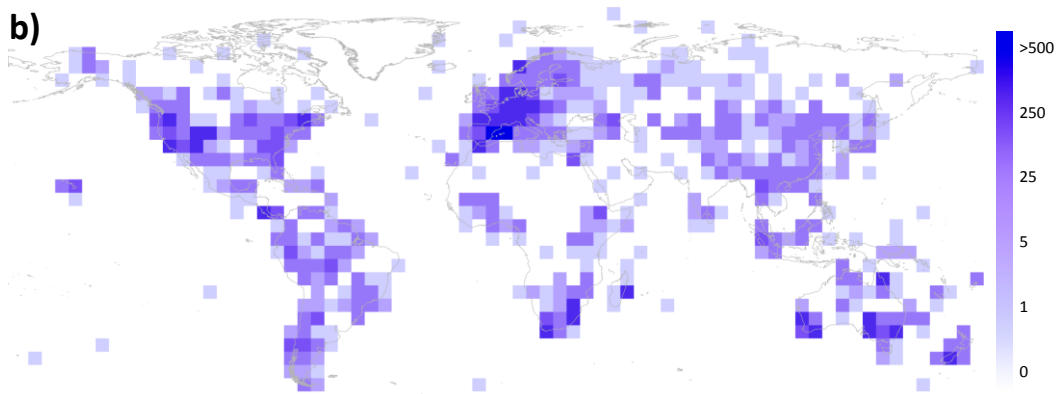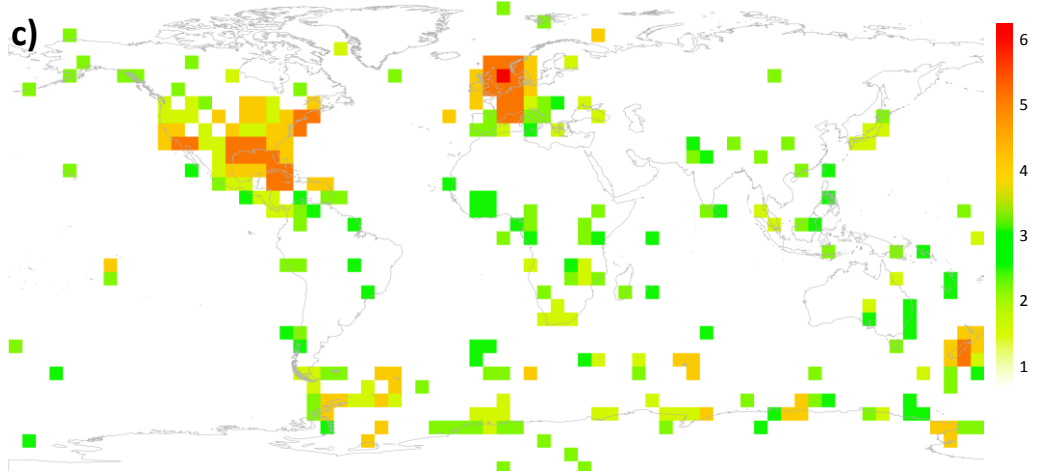
Figure 1. Examples of global distribution of biodiversity data. (a) Completeness of freshwater fish inventories at 1° x 1º grid cell resolution; data from *IPez* and *FishBase*, redrawn from Pelayo-Villamil et al. (2014); the scale depicts the percentage of the total (estimated) species that has been already observed at each cell. Data on (b) plant functional traits and (c) biotic interactions gathered into global databasing initiatives at each 5° x 5º grid cell. Plant functional traits come from TRY database (Kattge et al. 2011), and the scale indicates number of trait measurement sites (data courtesy of Jens Kattge, updated to January 2014; note that the scale has been transformed). Species' interactions come from GloBI (Poelen et al. 2014), and the scale indicates number of recorded interactions (data courtesy of Jorrit Poelen, extracted November 2014).

217

2.3. Prestonian shortfall

The Prestonian shortfall can be defined as lack of knowledge on the abundance of species and its population dynamics in space and time (Cardoso et al. 2011). It is named after Frank W. Preston (1896–1989) whose pioneering work on species commonness and rarity laid the conceptual groundwork for much contemporary ecological thought. Despite the fundamental importance of abundance data for addressing many ecological questions, such information is scarce for most species (Cardoso et al. 2011). This lack of data is due three main factors: i) the difficulty of producing accurate censuses for many animals; ii) the costs of gathering long-term data (Wolfe et al. 1987), and; iii) the rapid fluctuations in species abundance that necessitate frequent re-sampling. The Global Population Dynamics Database (GPDD) is addressing this shortfall by collating and compiling comparable data (Inchausti and Halley 2001). However, uncertainties in estimates of population size can affect the results (and conclusions) of analyses of the data in the GPDD (and most population time series) (Clark and Bjørnstad 2004, Knape and de Valpine 2012). The duration of studies on abundance dynamics also affects estimates of Minimum Viable Populations (Reed et al. 2003), demonstrating the need for long-term monitoring schemes capable of identifying population trends in time and space.

2.4. Darwinian shortfall

The Darwinian shortfall is named after Charles Darwin (1809–1882) and refers to the lack of knowledge about the tree of life and evolution of lineages, species and traits (Diniz-Filho et al. 2013). Since the 1950s, several methods to reconstruct phylogenetic relationships among species have been proposed (Felsenstein 1985, Hall 2011), and the entire field has advanced rapidly in the last 20 years. The increasing availability of molecular data and the development of powerful new computational methods triggered the wide application of phylogenetic comparative methods to understand trait evolution and biodiversity patterns (Nee & May 1997, Sechrest et al. 2002, Mace et al. 2003). Nevertheless, three issues still need to be sequentially solved to allow more consistent ecological comparative analysis (Diniz-Filho et al. 2013). First, we still have a limited understanding of the phylogenetic relationships among all living species; despite exponential growth of the number of phylogenies available for distinct taxa, many species are missing and many available phylogenies only establish relationships among higher taxonomic groups (e.g., Bininda-Emonds et al. 2007, Pyron & Wiens 2011, Jetz et al. 2012). Second, although phylogenies based on molecular data branch lengths are usually estimated with acceptable accuracy, there are no studies evaluating how the errors in estimating these branch lengths at distinct levels of the phylogeny will affect estimates of phylogenetic diversity based on different metrics. Moreover, calibrating these lengths to calculate absolute time is still challenging. This is important because understanding evolutionary rates and biogeographical patterns requires robust estimates of time since divergence (Dornburg et al. 2011, Lukoschek et al. 2012, Slater & Harmon 2013). Third, understanding how ecological traits relate to biodiversity patterns is hampered because current phylogenies provide limited information about trait evolution, even when they are accurate (Freckleton et al. 2002, Blomberg et al. 2003, Cadotte et al. 2013). Using phylogeny as a surrogate of trait variation assumes a linear relationship between trait differentiation among species and time since their divergence (see Hansen & Martins 1996). This is true only under

neutral processes, and fitting more complex models to describe evolutionary divergence in ecological traits and biodiversity patterns is still in its infancy (see Alfaro et al. 2009, Eastman et al. 2011, Morlon 2014). Furthermore, the information provided by the fossil record is often limited and uneven, affecting estimates of phylogenetic structure and trait evolution (Losos et al. 2011, Sansom et al. 2015).

2.5. Raunkiaeran shortfall

We define the Raunkiaeran shortfall as the lack of knowledge about ecologically relevant species traits. This includes trait variations both within (Kingsolver et al. 2001) and between species (Roy & Foote 1997), but also the ecological function (or functions) played by each trait, how these functions are affected by interactions with other traits, and which traits act together as bundles to perform specific ecosystem functions (Díaz et al. 2013). This shortfall is named after the Danish botanist Christen Raunkiaer (1860–1938), the creator of the homonymous plant life-form classification. Historically, interest has focused on documenting temporal trends in the evolution and selection of morphological and life history traits (Roy & Foote 1997, Kingsolver et al. 2001). This approach has often been done without an explicit or direct link to the organisms' fitness (e.g. Ricklefs 2012), which is implicitly of interest in this type of analyses. Functional trait-based approaches are increasingly being used for a wide range of applications in ecological and evolutionary research (McGill et al. 2006, de Bello et al. 2010). Such approaches are based on describing organisms in terms of their functional traits rather than their taxonomic or phylogenetic affiliations.

Functional traits are defined as any phenotypic attribute that affects a species' individual fitness and population dynamics and/or their influence on other organisms and ecosystem functions (Violle et al. 2007). Functional traits are either used for explicit comparisons between populations, species and communities subject to different environmental conditions (Cornelissen et al. 2003, Poorter et al. 2008, Albert et al. 2010, Laughlin and Laughlin 2013) or to investigate the effects of species on multiple ecosystem processes (de Bello et al. 2010, Díaz et al. 2013). A major characteristic of the Raunkiaeran shortfall is that the traits that are generally measured are often the most simple, rather than the most 'functional'. There is also some bias in the functional traits studied by plant and animal ecologists. While the former have traditionally focused on traits that either mediate the responses of species to environmental gradients and biotic interactions or are related to the ecosystem functions, the latter have mainly studied how traits mediate the responses of species to biotic and abiotic conditions (de Bello et al. 2010). Considerable progress has been achieved in defining a common set of useful traits for different taxa, especially plants, and in standardizing sampling protocols (e.g. Cornelissen et al. 2003). Such standardization has, in turn, promoted data-sharing and the development of online trait databases (e.g. Kattge et al. 2011). Nevertheless, significant taxonomic and geographic gaps remain (Figure 1b).

The clear standardization of trait definitions and measurements by Cornelissen et al. (2003) has fostered data gathering for plants (Kattge et al. 2010). However, such standardizations are yet not available for many taxa, hampering data gathering and sharing (Violle et al. 2007). Moreover, such standardization may result in concentrating research

303 efforts on a limited number of traits and, by extension, diverting research from other
304 potentially interesting traits for specific groups or ecological functions. In this context,
305 intraspecific variability in traits is often neglected, leading to biases (Albert et al. 2010) and
306 limiting the accuracy of traits retrieved from a database depending on the level of aggregation,
307 the trait and the habitat type (Cordlandwehr et al. 2013). The success of trait-based
308 approaches also depends on the selection of a sufficient number of informative functional
309 traits (Laughlin and Laughlin 2013), either linked to tolerance of abiotic and biotic conditions,
310 or to the effects of species on ecosystems (Cornelissen et al. 2003). This selection is often
311 arbitrary, reducing the generality of results. Easily measurable traits are usually chosen, even
312 though their relation to fitness components is often tenuous (Violle et al. 2007). Specifically,
313 this approach favours morphological traits over behavioural, physiological or life-history traits
314 (see, e.g., Kingsolver et al. 2001). Furthermore, while there is evidence that clusters of traits
315 influence demographic and fitness components (Poorter et al. 2008; Adler et al. 2014), the
316 functional consequences of trait covariation have been poorly studied.

317

2.6. Hutchinsonian shortfall

319 We define the Hutchinsonian shortfall as the lack of knowledge about the tolerances of species
320 to abiotic conditions – their Grinnellian niche (*sensu* Soberón et al. 2007). This shortfall is
321 named after George Evelyn Hutchinson (1903–1991), who established the modern concept of
322 the niche. This shortfall was originally defined by Cardoso et al. (2011) as gaps in the combined
323 knowledge of species' life histories, functional roles and responses to habitat changes.
324 However, for clarity and conceptual homogeneity we limit it here to the responses to
325 scenopoetic conditions (including habitat, climate, soil, water, among others), which are not
326 subject to depletion or modification by organisms (Hutchinson 1978). Correspondingly, we
327 assign lack of knowledge about other aspects of species functionality to either the Raunkiaeran
328 or Eltonian shortfalls.

329   The estimation of environmental preferences is a fundamental objective of ecology.
330 These estimates can be used to improve predictions of the responses of species to changing
331 conditions and, ultimately, increase understanding of abundance and occurrence patterns
332 (Brown 1984, Gaston 2003, Slatyer et al. 2013). Species tolerances are inferred from one of
333 two sources (Bozinovic et al. 2011, Sánchez-Fernández et al. 2012, Diamond et al. 2012): i)
334 physiologically-derived niche data –physiological data and performance curves generally
335 obtained under complicated and costly laboratory experiments in the case of animals
336 (Bozinovic et al. 2011), but also plants in greenhouse experiments– and/or field
337 experimentations for some plant and invertebrate species (Lambers et al. 1998, Diamond et al.
338 2012), and; ii) occurrence-derived niche data estimated from field observations, currently
339 obtained by the use of correlative niche models (Peterson et al. 2011). Although
340 physiologically-derived niches are generally more precise, they are not necessarily more
341 realistic due to variability in response curves among conspecific populations, phenotypic
342 plasticity and acclimation ability of each study population (e.g. McCann et al. 2014), and
343 evolutionary and heritable changes in physiological tolerances (Logan et al. 2014). Thus, niche

data obtained under laboratory conditions provide only a partial representation of Grinnellian niches, depending on the temporal and spatial span and range of conditions used.

Occurrence-derived niches can also produce misleading values of species tolerances and optima. The geographical distributions of species occurrences and environmental variables define the realized niche of the species (the so-called "Hutchinson duality"; Colwell and Rangel 2009). However, both species and environment typically show a spatially-autocorrelated structure (Diniz-Filho et al. 2003), which often results in an apparent correlation between species presences and environmental conditions. This is regardless of the origin of the spatial structure in species populations. However, many factors other than the environment can create such spatially-autocorrelated structure, leading to inaccurate estimates of climatic preferences. For example, when allopatric speciation processes are dominant, they tend to generate uneven environmental signatures, leading to incomplete and biased estimates of niches (Warren et al. 2014). Other biotic, historical or dispersal limitation processes may also generate non-equilibrium species distributions (Gouveia et al. 2014). The lack of agreement between the resolution of environmental data and the area effectively used by a single population (i.e., the mean home range of species; Dunning et al. 1995, Boyce 2006) –in addition to lack of information about microenvironmental variability within each locality (see McInerny and Purves 2011)– also compromises estimates of niche parameters from occurrence data.

2.7. Eltonian shortfall

The Eltonian shortfall is named after Charles Sutherland Elton (1900–1991), the pioneer of the concept of food chains and food webs. It refers to the lack of knowledge about interactions among species or among groups of species. It is qualitatively different from the other shortfalls which refer to the taxonomic, evolutionary, biogeographic or functional attributes of a taxon/clade. While such attributes are all recognized aspects of biodiversity, interactions and other processes can be viewed as an extraneous expansion of the concept. Interactions are often framed in terms of 'interaction diversity', a term coined by Thompson (1997) to refer to the set of biotic links between two sets of species. It is measured from matrices where each filled cell represents a recorded interaction between a pair of species, and values in cells may represent frequencies of interaction events or interaction strengths measured, for example, as consumed biomass or individuals. These matrices can be described using standard diversity metrics such as the Shannon or Simpson diversity indices (Blüthgen et al. 2008). Moreover, distributions of links across rows and columns can be explored as interaction patterns through categorical analysis, multivariate ordination or network structure (Lewinsohn et al. 2006). From these patterns metrics can be obtained which characterize functional aspects of community structure, such as distribution of trophic specialization or connectivity (Blüthgen et al. 2008). Interaction diversity and functional diversity complement species-based diversity research and act as bridges to study ecosystem processes such as food-web organization or species loss (Valiente-Banuet et al. 2014).

The Eltonian shortfall is sensitive to the degree of aggregation into higher taxonomic levels (Martinez 1993), trophospecies (Yodzis 1988) or other functional units (Ings et al. 2009).

11

386 Further, data on interactions gathered by different observers often differs due to variation in
387 collection methods, research objectives and criteria. This limits generalization of conclusions
388 from compilations and databases (e.g. Pimm et al. 1991). In particular, the choice of recording
389 method has a profound influence on the extent and quality of data and results. Interactions
390 which involve fleeting contacts have to be recorded by direct observation, e.g. of pollination
391 visits, which usually requires massive field effort (e.g. 171h of field observations in a desert
392 ecosystem recorded 55% of estimated total interactions, Chacoff et al. 2012). Plant-frugivore
393 interactions have been scored from vertebrate regurgitates or gut/fecal contents (Poulin et al.
394 1999), but identification of propagules is based on morphological recognition, sowing or,
395 increasingly, molecular studies (Marrero et al. 2009, González-Varo et al. 2014). More intimate
396 interactions often require collecting hosts (animals or plants) and dissecting or rearing from
397 them (Novotny et al. 2010, Poulin 2010). It is also almost impossible to distinguish a rare but
398 genuine externally-feeding herbivore from the occasional non-feeding 'tourist', without
399 further confirmation from field observations or feeding trials (Ødegaard 2004). All these
400 problems typically reduce the volume of interaction data that can be obtained, so studies using
401 standardized techniques to measure interactions throughout several biomes are scarce (see
402 Schemske et al. 2009). Due to this, interaction data are often insufficient for the intended
403 analyses, or may even produce biased results.

404

3. A unified framework for understanding and managing biodiversity shortfalls

407 The seven shortfalls compromise the accuracy, generality and realism of biodiversity
408 knowledge. It is therefore crucial to understand their characteristics and their implications for
409 biodiversity research. In this section we outline the main interactions between the shortfalls,
410 how are they are affected by the spatial and temporal scale of analysis, and how they can
411 result in biased and/or inaccurate knowledge.

412

413 *SIDEBAR: BIAS, PRECISION, ACCURACY AND UNCERTAINTY*

414 *Uncertainty can be defined as either lack of exactness in measures or predictions (Hortal et al. 2008), or*
415 *simply a lack of knowledge or ignorance (Jackson 2012). Such uncertainty can be decomposed into three*
416 *components (reviewed in Walther and Moore 2005): i) bias, referring to pervasive errors in the*
417 *measurement and/or unevenly-distributed gaps in the data or models that lead to consistent differences*
418 *between true values and observed/predicted ones; ii) precision, defined as the variability in measures or*
419 *estimates, regardless of how close they are to the true value, and; iii) accuracy, referring to the distance*
420 *between measures or estimates and the true value, which can be affected by precision and bias.*

421

3.1. Interactions between shortfalls

423 Biodiversity shortfalls interact with each other in several ways (Figure 2a). The Linnean
424 shortfall necessarily affects all the others since it represents a lack of knowledge of the basic

425 units of study in ecology and evolution. Clearly, we have no empirical data on the
426 characteristics of unknown species, although, arguably, some of these can be estimated using
427 models fitted to ecological and evolutionary data about related species (Raxworthy et al.
428 2003), or attributed to Operational Taxonomic Units (OTUs) (e.g. Yahara et al. 2010). However,
429 the biases in our knowledge of living species caused by the Linnean shortfall means that many
430 phylogenies have significant numbers of missing taxa, which can affect the final tree topology
431 and makes problematic the estimates of rates of evolution (Nee et al. 1994, FitzJohn et al.
432 2009). Thus, any incremental increase in knowledge of any given aspect of biodiversity must be
433 preceded, or at least accompanied, by filling of the Linnean shortfall (Figure 2b).

434   The Wallacean, Prestonian and Darwinian shortfalls are inextricably related to each
435 other, whereby lack of data in one necessarily affects the other two (Figure 2a). This
436 interdependence is less strong between the Prestonian and the Darwinian shortfalls because
437 limited knowledge of extant populations only creates uncertainty about current population
438 (and metapopulation) dynamics and short-term evolutionary processes, hampers accurate
439 predictions of variability in population size (e.g. McCarthy et al. 1994) or introduces errors in
440 coalescence analyses through the use of invalid estimates of demographic history (Pybus et al.
441 2000, Drummond et al. 2005). In contrast, the Darwinian shortfall is strongly influenced by the
442 Eltonian, Hutchinsonian and Raunkiaeran shortfalls, since limited data on interactions,
443 ecophysiological responses and functional traits hampers describing co-evolutionary
444 processes, and niche and trait evolution (Diniz-Filho et al. 2013).

445   Knowledge about abiotic and biotic components of the niche and the functional traits
446 of each species are also tightly linked. For example, most of the easily measurable functional
447 traits are under selection by a combination of abiotic and biotic forces (Cornelissen et al.
448 2003), so that their predictive power for specific stressors is reduced and contingent to a given
449 region (Díaz et al. 2007). Similarly, the complex trade-offs between traits within and across
450 species remains largely unknown, making it difficult to use a few simple traits to consistently
451 predict a wide array of phenomena, such as fitness components, biotic interactions or impact
452 of species on ecosystems (Laughlin 2014). Moreover, the lack of knowledge about the effects
453 of biotic interactions and behavioral and life-history traits on realized physioclimatic responses
454 makes it difficult to estimate Grinnellian niches. This is evident in the high intraspecific
455 variability in the responses of pest species to climate through the history of their invasions
456 (Sutherst 2014).

457

458

459

460
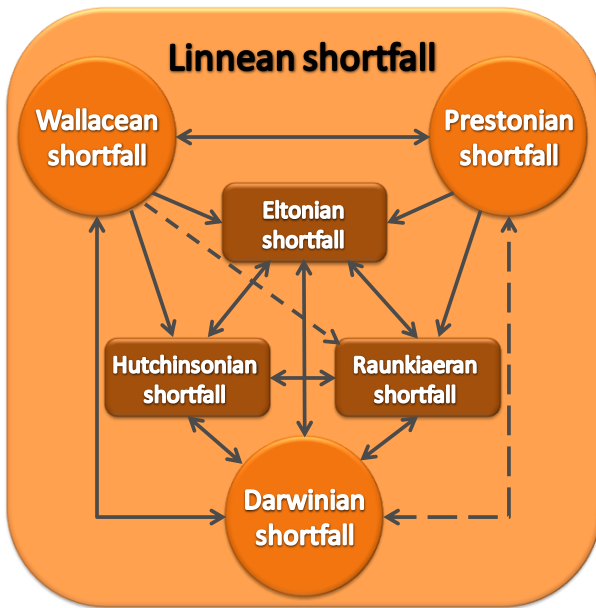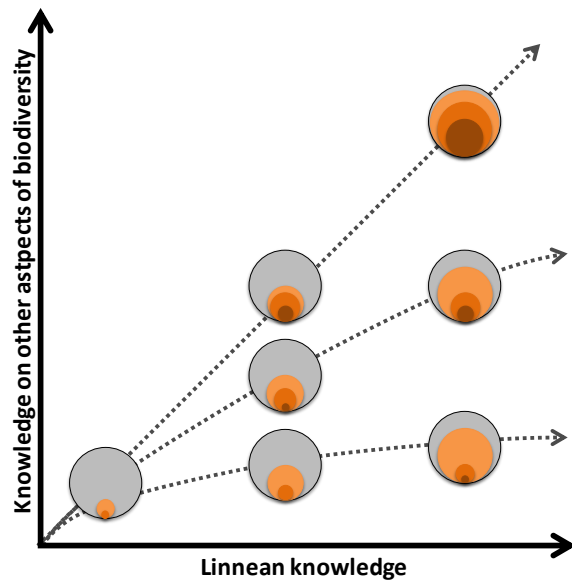
Figure 2. Relationships between shortfalls. Different shades of orange indicate different degrees of complexity in the knowledge about biodiversity, from an account of the basic entities (species, light shade) to knowledge about their extrinsic characteristics (i.e., their distribution in space and time; medium shade) and their intrinsic traits (i.e., ecological functionality as measured from species' niches and functional traits; dark shade). (a) Influence of the magnitude of each shortfall on the others; the Linnean shortfall affects all the others, and the direction of the influences between the rest is indicated by arrows (dashed arrows indicate non-critical effects). (b) Dependence of the shortfalls about other aspects of biodiversity on the filling in of the Linnean shortfall; the concentric circles represent the whole of biodiversity variations (grey), and the three kinds of shortfalls (shades of orange). The figure depicts three potential scenarios (indicated by dotted lines), where investment is devoted to either describing species diversity (below), documenting the distribution and variations in species diversity, or studying the ecological and evolutionary characteristics of all living entities (above).

473

474   The extrinsic shortfalls (Wallacean, Prestonian and Darwinian) have pervasive impacts
475   on the Raunkiaeran, Hutchinsonian and Eltonian shortfalls (Figure 2a). For example, the
476   Hutchinsonian shortfall is highly influenced by the Wallacean shortfall, which strongly biases
477   occurrence-derived niche data (Hortal et al. 2008). The lack of geographical coverage of trait
478   measurements (Figure 1b) can also affect the knowledge about within-species trait variations
479   and functional responses. The Raunkiaeran shortfall may be particularly sensitive to the lack of
480   detailed data on a representative subset of populations, since knowledge of within-population
481   trait variability is essential to determine functional effects and responses. Similarly, the
482   Wallacean shortfall compromises estimates of interactions occurring between any pair of
483   species due to the lack of geographical coverage of data (Poelen et al. 2014; Figure 1c). In fact,
484   the comparatively limited amount of high quality interaction data from the tropics makes it
485   difficult to determine whether latitudinal variations in interaction strength play a significant
486   role in the origin of diversity gradients (Schemske et al. 2009). Furthermore, well-resolved
487   phylogenies may be essential to determine whether the relationships between specific traits
488   and ecological functions are the result of other evolutionarily correlated (but unmeasured)
489   traits (Cadotte et al. 2012, Díaz et al. 2013, Thompson et al. 2015). Finally, knowledge of biotic
490   interactions is particularly sensitive to the taxonomic, spatial, temporal and ecological extent
491   of each data set. Quantitative interaction surveys are subject to the composite effect of the
492   problems in estimating the abundance within each interacting group and the interactions
493   between the groups themselves.

494

3.2. Scale-dependency of shortfalls

496   Biodiversity knowledge shortfalls are scale dependent in terms of resolution and the extent of
497   data coverage and analysis (Hartley & Kunin 2003, Whittaker et al. 2005). This can be most
498   easily illustrated with respect to the Linnean and Wallacean shortfalls (Figure 3). At the largest
499   possible grain size (the entire Earth) we necessarily have perfect knowledge of the distribution
500   of any species that has been described. However, at smaller grain sizes the Wallacean shortfall
501   begins to expand as increasingly precise information about distributions is required
502   (McPherson & Jetz 2007, Riddle et al. 2011, Pineda & Lobo 2012). Finally, at very small grain
503   sizes it becomes difficult to confirm the presence/absence of a species, especially for highly
504   mobile animals that range over wide areas and habitat types (e.g. Boyce 2006). This can also
505   be viewed in terms of species turnover at smaller grain sizes. The sensitivity of beta diversity to
506   sampling grain reflects the effect of local environmental heterogeneity: as sampling grain
507   increases, biotic assemblages appear more homogeneous (MacNally et al. 2004). There is also
508   a strong temporal signal at smaller spatial resolutions, with distributions fluctuating to
509   different degrees depending on the ecological characteristics of the species (Thomson et al.
510   2007).

511
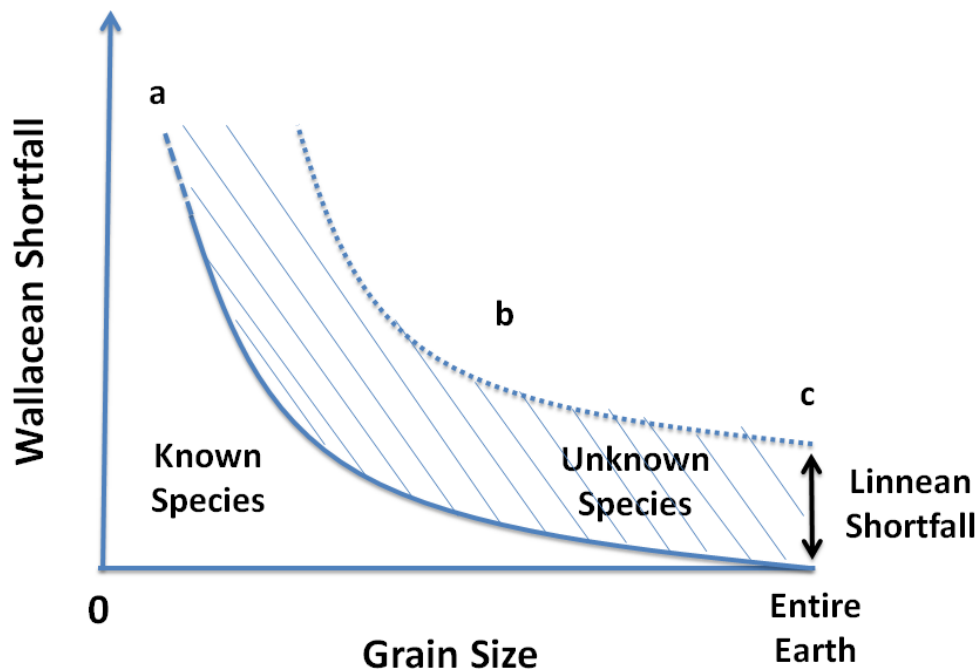
Figure 3. Relationship between grain size and the Wallacean shortfall. As grain size decreases the amount of knowledge needed to define the distribution of a species increases and the lack of such knowledge (the shortfall) correspondingly increases. When grain size is very small (a), there may be rapid shifts in presence/absence within sampling units putting practical and pragmatic limits on the shortfall. Depending on the characteristics of the species, the typical grain size for mapping species distributions on faunas and floras (i.e. distributional Atlases) is 1 to 100km^2 (b). At the grain size of the entire earth (c) we have perfect knowledge of the distribution of all known species and the Wallacean and Linnean shortfalls are equivalent.

Changes in biodiversity over time will strongly influence all shortfalls. The currently high rates of species and population extinctions are particularly important in this respect, since a reduction in global biodiversity truncates the shortfalls (see Costello et al. 2013) – partially alleviated by the much slower process of cladogenesis. The Raunkieran shortfall will also be influenced by anagenesis, which leads to shifting trait values over time. The process of adaptation and microevolution influences the Eltonian and Hutchinsonian shortfalls given the central role of natural selection in mediating abiotic and biotic interactions (McLachlan and Ladle 2011), combined with reshuffling of species co-occurrences over time. In summary, the size and nature of biodiversity shortfalls is highly dependent on the scale at which we collect, analyze and aggregate data.

3.3. Biases in Biodiversity Knowledge

Data on all aspects of biodiversity are typically taxonomically, geographically and temporally biased. Certain groups have received far more attention than others, to the point that only ~ 7% of fungi are estimated to have been discovered versus 70% of vascular plants (see Table 2 in Mora et al. 2011). Terrestrial vertebrates and vascular plants have been inventoried and described more exhaustively than almost all invertebrates or unicellular organisms. Moreover, within each taxon, data on larger, conspicuous and easily detectable species are recorded earlier and more extensively (Gaston and Blackburn 1994). Conversely, taxonomists tend to preferentially collect and record rare species, disregarding or under-representing common taxa (Garcillán & Ezcurra 2011). These biases also apply to other aspects of biodiversity. For example, there are more and better data on plant functional traits, due to the long botanical tradition of functional ecology (Díaz & Cabido 2001, Cornelissen et al. 2003), than for other organisms such as insects (Poff et al. 2006, Gibb & Parr 2013) or vertebrates (Villeger et al. 2010).

Species occurrence and assemblage data is also biased towards certain regions, habitats and environmental domains (Lobo et al. 2007; see Figure 2, Loiselle et al. 2008). This may be the most studied bias in biodiversity data, and involves a number of syndromes caused by the behavior of researchers (Sastre and Lobo 2009) and the historical patterns of colonization and inventorying (Bini et al. 2006, Stropp et al. 2015, Meyer et al. 2015). Thus, biodiversity inventories are typically more comprehensive near locations that offer convenient access, infrastructure and logistics (Dennis and Thomas 2000, Kadmon et al. 2004, Hortal et al. 2007). These geographical biases increase uncertainty about observed species distributions and the diversity of local assemblages (Beck and Kitching 2007, Boakes et al. 2010, Soria-Auza and Kessler 2008). Consequently, errors exist in the known distribution of endangered species and conservation targets, which can be several orders of magnitude larger in less studied groups. Furthermore, temporal shifts in the spatial coverage of surveys result in spurious changes in known distributions over time (Lobo et al. 2007, Hortal et al. 2008), affecting the ability to identify past range shifts (Huisman and Millar 2013) and discriminate current patterns of extinction (Lobo 2001).

Geographical variation in survey effort creates an uneven pattern of species descriptions through time (Baselga et al. 2010, Rich 2006), causing the Linnean shortfall to be spatially, temporally and taxonomically structured for most taxa (e.g. Diniz-Filho et al. 2005, Rodrigues et al. 2010). This structuring compromises reconstructions of the tree of life due to the uneven distribution of taxonomic information within extant lineages (Antonelli et al. 2014). The lack of data on particular lineages can, in turn, affect our knowledge on the evolution, diversity and interrelationships of traits (Figure 4).
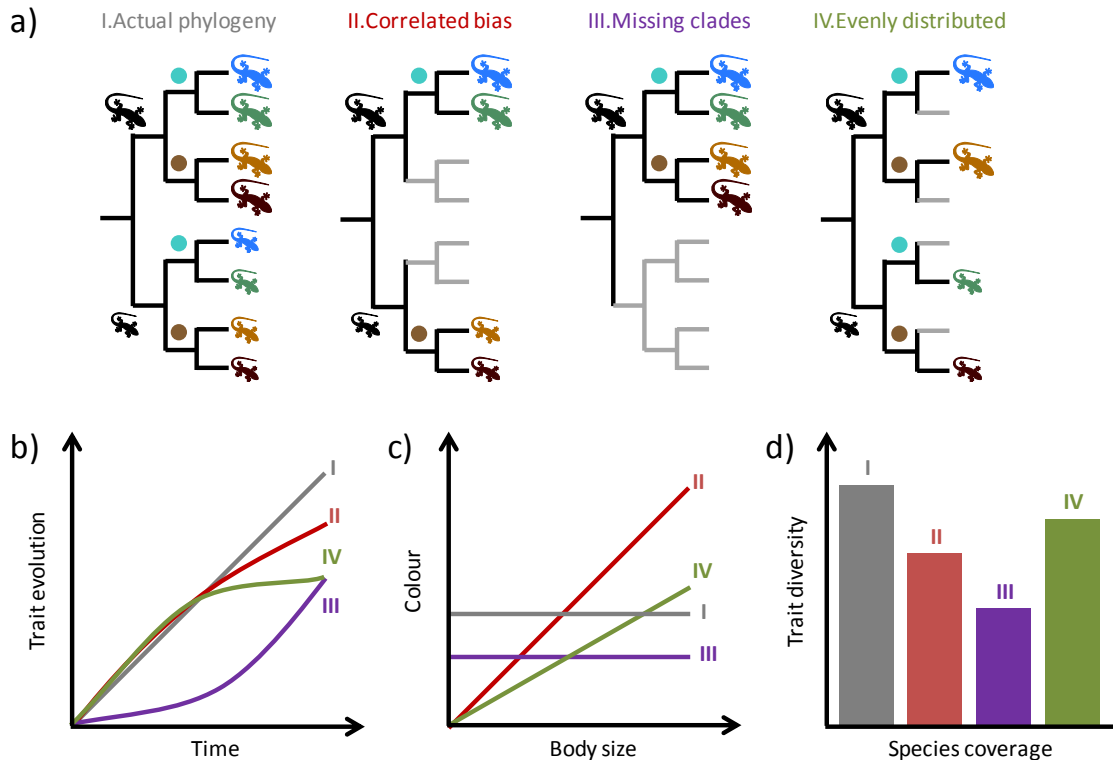
Figure 4. Effects of bias and completeness on the knowledge on trait evolution and trait diversity. (a) Different types of knowledge on the species diversity of a hypothetical lizard clade, and their impact on the perceived phylogeny and evolution of body size (large/small) and colour (from blue pigments to blue and green species, and from brown pigments to orange and dark brown species). In scenario (I) all species are known, while half of the species are missing in the other three scenarios, with different arrangements: (II) there is a phylogenetically correlated bias in the known species, consistent between both traits; (III) the whole clade of short-tailed species is unknown to science; and (IV) the missing species are evenly distributed through the phylogeny. (b) Perceived rates of trait evolution through time. (c) Perceived relationship between the two traits. (d) Perceived total trait diversity (i.e. number of different combinations of trait values).

The uneven spatial distribution of survey effort affects knowledge about the intrinsic characteristics of species. For example, bias in recording effort has a knock-on effect for occurrence-derived niche data for most species, particularly rare ones (Hortal et al. 2008). Known species interactions and trait values are also restricted to a few regions in the world which host large research institutions or field stations (Schemske et al. 2009). Hence, both Raunkiaeran and Eltonian shortfalls are more extensive outside of North America and Europe (see Figure 2). The same applies to marine systems, where the extensive and taxonomically comprehensive data on trophic interactions needed to develop well-resolved food webs are restricted to a limited number of estuaries (e.g. Raffaelli & Hall 1992) and fisheries (see Dunne et al. 2004). This limits our general understanding of the energy flows in the open ocean and the detection of global processes (Pauly and Palomares 2005).

4. Consequences and implications of knowledge shortfalls

Shortfalls in global biodiversity data have numerous implications for theory and practice (see Supplemental Table 1). The Linnean and Wallacean shortfalls have the farthest reaching influence, because data on the identity and distribution of species is vital for identifying broad-scale patterns in biodiversity and the processes that modify biodiversity (e.g. extinction). The Wallacean shortfall can also have profound impacts on estimates of conservation threat status. Range size is frequently used in conservation planning, with small ranges being given higher priority. Indeed, range restriction is an integral part of IUCN criteria to identify and classify species in danger of global extinction (IUCN 2011) with several conservation prioritization methods (e.g. Rodrigues et al. 2004) adopting an arbitrary criterion of < 50,000 km^2 to define range restriction/local endemism (Whittaker et al. 2005). Apart from the obvious problem that such a coarse-grained geographical category necessarily captures many species that are not under threat (e.g. many island endemics), the Wallacean shortfall means that prioritization for many taxa or for certain regions is highly uncertain (Riddle et al. 2011).

Bias in biodiversity data also critically influences the generality and realism of concepts or models. If the missing data are evenly distributed, the corresponding knowledge may lack precision, but it will be realistic and to some extent generalizable (see example IV in Figure 4). However, the more unevenly distributed the missing data, the less accurate the representation of the actual patterns or processes under study. This can lead to spurious hypotheses created to explain biased datasets. For example, Nelson demonstrated that supposed centers of plant endemism in Amazonia were partly an artefact of biased surveying for herbarium specimens (Nelson et al. 1990) – these patterns having been used to support the Pleistocene refuge theory for the origin of Amazonian plant diversity (Prance 1974).

Biodiversity data biases also strongly influence species distribution models (SDMs), a widely used analytical and predictive tool in conservation. SDMs typically relate field observations of species occurrences (and sometimes their absence) to environmental (usually climatic) predictors using statistically or theoretically derived response surfaces that are supposed to represent the tolerances of species to abiotic conditions. SDMs are routinely used for rare species, where accurate distributional data are typically missing. However, if the representation of the Grinnellian niche provided by occurrence data is biased (Hortal et al. 2008) then SDMs will consistently generate unrealistic results (Rocchini et al. 2011), particularly for rare species (Elith and Leathwick 2007). The effectiveness of SDMs is also hindered by the lack of knowledge about species interactions, since this limits their incorporation into the latest generation of models (e.g. Wisz et al. 2013). More generally, threat models could potentially be improved with a better understanding of how functional traits interact with environmental variables or more complete data on the abiotic tolerances of species. Further, there is mixed evidence that phylogeny can be used as a proxy of traits (Swenson & Enquist 2009, Pavoine et al. 2013). The lack of trait data and detailed phylogenies makes difficult to evaluate the extent to which phylogenetic information can be used as a proxy for trait variations (Díaz et al. 2013). In general, the creation of robust and practically useful models of the effects of threats on biodiversity and the impacts of its loss for ecosystem functioning remains a distant possibility.

5. Dealing with the shortfalls

The last decade has seen enormous advances in the collection and, especially, the collation and curation of biodiversity data at regional and global scales. The most recent of these 'mega-projects' is the 'Map of Life' (MoL), a web-based tool (www.mol.org) that aims to represent the distribution of every species on Earth (Jetz et al. 2012). At the time of writing the MoL is based on 196 data sets, 937,970 species and 371,807,519 records. If it is ever fully realized, this and similar initiatives such as the Catalogue of Life and the Encyclopedia of Life (www.eol.org) or the Paleobiology Database (paleobiodb.org) will provide an invaluable resource for the advancement of ecological research and conservation planning (Riddle et al. 2011). These global databasing projects are extremely important for diminishing the 'unknown knowns'; by cataloguing, organizing and making accessible the information they allow scientists to fully benefit from centuries of research on biodiversity. However, for maximum benefits the quality of the digitization process needs to be carefully vetted, ensuring that these data are accompanied by good meta-data providing an exhaustive account of the ancillary information associated to each record and measurement (Michener 2000, Hortal et al. 2007). Most importantly, for these data to constitute a reliable and generalizable source of information for theory and practice, three key challenges need to be met: i) the extent of the knowledge gaps and their patterns within the body of attainable knowledge must be described in a tractable way; ii) the major biodiversity data shortfalls (and their associated biases) need to be significantly reduced, and; iii) appropriate methods need to be developed that can deal with or account for the intrinsic limitations to the quality, longevity and coverage of biodiversity data (Ladle and Hortal 2013).

A necessary first step to deal with the missing information would be to circumscribe and inventory the ignorance that surrounds and is included within current knowledge; for example, through 'Atlases of ignorance' as proposed by Samuel W. Boggs (1949). In the era of big data, this may not be too problematic. Several works have described the extent of current knowledge, comparing it to estimates of global biodiversity (e.g. Mora et al. 2011, Costello et al. 2012). Despite controversies about the realism of the estimates (cf. Löbl and Leschen 2014), these assessments of the magnitude of the Linnean shortfall provide baselines that allow planning for the extent of the task. Equivalent estimates can be developed for other shortfalls, comparing for example estimated species geographic ranges with their occurrence records (Meyer et al. 2015), or evaluating the congruence of the estimated phylogenies with the available partial phylogenetic trees based on data (see Antonelli et al. 2014). Most importantly, these inventories allow the identification of biases in current data (see section 3.3).

New technologies and automated protocols will help diminish most shortfalls by increasing the effectiveness of the sampling effort devoted to each of them. Automatic analysis of phenotypes (Deans et al. 2015) and next-generation sequencing combined with bioinformatics (Taberlet et al. 2012) have the potential to identify species from large samples, documenting patterns of trait variation and linking genetic data with particular phenotypes. For some vertebrates, readily-available instruments allow the measurement of ecophysiological features of individuals during their activity in the field (e.g. Verdú et al. 2012). Other new technologies can potentially provide the data to fill in gaps in the Eltonian shortfall. Cheap, high resolution digital cameras are increasingly being deployed to survey elusive

organisms, particularly vertebrates (Rovero et al. 2014). They can also revolutionize the volume and standardization of recording pollinator or disperser visits to plants, during flowering (Celep et al. 2014) or fruiting (Prasad el al. 2010). DNA fingerprinting has not been widely used, but it promises a major breakthrough in large-scale recording of predator-prey or plant-herbivore interactions (Kress et al. 2015). For invertebrate herbivores in mass samples, gut contents may suffice to establish that they feed on a given plant without time-consuming observational or experimental confirmation (Pinzón-Navarro et al. 2010). Yet another powerful tool entails using stable isotopes for tracing trophic interactions or establishing trophic positions in food webs (Post 2002, Caut et al. 2009, Layman et al. 2012, Traugott et al. 2013).

Strategies to assess data quality and fill gaps in knowledge change dramatically if reducing bias becomes the main objective. Taxonomic misidentifications and recurrent errors in data gathering can result in meaningless results from complex analytical techniques (e.g. Lozier et al. 2009). Here, an initial evaluation of data quality and the biases themselves should be followed by a thoughtful redesign of surveys, measurements or models to offset them (Hortal et al. 2007). New data need to be strategically collected in ways that effectively reduce biases (Faith et al. 2013), even if this retards progress in filling the global shortfalls. Large databasing projects will be less useful if they do not diminish bias and increase representativeness of the available data. As in information technology, biodiversity science is particularly sensitive to the GIGO (garbage in, garbage out) problem, whereby biased input data (garbage in) will frequently produce undesired, often nonsensical, output (garbage out). Therefore, as stressed above, shortfalls need to be well mapped before they can be reduced. In the case of data on geographic distributions, the development of 'maps of ignorance' that provide information on where data are reliable or uncertain (Rocchini et al. 2011) can be used to inform the design and implementation of new surveys (Hortal and Lobo 2005). This kind of solution is potentially applicable to other shortfalls.

In the era of big data, modern biodiversity science needs to adopt strategies that recognize and embrace the unknown, incorporating the uncertainty produced by current gaps and biases in biodiversity data (Beck et al. 2012). A good starting point would be to routinely present uncertainty together with the data, results and/or model projections. This can be implemented with adequate visualizations and graphical representations in multimedia environments that allow simultaneous communicating of results and the associated uncertainty (McInerny et al. 2014). For example, maps of ignorance can be used to account for the uncertainty in the observed species richness patterns within a region, by blurring the colours of the map (scaled to represent the observed values) according to the level of uncertainty in each point. The development or deployment of visualization tools that represent uncertainty should be a priority for biodiversity informatics. Other options include restricting maps to well-sampled areas to avoid extrapolating beyond the limits of the known relationships and/or processes – although this would currently exclude most of the biologically richest and critically important regions of the world. Good practice in biodiversity science may entail clearly communicating how much a conclusion may be affected by each shortfall, and describing the limitations of any solutions adopted to deal with the associated uncertainty. In other words, how robust are the conclusions given the current level of uncertainty.

722   Quantitative descriptions of ignorance can also allow incorporating uncertainty
723   explicitly into modelling and decision making processes, for example by deploying fuzzy logic
724   (Petrou et al. 2014). Most regression techniques allow incorporating covariates to account for
725   measurement error, but an increasingly popular analytical strategy in ecological and
726   evolutionary research is the switch from significance testing to probabilistic approaches and
727   Bayesian analyses (Beck et al. 2012, Clark 2005). The utility of Bayesian statistics is currently
728   being put to the test in SDMs (Beale & Lennon 2012, F.Rodríguez-Sánchez et al. unpublished).
729   Here, maps of ignorance can be used as spatially explicit estimates of uncertainty for assessing
730   model sensitivity (Rocchini et al. 2011) or factoring such uncertainty directly into the models
731   through Bayesian techniques. Similar approaches could be used to account for the uncertainty
732   caused by other shortfalls.

733

734 6. Conclusions

735   The aim of this review was to describe the key remaining areas of ignorance about biodiversity,
736   the relationships between them, and their implications for research and conservation practice.
737   An important first step towards obtaining a better knowledge of biodiversity is recognizing
738   how much we do not know, and identifying critical gaps in knowledge. This explicit concern
739   with ignorance is not, in any sense, a criticism of the quality of the extensive work carried out
740   by past generations of naturalists and researchers. Rather, we are now able to recognize and
741   describe our ignorance precisely because their work has been progressively expanding and
742   defining the envelope of our knowledge. It is because we build on the preceding work of many
743   industrious researchers that we are now able to compile vast amounts of biodiversity data and
744   envisage what is yet to be discovered. Once we have sufficiently described biodiversity
745   knowledge and its biases and limitations, our task will be to improve the inventory of global
746   biodiversity in a way that maximizes coverage and which most effectively deploys the very
747   limited resources available for such work.

748   Whatever approach is adopted to fill in the gaps in knowledge – either targeting
749   strategically chosen biases or continuing the 'scattershot' approach of leaving survey decisions
750   to chance – the shortfalls are never likely to be filled in their entirety and they certainly will not
751   be filled in the near future. This poses a serious dilemma for conservationists and policy
752   makers, who cannot wait years or decades for data of sufficient quality to become available.
753   Of course, such dilemmas are not unique to conservation; it is commonly agreed that the
754   information revolution has made the world increasingly volatile, uncertain, complex, and
755   ambiguous. As a consequence the biggest challenges faced by modern societies are not
756   problems that can be 'solved', but take the form of dilemmas that must be managed (Johansen
757   2007). In this sense, we assert that, through careful analysis and inventive visualization of
758   ignorance, the uncertainty caused by biodiversity data shortfalls can be incorporated into
759   conservation planning and biodiversity research. This will bring awareness to the quality of our
760   current knowledge and, by extension, of the models developed with it. Even if this slows down
761   the pace of developing new analyses or conservation actions – which is not necessarily the
762   case – the conclusions produced by research where uncertainty is taken into account are likely
763   to be more robust and persistent, and likewise the long-term success of conservation actions

764 based upon them. To quote Daniel J. Boorstin, "The greatest enemy of knowledge is not
765 ignorance, it is the illusion of knowledge" (Boorstin 1983).

766

Summary points

768 1. Ecology, evolution and conservation science are entering the era of big data through the use
769 of massive databases on different aspects of biodiversity - it is time to assess the quality,
770 extent and representativeness of the available information.

771 2. We identify seven main shortfalls in biodiversity data. One represents the lack of
772 information on the total taxonomic extent of biodiversity (Linnean shortfall); three deal with
773 the most important extrinsic characteristics of the species – their geographic distribution,
774 population dynamics and evolutionary relationships (Wallacean, Prestonian and Darwinian
775 shortfalls, respectively); and three other shortfalls relate to ecological functioning – in terms of
776 functional traits, abiotic niche and biotic interactions (Raunkiaeran, Hutchinsonian and
777 Eltonian shortfalls, respectively).

778 3. All knowledge shortfalls are interconnected to varying degrees, according to scale and
779 spatial, temporal and taxonomic coverage. The Linnean shortfall critically affects all the others
780 because lack of information on unknown species necessarily prevents the description of any
781 other aspect of biodiversity. The lack of knowledge on extrinsic characteristics typically
782 aggravates shortfalls in knowledge of ecological functioning which, in turn, are tightly
783 entangled.

784 4. Global biodiversity data are generally incomplete and suffer from many biases. Most current
785 knowledge is about large, conspicuous, or economically valuable taxa and comes from
786 research in northern temperate regions and areas that are easy or convenient to access.

787 5. Bias generates uncertainty in all analyses of biodiversity, compromising the generality and
788 validity of theoretical knowledge on ecology and evolution, and the quality of conservation
789 assessments and actions.

790 6. Ignorance of biodiversity needs to be carefully described, quantified, and mapped. This will
791 identify knowledge gaps that require additional research effort, and allow the assessment of
792 uncertainty in estimates and model projections.

793

Future issues

795 1. Estimates of the extent and variability of all aspects of biodiversity will be used as a baseline
796 to compare with current knowledge stored in large databases, identifying gaps of knowledge
797 and directing research and exploratory works.

798 2. Maps of biogeographical ignorance based on the coverage of spatial and environmental
799 gradients through time provided by distributional data –as gathered in GBIF or Map of Life–
800 will be used to design surveys, assess model uncertainty and develop Bayesian Species
801 Distribution Models.

802 3. Technological advances (e.g. DNA bar-coding, automated species identification, Artificial
803 Inteligence-assisted remote sensing) will revolutionize the collection of information for all
804 domains of biodiversity knowledge, massively accelerating the rate of data capture.

805 4. Ongoing compilations of phylogenetic information such as the SUPERSMART project will be
806 used to identify gaps in the known tree of life, targeting clades in need of further study.

807 5. Assessments of the bias and representativeness of the information contained in long-term
808 data on population dynamics, functional traits –including ecophysiological information– and
809 biotic interactions stored on biodiversity databases will be developed to identify gaps in
810 knowledge and to better target taxa, biomes and ecosystems.

811 6. New ways of representing and communicating uncertainty will be developed to raise
812 awareness of the certainties behind the uncertainty, and the extent of current ignorance on
813 biodiversity pattern and process.

814

831

Literature cited

Adler PB, Salguero-Gómez R, Compagnoni A, Hsu JS, Ray-Mukherjee J, et al. 2014. Functional traits explain variation in plant life history strategies. *Proceedings of the National Academy of Sciences USA* 111: 740-45

Albert CH, Thuiller W, Yoccoz NG, Douzet R, Aubert S, Lavorel S. 2010. A multi-trait approach reveals the structure and the relative importance of intra-vs. interspecific variability in plant traits. *Functional Ecology* 24: 1192-201

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences USA* 106: 13410-14

Antonelli A, Condamine FL, Hettling H, Nilsson K, Nilsson RH et al. 2014. SUPERSMART: ecology and evolution in the era of big data. *PeerJ PrePrints* 2: e501v1

Baselga A, Lobo JM, Hortal J, Jiménez-Valverde A, Gómez JF. 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

Beale CM, Lennon JJ. 2012. Incorporating uncertainty in predictive species distribution modelling. *Philosophical Transactions of the Royal Society B* 367: 247-58

Beck J, Ballesteros-Mejia L, Buchmann CM, Dengler J, Fritz SA et al. 2012. What's on the horizon for macroecology? *Ecography* 35: 673-83

Beck JAN, Kitching IJ. 2007. Estimating regional species richness of tropical insects from museum data: a comparison of a geography-based and sample-based methods. *Journal of Applied Ecology* 44: 672-81

Bini LM, Diniz-Filho JAF, Rangel TFLVB, Bastos RP, Pinto MP. 2006. Challenging Wallacean and Linnean shortfalls: Knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions* 12: 475-82.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, et al. 2007. The delayed rise of present-day mammals. *Nature* 446: 507-12

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution* 57: 717-45

Blüthgen N, Fründ J, Vázquez DP, Menzel F. 2008. What do interaction network metrics tell us about specialization and biological traits. *Ecology* 89: 3387-99

Boakes EH, McGowan PJK, Fuller RA, Chang-qing D, Clark NE et al. 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biol* 8: e1000385

Boggs SW. 1949. An atlas of ignorance: A needed stimulus to honest thinking and hard work. *Proceedings of the American Philosophical Society* 93: 253-58

Boorstin DJ. 1983. *The Discoverers: A History of Man's Search to Know His World and Himself*. Los Angeles, CA: The Publishing Mills.

Boyce MS. 2006. Scale for resource selection functions. *Diversity and Distributions* 12: 269-76

Bozinovic F, Calosi P, Spicer JI. 2011. Physiological correlates of geographic range in animals. *Annual Review of Ecology, Evolution, and Systematics* 42: 155-79

Brown JH. 1984. On the relationship between abundance and distribution of species. *American Naturalist* 124: 255-79

Brown JH, Lomolino MV. 1998. *Biogeography. Second edition*. Sunderland, Massachussets: Sinauer Associates, Inc. 691 pp.

Bush MB, Lovejoy TE. 2007. Amazonian conservation: pushing the limits of biogeographical knowledge. *Journal of Biogeography* 34: 1291-93

Cadotte M, Albert CH, Walker SC. 2013. The ecology of differences: Assessing community assembly with trait and evolutionary distances. *Ecology Letters* 16: 1234-44

882   Cadotte MW, Dinnage R, Tilman D. 2012. Phylogenetic diversity promotes ecosystem stability.
883   *Ecology* 93: S223-S33
884   Caley MJ, Fisher R, Mengersen K. 2014. Global species richness estimates have not converged.
885   *Trends in Ecology & Evolution* 29: 187-88
886   Cardoso P, Erwin TL, Borges PA, New TR. 2011. The seven impediments in invertebrate
887   conservation and how to overcome them. *Biological Conservation* 144: 2647-55
888   Caut S, Angulo E, Courchamp F. 2009. Variation in discrimination factors (Δ15N and Δ13C): the
889   effect of diet isotopic values and applications for diet reconstruction. *Journal of
890   Applied Ecology* 46: 443-53
891   Celep F, Atalay Z, Dikmen F, Do an M, Classen-Bockhoff R. 2014. Flies as pollinators of
892   melittophilous Salvia species (Lamiaceae). *American Journal of Botany* 101: 2148-59
893   Chacoff NP, Vázquez D, Lomáscolo S, Stevani E, Dorado J, Padrón B. 2012. Evaluating sampling
894   completeness in a desert plant-pollinator network. *Journal of Animal Ecology* 81: 190-
895   200
896   Chapman AD. 2009. *Numbers of living species in Australia and the World*. 2nd edition.
897   Australian Biological Resources Study, Canberra
898   Clark JS, Bjørnstad ON. 2004. Population time series: Process variability, observation errors,
899   missing values, lags, and hidden states. *Ecology* 85: 3140-50
900   Clark JS. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters* 8: 2-14
901   Colwell RK, Rangel TF. 2009. Hutchinson's duality: The once and future niche. *Proceedings of
902   the National Academy of Sciences USA* 106: 19651–58
903   Cordlandwehr V, Meredith RL, Ozinga WA, Bekker RM, van Groenendael JM, Bakker JP. 2013.
904   Do plant traits retrieved from a database accurately predict on-site measurements?
905   *Journal of Ecology* 101: 662-70
906   Cornelissen JHC, Lavorel S, Garnier E, Diaz S, Buchmann N et al. 2003. A handbook of protocols
907   for standardised and easy measurement of plant functional traits worldwide.
908   *Australian Journal of Botany* 51: 335-80
909   Costello MJ, Wilson S, Houlding B. 2012. Predicting Total Global Species Richness Using Rates
910   of Species Description and Estimates of Taxonomic Effort. *Systematic Biology* 61: 871-
911   83
912   Costello MJ, Wilson S, Houlding B. 2012. Predicting total global species richness using rates of
913   species description and estimates of taxonomic effort. *Systematic Biology* 61: 871-83
914   Costello MJ, May RM, Stork NE. 2013. Can we name Earth's species before they go extinct?
915   *Science* 339: 413-16
916   Dayrat B. 2005. Towards integrative taxonomy. *Biological Journal of the Linnean Society* 85:
917   407-15
918   de Bello F, Lavorel S, Díaz S, Harrington R, Cornelissen JC et al. 2010. Towards an assessment of
919   multiple ecosystem processes and services via functional traits. *Biodiversity and
920   Conservation* 19: 2873-93
921   Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M et al. 2015. Finding Our Way through
922   Phenotypes. *PLoS Biol* 13: e1002033
923   Dennis R, Thomas C. 2000. Bias in butterfly distribution maps: the influence of hot spots and
924   recorder's home range. *Journal of Insect Conservation* 4: 73-77
925   Diamond SE, Nichols LM, McCoy N, Hirsch C, Pelini SL, et al. 2012. A physiological trait-based
926   approach to predicting the responses of species to experimental climate warming.
927   *Ecology* 93: 2305-12
928   Diaz S, Cabido M. 2001. Vive la différence: plant functional diversity matters to ecosystem
929   processes. *Trends in Ecology and Evolution* 16: 646-55
930   Díaz S, Lavorel S, de Bello F, Quétier F, Grigulis K, Robson TM. 2007. Incorporating plant
931   functional diversity effects in ecosystem service assessments. *Proceedings of the
932   National Academy of Sciences* 104: 20684-89

933 Díaz S, Purvis A, Cornelissen JHC, Mace GM, Donoghue MJ et al. 2013. Functional traits, the
934   phylogeny of function, and ecosystem service vulnerability. *Ecology and Evolution* 3:
935   2958-75
936 Diniz-Filho JAF, Bastos RP, Rangel TFLVB, Bini LM, Carvalho P, Silva RJ. 2005. Macroecological
937   correlates and spatial patterns of anuran description dates in the Brazilian Cerrado.
938   *Global Ecology and Biogeography* 14: 469-77
939 Diniz-Filho JAF, Bini LM, Hawkins BA. 2003. Spatial autocorrelation and red herrings in
940   geographical ecology. *Global Ecology and Biogeography* 12: 53-64
941 Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM. 2013. Darwinian shortfalls in
942   biodiversity conservation. *Trends in Ecology & Evolution* 28: 689-95
943 Dornburg A, Beaulieu JM, Oliver JC, Near TJ. 2011. Integrating fossil preservation biases in the
944   selection of calibrations for molecular divergence time estimation. *Systematic Biology*
945   60: 519-27
946 Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past
947   population dynamics from molecular sequences. *Molecular Biology and Evolution* 22:
948   1185-92
949 Dunne J, Williams R, Martinez N. 2004. Network structure and robustness of marine food
950   webs. *Marine Ecology Progress Series* 273: 291-302
951 Dunning JB, Stewart DJ, Danielson BJ, Noon BR, Root TL, et al. 1995. Spatially explicit
952   population models: current forms and future uses. *Ecological Applications* 5: 3-11
953 Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ. 2011. A novel comparative method for
954   identifying shifts in the rate of character evolution on trees. *Evolution* 65: 3578-89
955 Elith J, Leathwick J. 2007. Predicting species distributions from museum and herbarium records
956   using multiresponse models fitted with multivariate adaptive regression splines.
957   *Diversity and Distributions* 13: 265-75
958 Ellwood MDF, Foster WA. 2004. Doubling the estimate of invertebrate biomass in a rainforest
959   canopy. *Nature* 429: 549-51
960 Faith D, Collen B, Ariño A, Patricia Koleff PK, Guinotte J, et al. 2013. Bridging the biodiversity
961   data gaps: Recommendations to meet users' data needs. *Biodiversity Informatics* 8:
962   4126
963 Felsenstein J. 1985. Phylogenies and the comparative method. *American Naturalist* 125: 1-15
964 Firestein S. 2012. *Ignorance. How it drives science*. Oxford: Oxford University Press. 195 pp.
965 FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and
966   extinction rates from incompletely resolved phylogenies. *Systematic Biology* 58: 595-
967   611
968 Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: A test
969   and review of evidence. *American Naturalist* 160: 712-26
970 Garcillán PP, Ezcurra E. 2011. Sampling procedures and species estimation: testing the
971   effectiveness of herbarium data against vegetation sampling in an oceanic island.
972   *Journal of Vegetation Science* 22: 273-80.
973 Gaston KJ. 2003. *The structure and dynamics of geographic ranges*. Oxford: Oxford University
974   Press
975 Gaston KJ, Blackburn TM. 1994. Are newly described bird species small-bodied? *Biodiversity*
976   *Letters* 2: 16–20
977 Gaston KJ, Rodrigues ASL. 2003. Reserve selection in regions with poor biological data.
978   *Conservation Biology* 17: 188-95
979 Gibb H, Parr CL. 2013. Does structural complexity determine the morphology of assemblages?
980   An experimental test on three continents. *PLoS ONE* 8: e64005
981 Gleason HA. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey*
982   *Botanical Club* 53: 7-26

983 González-Varo JP, Arroyo JM, Jordano P. 2014. Who dispersed the seeds? The use of DNA
984   barcoding in frugivory and seed dispersal studies. *Methods in Ecology and Evolution* 5:
985   806-14
986 Gouveia SF, Hortal J, Tejedo M, Duarte H, Cassemiro FAS et al. 2014. Climatic niche at
987   physiological and macroecological scales: the thermal tolerance–geographical range
988   interface and niche dimensionality. *Global Ecology and Biogeography* 23: 446-56
989 Hall BG. 2011. *Phylogenetic trees made easy: A how-to manual. Fourth edition.* Sunderland,
990   MA: Sinauer Associates. 282 pp.
991 Hall SJ, Raffaelli D. 1991. Food-web patterns: lessons from a species-rich web. *Journal of
992   Animal Ecology* 60: 823-41
993 Hansen TF, Martins EP. 1996. Translating between microevolutionary process and
994   macroevolutionary patterns: The correlation structure of interspecific cata. *Evolution*
995   50: 1404-17
996 Hartley S, Kunin WE. 2003. Scale dependency of rarity, extinction risk, and conservation
997   priority. *Conservation Biology* 17: 1559-70
998 Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Systematic
999   Biology* 54: 852-59
1000 Hopkins GW, Freckleton RP. 2002. Declines in the numbers of amateur and professional
1001   taxonomists: implications for conservation. *Animal Conservation* 5: 245-49
1002 Hopkins MJG. 2007. Modelling the known and unknown plant biodiversity of the Amazon
1003   Basin. *Journal of Biogeography* 34: 1400-11
1004 Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A. 2008. Historical bias in
1005   biodiversity inventories affects the observed environmental niche of the species. *Oikos*
1006   117: 847-58
1007 Hortal J, Lobo JM, Jimenez-Valverde A. 2007. Limitations of Biodiversity Databases: Case Study
1008   on Seed-Plant Diversity in Tenerife, Canary Islands. *Conservation Biology* 21: 853-63
1009 Hortal J, Lobo JM. 2005. An ED-based protocol for optimal sampling of biodiversity. *Biodiversity
1010   & Conservation* 14: 2913-47
1011 Huisman JM, Millar AJK. 2013. Australian seaweed collections: use and misuse. *Phycologia* 52:
1012   2-5
1013 Hutchinson GE. 1978. *An introduction to population biology*. New Haven, CT: Yale University
1014   Press
1015 Inchausti P, Halley J. 2001. Investigating long-term ecological variability using the global
1016   population dynamics database. *Science* 293: 655-57
1017 Ings TC, Montoya JM, Bascompte J, Blüthgen N, Brown L, et al. 2009. Ecological networks -
1018   beyond food webs. *Journal of Animal Ecology* 78: 253-69
1019 IUCN. 2011. *IUCN Red List of threatened species*. Gland: IUCN
1020 Jackson ST. 2012. Representation of flora and vegetation in Quaternary fossil assemblages:
1021   known and unknown knowns and unknowns. *Quaternary Science Reviews* 49: 1-15
1022 Jetz W, McPherson JM, Guralnick RP. 2012. Integrating biodiversity distribution knowledge:
1023   toward a global map of life. *Trends in ecology & evolution* 27: 151-59
1024 Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in
1025   space and time. *Nature* 491: 444-48
1026 Johansen B. 2007. *Get there early*. Oakland: Berrett-Koehler
1027 Kadmon R, Farber O, Danin A. 2004. Effect of roadside bias on the accuracy of predictive maps
1028   produced by bioclimatic models. *Ecological Applications* 14: 401-13
1029 Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P et al. 2011. TRY – a global database of plant
1030   traits. *Global Change Biology* 17: 2905-35
1031 Kattge J, Ogle K, Bönisch G, Díaz S, Lavorel S, et al. 2011. A generic structure for plant trait
1032   databases. *Methods in Ecology and Evolution* 2: 202-13
1033 Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, et al. 2001. The strength of
1034   phenotypic selection in natural populations. *American Naturalist* 157: 245-61

1035 Knape J, de Valpine P. 2012. Are patterns of density dependence in the Global Population
1036   Dynamics Database driven by uncertainty about population abundance? *Ecology*
1037   *Letters* 15: 17-23
1038 Kress WJ, García-Robledo C, Uriarte M, Erickson DL. 2015. DNA barcodes for ecology,
1039   evolution, and conservation. *Trends in Ecology & Evolution* 30: 25-35
1040 Ladle R, Hortal J. 2013. Mapping species distributions: living with uncertainty. *Frontiers of*
1041   *Biogeography* 5: 8-9
1042 Lambers H, Chapin III FS, Pons TL. 2008. *Plant Physiological Ecology. Second edition.* New York:
1043   Springer-Verlag. 605 pp.
1044 Laughlin DC, Laughlin DE. 2013. Advances in modeling trait-based plant community assembly.
1045   *Trends in Plant Science* 18: 584-93
1046 Laughlin DC. 2014. The intrinsic dimensionality of plant traits and its relevance to community
1047   assembly. *Journal of Ecology* 102: 186-93
1048 Layman CA, Araujo MS, Boucek R, Hammerschlag-Peyer CM, Harrison E, et al. 2012. Applying
1049   stable isotopes to examine food-web structure: an overview of analytical tools.
1050   *Biological Reviews* 87: 545-62
1051 Lewinsohn TM, Inácio Prado P, Jordano P, Bascompte J, M Olesen J. 2006. Structure in plant–
1052   animal interaction assemblages. *Oikos* 113: 174-84
1053 Löbl I, Leschen RAB. 2014. Misinterpreting global species numbers: examples from Coleoptera.
1054   *Systematic Entomology* 39: 2-6
1055 Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF. 2007. How does the knowledge
1056   about the spatial distribution of Iberian dung beetle species accumulate over time?
1057   *Diversity and Distributions* 13: 772-80
1058 Lobo JM. 2001. Decline of roller dung beetle (Scarabaeinae) populations in the Iberian
1059   peninsula during the 20th century. *Biological Conservation* 97: 43-50
1060 Logan ML, Cox RM, Calsbeek R. 2014. Natural selection on thermal performance in a novel
1061   thermal environment. *Proceedings of the National Academy of Sciences* 111: 14165-69
1062 Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG et al. 2008. Predicting species
1063   distributions from herbarium collections: does climate bias in collection sampling
1064   influence model outcomes? *Journal of Biogeography* 35: 105-16
1065 Lomolino MV. 2004. Conservation Biogeography. In *Frontiers of Biogeography: new directions*
1066   *in the geography of nature*, ed. MV Lomolino, LR Heaney, pp. 293-96. Sunderland,
1067   Massachussets: Sinauer Associates, Inc.
1068 Losos JB. 2011. Seeing the forest for the trees: The limitations of phylogenies in comparative
1069   biology. *American Naturalist* 177: 709-27
1070 Lozier JD, Aniello P, Hickerson MJ. 2009. Predicting the distribution of Sasquatch in western
1071   North America: anything goes with ecological niche modelling. *Journal of*
1072   *Biogeography* 36: 1623-27
1073 Lukoschek V, Scott Keogh J, Avise JC. 2012. Evaluating fossil calibrations for dating phylogenies
1074   in light of rates of molecular evolution: A comparison of three approaches. *Systematic*
1075   *Biology* 61: 22-43
1076 MacNally R, Fleishman E, Bulluck LP, Betrus CJ. 2004. Comparative influence of spatial scale on
1077   beta diversity within regional assemblages of birds and butterflies. *Journal of*
1078   *Biogeography* 31: 917-29
1079 Mace GM, Gittleman JL, Purvis A. 2003. Preserving the Tree of Life. *Science* 300: 1707-09
1080 Mace GM. 2004. The role of taxonomy in species conservation. *Philosophical Transactions of*
1081   *the Royal Society of London B: Biological Sciences* 359: 711-19
1082 Marrero P, Fregel R, Cabrera VM, Nogales M. 2009. Extraction of high-quality host DNA from
1083   feces and regurgitated seeds: A useful tool for vertebrate ecological studies. *Biological*
1084   *Research* 42: 147-51
1085 Martinez ND. 1993. Effects of resolution on food web structure. *Oikos* 66: 403-12
1086 May RM. 2010. Tropical arthropod species, more or less? *Science* 329: 41-42

1087 McCann S, Greenlees MJ, Newell D, Shine R. 2014. Rapid acclimation to cold allows the cane
1088   toad to invade montane areas within its Australian range. *Functional Ecology* 28: 1166-
1089   74

1090 McCarthy MA, Franklin DC, Burgman MA. 1994. The importance of demographic uncertainty:
1091   An example from the helmeted honeyeater Lichenostomus melanops cassidix.
1092   *Biological Conservation* 67: 135-42

1093 McGill BJ, Enquist BJ, Weiher E, Westoby M. 2006. Rebuilding community ecology from
1094   functional traits. *Trends in Ecology & Evolution* 21: 178-85

1095 McInerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M et al. 2014. Information visualisation
1096   for science and policy: engaging users and avoiding bias. *Trends in Ecology & Evolution*
1097   29: 148-57

1098 McInerny GJ, Purves DW. 2011. Fine-scale environmental variation in species distribution
1099   modelling: regression dilution, latent variables and neighbourly advice. *Methods in*
1100   *Ecology and Evolution* 2: 248-57

1101 McLachlan AJ, Ladle RJ. 2011. Barriers to adaptive reasoning in community ecology. *Biological*
1102   *Reviews* 86: 543-48

1103 McPherson JM, Jetz W. 2007. Type and spatial structure of distribution data and the perceived
1104   determinants of geographical gradients in ecology: the species richness of African
1105   birds. *Global Ecology and Biogeography* 16: 657-67

1106 Meyer C, Kreft H, Guralnick R, Jetz W. 2015. Global priorities for an effective information basis
1107   of biodiversity distributions. *PeerJ PrePrints* 3: e1057

1108 Michener WK. 2000. Metadata. In *Ecological Data: Design, Management and Processing*, ed.
1109   WK Michener, J Brunt, pp. 92-116. London: Blackwell Science

1110 Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. 2011. How many species are there on Earth
1111   and in the ocean? *PLoS biology* 9: e1001127

1112 Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters* 17: 508-
1113   25

1114 Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philosophical*
1115   *Transactions of the Royal Society of London B* 344: 305-11

1116 Nee S, May RM. 1997. Extinction and the loss of evolutionary history. *Science* 278: 692-94

1117 Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML. 1990. Endemism centres, refugia and
1118   botanical collection density in Brazilian Amazonia.

1119 Novotny V, Miller SE, Baje L, Balagawi S, Basset Y, et al. 2010. Guild-specific patterns of species
1120   richness and host specialization in plant-herbivore food webs from a tropical forest.
1121   *Journal of Animal Ecology* 79: 1193-203

1122 Ødegaard F. 2004. Species richness of phytophagous beetles in the tropical tree *Brosimum*
1123   *utile* (Moraceae): The effects of sampling strategy and the problem of tourists.
1124   *Ecological Entomology* 29: 76-88

1125 Pauly D, Palomares M-L. 2005. Fishing down marine food web: It is far more pervasive than we
1126   thought. *Bulletin of Marine Science* 76: 197-212

1127 Pavoine S, Gasc A, Bonsall MB, Mason NWH. 2013. Correlations between phylogenetic and
1128   functional diversity: mathematical artefacts or true ecological and evolutionary
1129   processes? *Journal of Vegetation Science* 24: 781-93

1130 Pelayo-Villamil P, Guisande C, Vari RP, Manjarrés-Hernández A, García-Roselló E et al. 2014.
1131   Global diversity patterns of freshwater fishes – potential victims of their own success.
1132   *Diversity and Distributions* in press, doi: 10.1111/ddi.12271

1133 Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, et al. 2011. *Ecological*
1134   *niches and geographic distributions*. Princeton: Princeton University Press. 328 pp.

1135 Petrou ZI, Kosmidou V, Manakos I, Stathaki T, Adamo M et al. 2014. A rule-based classification
1136   methodology to handle uncertainty in habitat mapping employing evidential reasoning
1137   and fuzzy logic. *Pattern Recognition Letters* 48: 24-33

Pimm SL, Lawton JH, Cohen JE. 1991. Food web patterns and their consequences. *Nature* 350: 669-74

Pineda E, Lobo JM. 2012. The performance of range maps and species distribution models representing the geographic variation of species richness at different resolutions. *Global Ecology and Biogeography* 21: 935-44

Pinzón-Navarro S, Jurado-Rivera JA, Gomez-Zurita J, Lyal CHC, Vogler AP. 2010. DNA profiling of host–herbivore interactions in tropical forests. *Systematic Entomology* 35: 18-32

Poelen JH, Simons JD, Mungall CJ. 2014. Global biotic interactions: An open infrastructure to share and analyze species-interaction datasets. *Ecological Informatics* 24: 148-59

Poff NL, Olden JD, Vieira NKM, Finn DS, Simmons MP, Kondratieff BC. 2006. Functional trait niches of North American lotic insects: Traits-based ecological applications in light of phylogenetic relationships. *Journal of the North American Benthological Society* 25: 730-55

Poorter L, Wright SJ, Paz H, Ackerly D, Condit R et al. 2008. Are functional traits good predictors of demographic rates? Evidence from five neotropical forests. *Ecology* 89: 1908-20

Post DM. 2002. Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology* 83: 703-18

Poulin B, Wright SJ, Lefebvre G, Calderon O. 1999. Interspecific synchrony and asynchrony in the fruiting phenologies of congeneric bird-dispersed plants in Panama. *Journal of Tropical Ecology* 15: 213-27

Poulin R. 2010. Network analysis shining light on parasite ecology and diversity. *Trends in Parasitology* 26: 492-98

Prance GT. 1974. Phytogeographic support for the theory of Pleistocene forest refuges in the Amazon basin, based on evidence from distribution patterns in Caryocaraceae, Chrysobalanaceae, Dichapetalaceae and Lecythidaceae. *Acta Amazonica* 3: 5-28

Prasad S, Pittet A, Sukumar R. 2009. Who really ate the fruit? A novel approach to camera trapping for quantifying frugivory by ruminants. *Ecological Research* 25: 225-31

Pybus OG, Rambaut A, Harvey PH. 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155: 1429-37

Pyron RA, Wiens JJ. 2011. A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution* 61: 543-83

Raffaelli D, Hall SJ. 1992. Compartments and predation in an estuarine food web. *Journal of Animal Ecology* 61: 551-60

Raxworthy CJ, Martínez-Meyer E, Horning N, Nussbaum RA, Schneider GE et al. 2003. Predicting distributions of known and unknown reptile species in Madagascar. *Nature* 426: 837-41

Reed DH, O'Grady JJ, Brook BW, Ballou JD, Frankham R. 2003. Estimates of minimum viable population sizes for vertebrates and factors influencing those estimates. *Biological Conservation* 113: 23-34

Rex MA, Etter RJ. 2010. *Deep-sea biodiversity: pattern and scale*. Cambridge: Harvard University Press. 368 pp.

Rich TCG. 2006. Floristic changes in vascular plants in the British Isles: geographical and temporal variation in botanical activity 1836-1988. *Botanical Journal of the Linnean Society* 152: 303-30

Ricklefs RE. 2012. Species richness and morphological diversity of passerine birds. *Proceedings of the National Academy of Sciences USA* 109: 14482-87

Riddle BR, Ladle RJ, Lourie SA, Whittaker RJ. 2011. Basic biogeography: estimating biodiversity and mapping nature. In *Conservation Biogeography*, ed. RJ Ladle, RJ Whittaker, pp. 45-92. Oxford: Wiley

1189 Rocchini D, Hortal J, Lengyel S, Lobo JM, Jimenez-Valverde A et al. 2011. Accounting for
1190   uncertainty when mapping species distributions: the need for maps of ignorance.
1191   *Progress in Physical Geography* 35: 211-26
1192 Rodrigues AS, Andelman SJ, Bakarr MI, Boitani L, Brooks TM et al. 2004. Effectiveness of the
1193   global protected area network in representing species diversity. *Nature* 428: 640-43
1194 Rodrigues ASL, Gray CL, Crowter BJ, Ewers RM, Stuart SN et al. 2010. A global assessment of
1195   amphibian taxonomic effort and expertise. *Bioscience* 60: 798-806
1196 Rosen R. 1996. On the limitations of scientific knowledge. In *Boundaries and barriers: On the*
1197   *limits to scientific knowledge*, ed. JL Casti, A Karlqvist, pp. 199-214. Reading, MA:
1198   Perseus Books
1199 Roskov Y, Kunze T, Orrell T, Abucay L, Paglinawan L, et al, eds. 2014. *Species 2000 & ITIS*
1200   *Catalogue of Life, 2014 Annual Checklist*. Digital resource at
1201   www.catalogueoflife.org/annual-checklist/2014. Leiden: Species 2000: Naturalis
1202 Rovero F, Martin E, Rosa M, Ahumada JA, Spitale D. 2014. Estimating species richness and
1203   modelling habitat preferences of tropical forest mammals from camera trap data. *PloS*
1204   *one* 9: e103300
1205 Roy K, Foote M. 1997. Morphological approaches to measuring biodiversity. *Trends in Ecology*
1206   *& Evolution* 12: 277-81
1207 Sánchez-Fernández D, Aragón P, Bilton DT, Lobo JM. 2012. Assesing the congruence of thermal
1208   niche estimations derived from distribution and physiological data. A test using diving
1209   beetles. *PLoS ONE* 7: e48163
1210 Sansom RS, Randle E, Donoghue PC. 2015. Discriminating signal from noise in the fossil record
1211   of early vertebrates reveals cryptic evolutionary history. *Proceedings of the Royal*
1212   *Society of London B: Biological Sciences* 282: 20142245
1213 Sastre P, Lobo JM. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns.
1214   *Biological Conservation* 142: 462-67.
1215 Schemske DW, Mittelbach GG, Cornell HV, Sobel JM, Roy K. 2009. Is there a latitudinal gradient
1216   in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and*
1217   *Systematics* 40: 245-69
1218 Sechrest W, Brooks TM, da Fonseca GAB, Konstant WR, Mittermeier RA, et al. 2002. Hotspots
1219   and the conservation of evolutionary history. *Proceedings of the National Academy of*
1220   *Sciences USA* 99: 2067-71
1221 Slater GJ, Harmon LJ. 2013. Unifying fossils and phylogenies for comparative analyses of
1222   diversification and trait evolution. *Methods in Ecology and Evolution* 4: 699-702
1223 Slatyer RA, Hirst M, Sexton JP. 2013. Niche breadth predicts geographical range size: a general
1224   ecological pattern. *Ecology Letters* 16: 1104-14
1225 Soberón J, Jiménez R, Golubov J, Koleff P. 2007. Assessing completeness of biodiversity
1226   databases at different spatial scales. *Ecography* 30: 152-60
1227 Soria-Auza RW, Kessler M. 2008. The influence of sampling intensity on the perception of the
1228   spatial distribution of tropical diversity and endemism: a case study of ferns from
1229   Bolivia. *Diversity and Distributions* 14: 123-30
1230 Stropp J, Ladle RJ, Malhado ACM, Hortal J, Gaffuri J, et al. 2015. Mapping ignorance: 300 years
1231   of knowledge of flowering plants in Africa. *Global Ecology and Biogeography*:
1232   conditionally accepted
1233 Sutherst R. 2014. Pest species distribution modelling: origins and lessons from history.
1234   *Biological Invasions* 16: 239-56
1235 Swenson NG, Enquist BJ. 2009. Opposing assembly mechanisms in a Neotropical dry forest:
1236   implications for phylogenetic and functional community ecology. *Ecology* 90: 2161-70
1237 Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation
1238   biodiversity assessment using DNA metabarcoding. *Molecular Ecology* 21: 2045-50

1239 Thompson JN. 1997. Conserving interaction biodiversity. In *The ecological basis of*
1240   *conservation: heterogeneity, ecosystems, and biodiversity*, ed. STA Pickett, RS Ostfeld,
1241   M Shachak, GE Likens, pp. 285-93. New York: Chapman & Hall
1242 Thompson PL, Davies TJ, Gonzalez A. 2015. Ecosystem functions across trophic levels are linked
1243   to functional and phylogenetic diversity. *PLoS ONE* 10: e0117595
1244 Thomson JR, Mac Nally R, Fleishman E, Horrocks G. 2007. Predicting bird species distributions
1245   in reconstructed landscapes. *Conservation Biology* 21: 752-66
1246 Traugott M, Kamenova S, Ruess L, Seeber J, Plantegenest M. 2013. Empirically characterising
1247   trophic networks: what emerging DNA-based methods, stable isotope and fatty acid
1248   analyses can offer. *Advances in Ecological Research* 49: 177-224
1249 Valiente-Banuet A, Aizen MA, Alcántara JM, Arroyo J, Cocucci A et al. 2014. Beyond species
1250   loss: the extinction of ecological interactions in a changing world. *Functional Ecology* in
1251   press, doi:10.1111/1365-2435.12356
1252 Verdú JR, Alba-Tercedor J, Jiménez-Manrique M. 2012. Evidence of different thermoregulatory
1253   mechanisms between tywo sympatric *Scarabaeus* species using infrared thermography
1254   and micro-computer tomography. *PLoS ONE* 7: e33914
1255 Villeger S, Miranda JR, Hernández DF, Mouillot D. 2010. Contrasting changes in taxonomie vs.
1256   functional diversity of tropical fish communities after habitat degradation. *Ecological*
1257   *Applications* 20: 1512-22
1258 Violle C, Navas ML, Vile D, E. K, Fortunel C et al. 2007. Let the concept of trait be functional!
1259   *Oikos* 116: 882-92
1260 Walther BA, Moore JL. 2005. The definitions of bias, precision, and accuracy, and their use in
1261   testing the performance of species richness estimators, with a literature review of
1262   estimator performance. *Ecography* 28: 815-29
1263 Warren DL, Cardillo M, Rosauer DF, Bolnick DI. 2014. Mistaking geography for biology: inferring
1264   processes from species distributions. *Trends in Ecology & Evolution* 29: 572-80
1265 Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ. 2005. Conservation
1266   biogeography: assessment and prospect. *Diversity and Distributions* 11: 3-23
1267 Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J et al. 2013. The role of biotic interactions in
1268   shaping distributions and realised assemblages of species: implications for species
1269   distribution modelling. *Biological Reviews* 88: 15-30
1270 Wolfe DA, Champ MA, Flemer DA, Mearns AJ. 1987. Long-term biological data sets: Their role
1271   in research, monitoring, and management of estuarine and coastal marine systems.
1272   *Estuaries* 10: 181-93
1273 Yahara T, Donoghue M, Zardoya R, Faith DP, Cracraft J. 2010. Genetic diversity assessments in
1274   the century of genome science. *Current Opinion in Environmental Sustainability* 2: 43-
1275   49
1276 Yodzis P. 1988. The indeterminacy of ecological interactions as perceived through perturbation
1277   experiments. *Ecology* 69: 508-15
1278

Supplementary materials

Supplemental Table 1. Main consequences of the seven shortfalls of biodiversity knowledge, the long-term strategies necessary to fill them and the short-term strategies to account for the uncertainty they produce.

Shortfall	Consequences	Short-term strategies to account for uncertainty	Long-term strategies for filling in the shortfall
Linnean	Misidentifications		
	Inaccurate estimates of global extinction rates		
	Inaccurate estimates of species diversity gradients	Estimates based on existing species richness Conduct sensitivity analyses Extrapolate across taxa or regions	Achieve a consensus about species concepts and taxon delimitation for extinct and extant organisms Increase taxonomical effort and expertise Planned survey designs Global sharing of taxonomical information
Wallacean	Lack of knowledge of species' responses to global change		
	Inaccurate estimates of threat		
	Uncertainty about conservation prioritization schemes and systematic conservation planning Compromised efficacy of Species Distribution Models Biased knowledge on diversity gradients	Use large-scale databases Include data on species range limits Develop theoretical simulations involving niche modelling and relevant traits Incorporate maps of ignorance to account for geographical uncertainty in the knowledge	Expand sampling effort and reliability Assess the degree of relevance of dispersal limitations in current distributions Define short-term effects of population dynamics on geographic range structure
Prestonian	Inability to identify population declines and PVA estimates Inability to predict pest outbreaks	Standardize and optimize sampling protocols developed for inventorying and monitoring	Adopt analogous protocols by different teams Increase support for natural history collections and expeditions
Darwinian	Inaccurate estimates of evolutionary patterns Unrealistic estimates of the history of	Use taxonomic classifications as a proxy for evolutionary relationships	Obtain molecular data for poorly-studied groups Combine molecular and fossil information

	diversity Creationism Ineffectiveness of comparative method Lack of understanding of species' adaptive reactions to global change	Use simulations of evolutionary processes Use super- and mega-trees	to better estimate branch length and calibration Create more realistic evolutionary models, including trait evolution
Raunkiæran	Lack of ability to predict species' responses to change Inability to predict changes in ecosystem functions Use of inadequate traits Lack of knowledge about trait bundles	Use phylogeny as a proxy for trait diversity Use simple traits or combinations of traits as proxies for functions Use of large-scale databases instead of local measures Careful selection of traits according to the function assessed	Expand standardization of definitions and protocols for more organisms Collect traits for more species and along environmental gradients and temporal series Perform more experimental tests
Hutchinsonian	Lack of ability to predict species' responses to global change Inability to understand the causes of species distributions and abundance Less effective translocations of threatened species Incapacity to forecast the spread of invasive species	Define bioclimatic envelopes as a surrogate of the niche Examine the agreement between physiologically- and distribution-defined estimates of the scenopoetic niche	Obtain more physiological data to delimit species' tolerances Measure intraspecific variability in tolerance limits Delimit the non-equilibrium status of species with the environmental conditions
Eltonian	Lack of ability to predict species' responses to global change Lack of knowledge about assembly rules Inability to predict processes in non-analog communities Difficulty of restoration processes Inability to predict diseases Inability to characterize community structure	Concentrate efforts on the best-studied interactions and well-resolved taxa Produce careful meta-analyses of the best data sets Prioritize studies on interaction networks at sites which hold basic data from other studies (e.g. permanent forest plots)	Set clear and widely applicable definitions of interaction types Develop standards for field procedures to ensure minimum comparability, either longitudinal, across sites or across systems Allocate resources for large-scale field work, prioritizing interactions that are clearly linked to key ecosystem processes and services (e.g. pollination) Invest in applying new technologies to interaction surveys (e.g. fingerprinting or molecular profiling of gut contents)