Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity

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

[1] Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN-CSIC), 28006 Madrid, Spain; email: jhortal@mncn.csic.es, mcnj117@mncn.csic.es

[2] Department of Ecology, Instituto de Ciências Biologicas, Universidade Federal de Goiás, 74001-970 Goiânia, Brazil; email: diniz@icb.ufg.br

[3] Center for Ecology, Evolution, and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

[4] Institute of Botany, Czech Academy of Sciences, CZ-379 82 Třeboň, Czech Republic; email: fradebello@ctfc.es

[5] Department of Botany, Faculty of Sciences, University of South Bohemia, 370 05 Budějovice, Czech Republic

[6] Departamento de Biologia Animal, Instituto de Biologia, Universidade Estadual de Campinas, 13083-862 Campinas, Brazil; email: thomasl@unicamp.br

[7] Instituto de Ciências Biológicas e da Saúde (ICBS), Universidade Federal de Alagoas, 57072-900 Maceió, Brazil

[8] School of Geography and the Environment, University of Oxford, OX1 3QY Oxford, United Kingdom; email: richard.ladle@ouce.ox.ac.uk

Annu. Rev. Ecol. Evol. Syst. 2015. 46:523–49

First published online as a Review in Advance on October 28, 2015

The *Annual Review of Ecology, Evolution, and Systematics* is online at ecolsys.annualreviews.org

This article's doi:
10.1146/annurev-ecolsys-112414-054400

*Corresponding authors

Keywords

scientific ignorance, biodiversity data, knowledge shortfalls, bias, uncertainty, macroecology, functional ecology

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), distribution (Wallacean), abundance (Prestonian), and evolutionary patterns (Darwinian). We also redefine the Hutchinsonian shortfall to apply to the abiotic tolerances of species and propose new shortfalls relating to limited knowledge of species traits (Raunkiæran) 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; classifications are abstractions that researchers use to represent the real world and produce scientific knowledge (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) 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 of any given characteristic of biodiversity is practically unachievable, owing to the interaction between the complex temporal and spatial dynamics of nature and human capacity to survey it (Ladle & 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 & Rodrigues 2003, Mace 2004).

The fundamental and practical limits of biodiversity knowledge mean that scientists must 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 (see the sidebar The Importance of Ignorance), as biased and unrepresentative knowledge compromises our 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 (*a*) identify key shortfalls in biodiversity knowledge; (*b*) review the origins, drivers, and current explanations for these shortfalls; (*c*) assess the consequences of these shortfalls for ecological, evolutionary, and conservation research; and (*d*) propose strategies and tools to overcome these shortfalls and, in parallel, suggest ways in which uncertainties and biases in biodiversity data may be most effectively allowed for in research and conservation practice.

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 rate of production of novel entities exceeds the maximum rate at which we can describe them. In this sense, knowledge shortfalls can be defined as the gaps between realized/extant knowledge and complete knowledge within a biological domain at a given moment in time (normally the 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

THE IMPORTANCE OF IGNORANCE

"Thoroughly conscious ignorance is the prelude to every real advance in science." (*James Clerk Maxwell, quoted in Firestein 2012, p. 48*)

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. Conscious ignorance may help researchers identify the right questions, avoid developing theories based on insufficient or misleading data, and, ultimately, direct scientific research toward significant advancements (Firestein 2012). 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 addressed through further research).

interactions. Five of these shortfalls have 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 aspects studied from a source of information. A suitable example is the fossil record, to which virtually all of the shortfalls apply.

2.1. Linnean Shortfall

The Linnean shortfall is named after Karl von Linné, or Linnaeus (1707–1778), and refers to the discrepancy between formally described species and the number of species that actually exist (Lomolino 2004). We propose that this shortfall should also include the knowledge gap for extinct species. The Linnean shortfall is more severe for organisms that are smaller in size, niche width, or distributional range and which are less complex or phenotypically conspicuous, with this pattern holding both between and within taxonomic groups (Riddle et al. 2011). The magnitude of the Linnean shortfall is unknown for two reasons. First, the number of formally described species is

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

Shortfall	Aspect of biodiversity	Definition
Linnean	Species	Most of the species on Earth have not been described and cataloged (Brown & Lomolino 1998); this concept can be extended to extinct species (discussed in this review)
Wallacean	Geographic distribution	Knowledge about the geographic distribution of most species is incomplete; it is inadequate at all scales most of the time (Lomolino 2004)
Prestonian	Populations	Data on species abundance and population dynamics in space and time are often scarce (Cardoso et al. 2011)
Darwinian	Evolution	Lack of knowledge about the tree of life and the evolution of species and their traits (Diniz-Filho et al. 2013)
Raunkiæran	Functional traits and ecological functions	Lack of knowledge about species' traits and their ecological functions (discussed in this review)
Hutchinsonian	Abiotic tolerances	Lack of knowledge about the responses and tolerances of species to abiotic conditions (i.e., their scenopoetic niche; discussed in this review, redefined from Cardoso et al. 2011)
Eltonian	Ecological interactions	Lack of knowledge on species' interactions and these interactions' effects on individual survival and fitness (discussed in this review)

ON KNOWNS AND UNKNOWNS

"To know that we know what we know, and that we do not know what we do not know, this is true knowledge." (*Henry David Thoreau*, *Walden*, *1854*)

Creating an overview of knowledge (and lack of it) for a given topic involves determining what we do know and what we are aware that we do not know and recognizing that there are facts far beyond our current knowledge. Or, according to Jackson's (2012) bestiary of ignorance, determining the known knowns, the known unknowns, and the unknown unknowns. Jackson's classification tacitly includes a fourth category, the unknown knowns: facts that we have recorded, but which are not easily accessible, or are so basic that we are unaware that we know them. Although not a shortfall, this latter category is particularly important for biodiversity research because it includes information stored in natural history collections and the gray literature that could be made accessible with new bioinformatics tools.

constantly changing as a result of new descriptions, revisions, and unresolved synonyms (Chapman 2009, Baselga et al. 2010, May 2010), as well as difficulties in establishing a unified species concept or agreement on operational tools to delimit different taxa (Dayrat 2005, Hebert & Gregory 2005). The most comprehensive and authoritative global index of species is the Catalogue of Life (**http://www.catalogueoflife.org**), which currently (September 2015) has records for more than 1.6 million species (Roskov et al. 2015). Second, the predicted number of species is highly sensitive to the estimation method adopted and to parameter values: Estimates have ranged from 2 to 100 million eukaryotic species (May 2010), with most recent global species richness estimates converging on a narrower band around 10 million species (Mora et al. 2011, Caley et al. 2014; but see Costello et al. 2012).

The Linnean shortfall incorporates two distinct categories of unknown species: those yet to be sampled, and collected species that have not yet been described. Species in the former category are most frequent in the large unsurveyed regions of the world. These include remote regions such as the forests of southwest Amazonia (Bush & Lovejoy 2007, Hopkins 2007) but also poorly studied ecosystems such as the deep sea (Rex & Etter 2010) or the upper canopies of rainforests (Ellwood & Foster 2004). Collected, but as yet undescribed, species may run into the hundreds of thousands, and our ignorance of them is largely a consequence of a lack of funding for and capacity in global taxonomy (see the sidebar On Knowns and Unknowns). Including these undocumented species in the Catalogue of Life may be problematic, as new entries should be validated by an expert who has scrutinized descriptions and specimens—in many cases, although the taxonomic groups are extant, their corresponding experts are not (Hopkins & Freckleton 2002).

2.2. Wallacean Shortfall

The Wallacean shortfall is named after Alfred Russel Wallace (1823–1913) and refers to lack of knowledge about the geographical distribution of species (Lomolino 2004). This shortfall stems from geographic biases in the information on species distributions (**Figure 1*a***), which cause many maps of observed biodiversity to closely resemble maps of survey effort (Hortal et al. 2007). Such similarity also depends on the grain at which species' distributions are recorded and analyzed (see Section 3.2). Specifically, knowledge of species distributions is intimately connected with temporal and spatial variation in surveying effort (Hortal et al. 2008, Boakes et al. 2010). That some regions

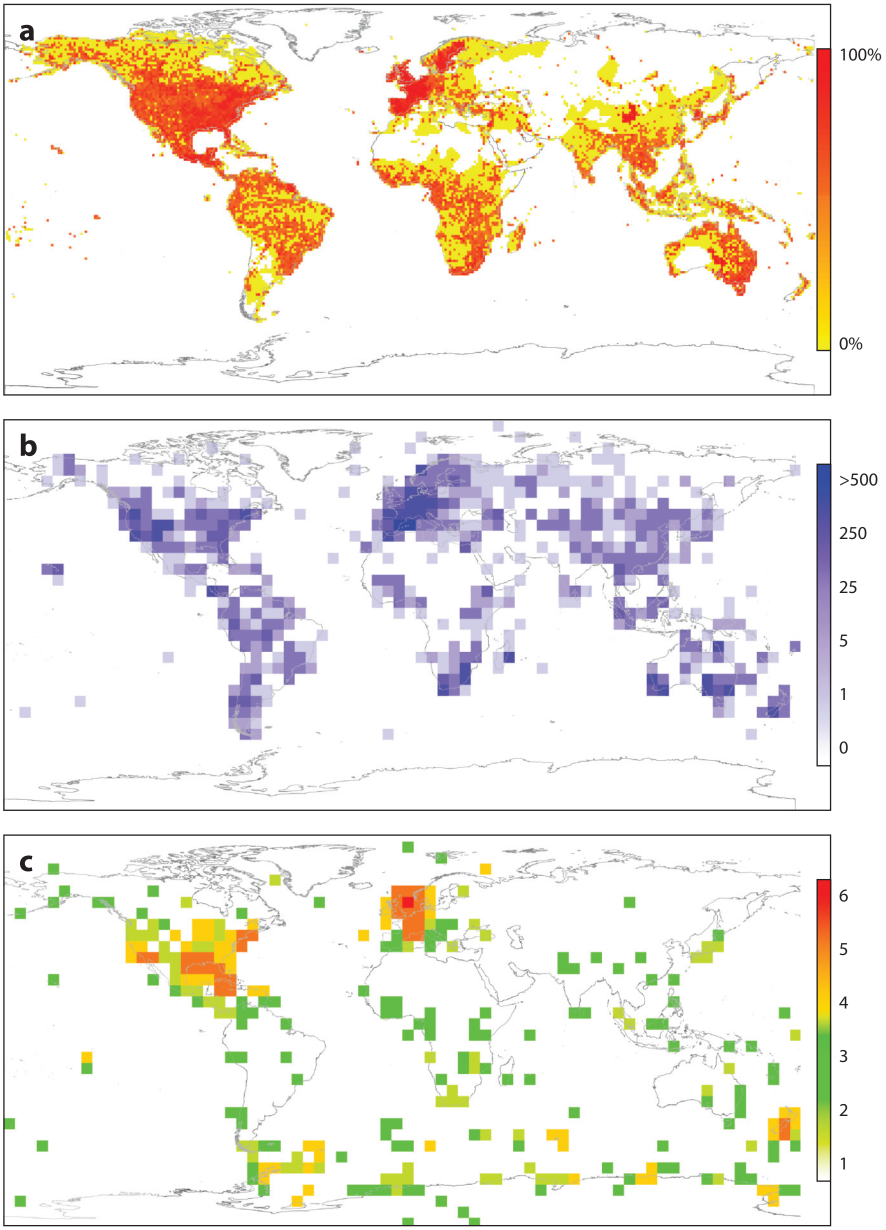

Figure 1

Examples of the global distribution of biodiversity data. (*a*) Completeness of freshwater fish inventories at 1° × 1°-grid-cell resolution; the scale depicts the percentage of total (estimated) species that have already been observed in each cell. Data in panel *a* are from *IPez* and *FishBase*, redrawn from Pelayo-Villamil et al. (2015). (*b*) Data on plant functional traits, presented at 5° × 5°-grid-cell resolution; the transformed scale indicates the number of trait measurement sites. Data in panel *b* are from the TRY database (Kattge et al. 2011a), courtesy of Jens Kattge, updated to January 2014. (*c*) Biotic interactions gathered in global database initiatives, presented at 5° × 5°-grid-cell resolution. The scale indicates the number of recorded interactions. Data in panel *c* come from GloBI (Poelen et al. 2014), courtesy of Jorrit Poelen, extracted November 2014.

are better sampled than others is inevitable given the stark differences in scientific capacity and accessibility between countries and regions (Rodrigues et al. 2010). For example, the Wallacean shortfall is particularly acute in remote and inaccessible regions, such as the forests of southwest Amazonia and the Congo basin. Approximately 40% of Amazonia has never been surveyed, and we do not have an accurate geographic distribution for any of the plant species that occur in this region (Bush & Lovejoy 2007). It should also be noted that distribution data typically vary among political rather than ecological units and may therefore be heavily biased in relation to historical patterns of collecting, analyzing, and collating biogeographical data (Rodrigues et al. 2010, Meyer et al. 2015).

2.3. Prestonian Shortfall

The Prestonian shortfall can be defined as lack of knowledge about the abundance of species and their 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 results from three main factors: (*a*) the difficulty of producing an accurate census for many organisms, (*b*) the costs of gathering long-term data (Wolfe et al. 1987), and (*c*) rapid fluctuations in species abundance that necessitate frequent resampling. The Global Population Dynamics Database (GPDD) addresses this shortfall by collating and compiling comparable data (Inchausti & Halley 2001). However, uncertainties in estimates of population size can affect the results (and conclusions) of analyses of the data in the GPDD and in most population time series (Clark & Bjørnstad 2004, Knape & de Valpine 2012). The duration of studies on abundance dynamics also affects estimates of minimum viable populations (Reed et al. 2003), accentuating 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 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 past 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 need to be consecutively 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 in 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. 2012b). Second, although phylogenies based on branch lengths from molecular data are usually estimated with acceptable accuracy, no studies have yet evaluated how errors in estimating branch lengths at distinct levels of the phylogeny 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, our understanding of how ecological traits relate to biodiversity patterns is hampered by current phylogenies, which provide limited information about trait evolution, even when they are accurate (Freckleton et al. 2002, Blomberg et al. 2003, Cadotte et al. 2013). When researchers use phylogeny as a surrogate of trait variation, they assume a linear relationship between trait differentiation among species and time since divergence (see Hansen & Martins 1996). This assumption is valid only under neutral processes, and our ability to fit 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 2011, Sansom et al. 2015).

2.5. Raunkiæran Shortfall

We define the Raunkiæran shortfall as lack of knowledge about species traits and their functions. This includes not only trait variation both within (Kingsolver et al. 2001) and between (Roy & Foote 1997) species, but also the ecological function (or functions) that ensues as a result of 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 Raunkiær (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 followed without an explicit or direct link to the organisms' fitness (e.g., Ricklefs 2012), which is of implicit interest in this type of analysis. Functional trait-based approaches are increasingly being used in 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.

A functional trait is defined as any phenotypic attribute that affects the fitness of organisms and/or their influence on other organisms and on ecosystem functions (Violle et al. 2007). Functional traits are measured at the individual level when possible. They are used either 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 & Laughlin 2013) or to investigate the effects of species on various ecosystem processes (de Bello et al. 2010, Díaz et al. 2013). A major characteristic of the Raunkiæran shortfall is that the traits that are generally measured are often the most simple, rather than the most functional. There is also a noticeable difference between the functional traits studied by plant and animal ecologists. Whereas plant ecologists have traditionally focused on traits that either mediate the response of species to environmental gradients and biotic interactions or are related to ecosystem functions, animal ecologists have primarily studied how traits mediate the response 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 for 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. 2011a). Nevertheless, significant taxonomic and geographic gaps remain (**Figure 1*b***).

The clear standardization of trait definitions and measurements by Cornelissen et al. (2003) has fostered data gathering for plants (Kattge et al. 2011b). However, such standardizations are not yet available for many taxa, hampering data gathering and sharing (Violle et al. 2007). Moreover, such standardization may result in researchers concentrating their efforts on a limited number of traits and, therefore, reducing research of other traits that may be revealing for particular groups or poorly understood ecological functions. In this context, intraspecific variability in traits is often neglected, leading to biases (Albert et al. 2010) and, depending on the level of aggregation, the trait, and the habitat type, limiting the accuracy of traits retrieved from a database (Cordlandwehr et al. 2013). The success of trait-based approaches also depends on the selection of a sufficient number of informative functional traits (Laughlin & Laughlin 2013), either linked to species' tolerance of abiotic and biotic conditions or to the effects of species on ecosystems (Cornelissen et al. 2003). This selection is often arbitrary, reducing the generality of results. Easily measurable traits are usually chosen, even though their relation to fitness components is often tenuous (Violle et al. 2007). Specifically, this approach favors morphological traits over behavioral, physiological, or life history traits (see Kingsolver et al. 2001). Furthermore, although evidence exists that clusters of traits can influence demographic and fitness components (Poorter et al. 2008, Adler et al. 2014), the functional consequences of trait covariation have been poorly studied.

2.6. Hutchinsonian Shortfall

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

The estimation of environmental preferences and tolerances is a fundamental objective of ecology. These estimates can be used to improve predictions of the responses of species to changing conditions and, ultimately, to increase understanding of the causes of abundance and occurrence patterns (Brown 1984, Gaston 2003, Slatyer et al. 2013). Species tolerances are inferred from one of two sources (Bozinovic et al. 2011, Diamond et al. 2012, Sánchez-Fernández et al. 2012). The first source is physiologically derived niche data—physiological data and performance curves generally obtained from complicated and costly laboratory experiments in the case of animals (Bozinovic et al. 2011) or greenhouse experiments in plants—and/or field experimentation for some plant and invertebrate species (Lambers et al. 2008, Diamond et al. 2012). The second source is occurrence-derived niche data estimated from field observations, currently obtained from correlative niche models (Peterson et al. 2011). Although physiologically derived niche data are generally more precise than occurrence-derived niche data, they are not necessarily more realistic, owing to variability in response curves among conspecific populations, phenotypic plasticity, and the acclimation ability of each study population (e.g., McCann et al. 2014), as well as evolutionary and heritable change in physiological tolerance (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 studied.

However, occurrence-derived niche data can also produce misleading values for species' tolerances and optima. The geographical distributions of species occurrences and environmental variables define the realized niche of a species (so-called Hutchinson's duality; Colwell & 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 presence and environmental conditions. This occurs regardless of the origin of the spatial structure in populations. Indeed, many factors other than the environment can create such a 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 nonequilibrium 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 a species; Dunning et al. 1995, Boyce 2006)—in addition to lack of information about microenvironmental variability within each locality (see McInerny & Purves 2011)—also compromises the quality of estimates of niche parameters based on occurrence data.

2.7. Eltonian Shortfall

The Eltonian shortfall is named after Charles Sutherland Elton (1900–1991), the pioneer of the concepts of food chains and food webs. This shortfall refers to 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. Although such attributes are all recognized aspects of biodiversity, interactions and other processes can be viewed as extraneous expansions 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 using matrices in which each filled cell represents a recorded interaction between a pair of species; values in cells may represent frequencies of interaction events or interaction strengths measured as consumed biomass or number of interacting individuals, for example. These matrices can be described using standard diversity metrics such as the Shannon or Simpson diversity indices (Blüthgen et al. 2008). Moreover, the distribution 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, such as the distribution of trophic specialization or connectivity, can be obtained that characterize functional aspects of community structure (Blüthgen et al. 2008, Jorge et al. 2014). Interaction diversity research and functional diversity research 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). Further, data on interactions gathered by different observers often differ owing to variation in collection methods, research objectives, and criteria. This limits the generalizability of conclusions from compilations and databases (e.g., Pimm et al. 1991). In particular, the choice of recording method has a profound influence on the extent and quality of data and results. Interactions that involve fleeting contacts, e.g., pollination visits, have to be recorded by direct observation, which usually requires massive field effort (e.g., 171 h of field observation in a desert ecosystem recorded 55% of the total plant-pollinator interactions estimated to exist in that system; Chacoff et al. 2012). Plant-frugivore interactions have been scored from vertebrate regurgitates or gut/fecal contents (Poulin et al. 1999), but identification of propagules is based on morphological recognition, sowing, or, increasingly, molecular studies (Marrero et al. 2009, González-Varo et al. 2014). More intimate interactions often require collecting hosts (animals or plants) and dissecting or rearing from them (Novotny et al. 2010, Poulin 2010). It is also almost impossible to distinguish a rare but genuine externally feeding herbivore, such as a mobile leaf-eater, from the occasional nonfeeding tourist

BIAS, PRECISION, ACCURACY, AND UNCERTAINTY

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

without further confirmation from field observations or feeding trials (Ødegaard 2004). All of these problems typically reduce the volume of interaction data that can be obtained, so studies using standardized techniques to measure interactions throughout several biomes are scarce (see Schemske et al. 2009). As a result, interaction data are often insufficient for the intended analyses or may even produce biased results.

3. A UNIFIED FRAMEWORK FOR UNDERSTANDING AND MANAGING BIODIVERSITY SHORTFALLS

The seven shortfalls compromise the accuracy, generality, and realism of biodiversity knowledge. It is therefore crucial to understand their characteristics and their implications for biodiversity research. In this section, we outline the main interactions between the shortfalls, how they are affected by the spatial and temporal scale of analysis, and how they can result in biased and/or inaccurate knowledge (see the sidebar Bias, Precision, Accuracy, and Uncertainty).

3.1. Interactions Between Shortfalls

Biodiversity shortfalls combine with each other in several ways (**Figure 2*a***). The Linnean shortfall necessarily affects all the others because it represents a lack of knowledge about the basic units of study in ecology and evolution. Clearly, we have no empirical data on the characteristics of unknown species, although, arguably, some of these characteristics can be estimated using models fitted to ecological and evolutionary data about related species (Raxworthy et al. 2003) or attributed to operational taxonomic units (e.g., Yahara et al. 2010). However, the biases in our knowledge of living species caused by the Linnean shortfall mean that many phylogenies have significant numbers of missing taxa, which can affect the final tree topology and compromise the estimated rates of evolution (Nee et al. 1994, FitzJohn et al. 2009). Thus, an incremental increase in knowledge of any given aspect of biodiversity must be preceded, or at least accompanied, by filling in the Linnean shortfall (**Figure 2*b***).

The Wallacean, Prestonian, and Darwinian shortfalls are inextricably related to each other, whereby lack of data for one necessarily affects the other two (**Figure 2*a***). The interdependence between the Prestonian and the Darwinian shortfalls is less strong because limited knowledge of extant populations only creates uncertainty about current population (and metapopulation) dynamics and short-term evolutionary processes. However, this uncertainty can hamper accurate predictions of variability in population size (e.g., McCarthy et al. 1994) or introduce errors in coalescence analyses through the use of invalid estimates of demographic history (Pybus et al. 2000, Drummond et al. 2005). By contrast, the Darwinian shortfall is strongly influenced

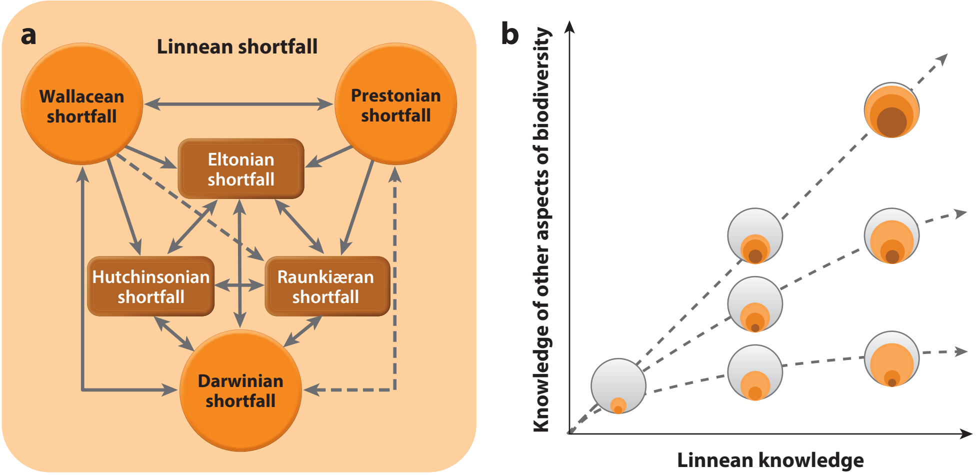

Figure 2

Relationships between shortfalls. Different shades of orange indicate different degrees of complexity in our knowledge about biodiversity, from knowledge of the basic entities (species; *light orange*) to knowledge about their extrinsic characteristics (i.e., their distribution in space and time; *medium orange*) to knowledge about their intrinsic traits (i.e., ecological functionality as measured from species' niches and functional traits; *dark orange*). (*a*) Magnitude of the influence of each shortfall on the others; the Linnean shortfall affects all the others, and the directions of the influences between the rest are indicated by arrows (*dashed arrows* indicate noncritical effects). (*b*) Dependence of shortfalls regarding other aspects of biodiversity on the filling in of the Linnean shortfall; the concentric circles represent the whole range of biodiversity variation (*gray*) and the three kinds of shortfalls (*shades of orange*). The figure depicts three potential scenarios (*dotted lines*) in which investment is devoted to describing species diversity (*below*), documenting the distribution of and variation in species diversity (*middle*), or studying the ecological and evolutionary characteristics of all living entities (*above*).

by the Eltonian, Hutchinsonian, and Raunkiæran shortfalls, as limited data on interactions, ecophysiological responses, and functional traits hamper the description of coevolutionary processes and niche and trait evolution (Diniz-Filho et al. 2013).

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

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

3.2. Scale Dependency of Shortfalls

Biodiversity knowledge shortfalls are scale-dependent in terms of resolution and the extent of data coverage and analysis (Hartley & Kunin 2003, Whittaker et al. 2005). This is most easily illustrated with respect to the Linnean and Wallacean shortfalls (**Figure 3**). At the largest possible grain size (the entire Earth), we have perfect knowledge of the distribution of any species that has been described. However, at smaller grain sizes, the Wallacean shortfall begins to expand, as increasingly precise information about distributions is required (McPherson & Jetz 2007, Riddle et al. 2011, Pineda & Lobo 2012). At very small grain sizes, it becomes difficult to even confirm the presence/absence of a species, especially for highly mobile animals that range over wide areas and habitat types (e.g., Boyce 2006). This problem can also be viewed in terms of species turnover at smaller grain sizes. The sensitivity of beta diversity to sampling grain reflects the effect of local environmental heterogeneity: As sampling grain increases, biotic assemblages

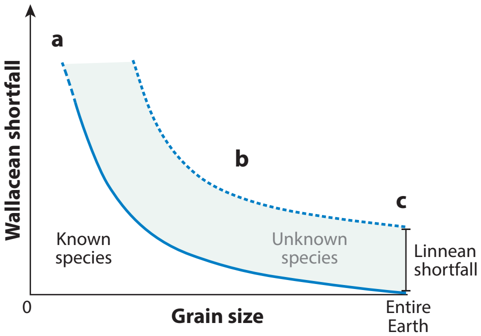

Figure 3

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 of faunas and floras (i.e., distributional atlases) is 1 to 100 km^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.

appear more homogeneous (MacNally et al. 2004). A strong temporal signal also exists at smaller spatial resolutions, with distributions fluctuating to different degrees depending on the ecological characteristics of the species (Thomson et al. 2007).

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, as a reduction in global biodiversity truncates the shortfalls (see Costello et al. 2013), which are also partially alleviated by the much slower process of cladogenesis. The Raunkiæran shortfall will also be influenced by anagenesis, which leads to shifting trait values over time. The processes of adaptation and microevolution influence the Eltonian and Hutchinsonian shortfalls, given the central role of natural selection in mediating abiotic and biotic interactions (McLachlan & Ladle 2011) combined with the shuffling of species co-occurrences over time. In summary, the size and nature of biodiversity shortfalls are highly dependent on the scale at which we collect, analyze, and aggregate data.

3.3. Biases in Biodiversity Knowledge

As a rule, data on all aspects of biodiversity are taxonomically, geographically, and temporally biased. Certain groups have received far more attention than others, to the point that it is estimated that only ~7% of fungi 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, more conspicuous, and more easily detectable species are recorded earlier and more extensively (Gaston & 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, more and better data exist for functional traits in plants than for other organisms, such as insects (Poff et al. 2006, Gibb & Parr 2013) or vertebrates (Villeger et al. 2010), owing to the long botanical tradition of functional ecology (Díaz & Cabido 2001, Cornelissen et al. 2003).

Species occurrence and assemblage data are also biased toward certain regions, habitats, and environmental domains (Lobo et al. 2007, Loiselle et al. 2008; see **Figure 1**). This may be the most studied bias in biodiversity data and involves several syndromes caused by the behavior of researchers (Sastre & Lobo 2009) and historical patterns of colonization and inventorying (Bini et al. 2006, Meyer et al. 2015). Thus, biodiversity inventories are typically more comprehensive near locations that offer convenient access, infrastructure, and logistics (Dennis & 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 & Kitching 2007, Soria-Auza & Kessler 2008, Boakes et al. 2010). Consequently, errors exist in the data on the distribution of endangered species and conservation targets, and these errors can be several orders of magnitude larger in less-studied groups. Furthermore, temporal shifts in the spatial coverage of surveys result in spurious data on changes in distributions over time (Lobo et al. 2007, Hortal et al. 2008), affecting our ability to identify past range shifts (Huisman & Millar 2013) and ascertain current patterns of extinction (Lobo 2001).

Geographical variation in survey effort creates an uneven pattern of species descriptions through time (Rich 2006, Baselga et al. 2010), 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 owing 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 about the evolution, diversity, and interrelationships of traits (**Figure 4**).

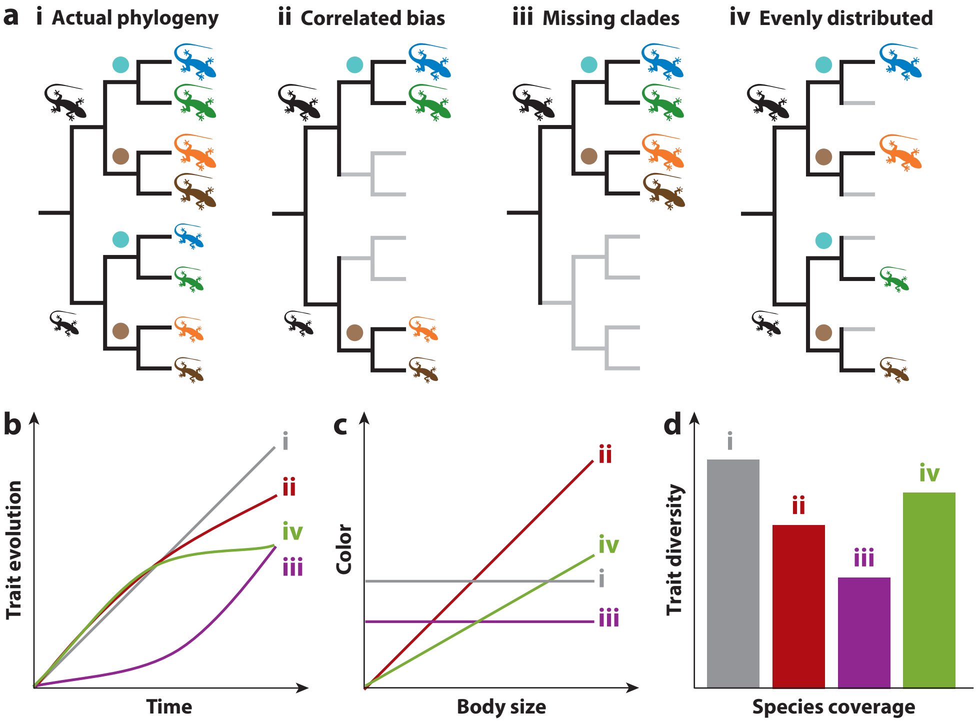

Figure 4

Effects of bias and completeness on our knowledge of trait evolution and trait diversity. (*a*) Different types of knowledge about the species diversity of a hypothetical lizard clade and their impact on the perceived phylogeny and evolution of body size (large/small) and color (from *blue* species to *blue* and *green* species, and from *brown* species to *orange* and *dark brown* species). In scenario *i*, all species are known, whereas half of the species are missing in the other three scenarios, in different arrangements: (*ii*) A phylogenetically correlated bias exists 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 throughout the phylogeny. (*b*) Perceived rates of trait evolution through time for each scenario in panel *a*. (*c*) Perceived relationship between the two traits for each scenario in panel *a*. (*d*) Perceived total trait diversity (i.e., number of different combinations of trait values) for each scenario in panel *a*.

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 that host large research institutions or field stations (Schemske et al. 2009). Hence, both Raunkiæran and Eltonian shortfalls are more extensive outside of North America and Europe (see **Figure 1**). The same applies to marine systems, in which the extensive and taxonomically comprehensive data on trophic interactions that are 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 energy flows in the open ocean and impedes our ability to detect global processes (Pauly & Palomares 2005).

4. CONSEQUENCES AND IMPLICATIONS OF KNOWLEDGE SHORTFALLS

Supplemental Material

Shortfalls in global biodiversity data have numerous implications for theory and practice (see **Supplemental Table 1**; follow the **Supplemental Material link** from the Annual Reviews home page at **http://www.annualreviews.org**). The Linnean and Wallacean shortfalls have the farthest-reaching influence because data on the identity and distribution of species are 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, in which species with small ranges are given higher priority. Indeed, range restriction is an integral part of IUCN (2011) criteria to identify and classify species in danger of global extinction. Several conservation prioritization methods (e.g., Rodrigues et al. 2004) adopt an arbitrary criterion of $<50,000\ \text{km}^2$ 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 certain taxa and/or regions may be unduly prioritized for conservation action (Riddle et al. 2011).

Bias in biodiversity data also critically influences the generality and realism of concepts or models. If missing data are evenly distributed, the corresponding knowledge may lack precision, but it will be realistic and to some extent generalizable (see scenario IV in **Figure 4**). However, the more unevenly distributed the missing data are, the less accurate the representation of the actual patterns or processes under study will be. This can lead to spurious hypotheses being created to explain biased data sets. For example, Nelson et al. (1990) demonstrated that supposed centers of plant endemism in Amazonia were partly an artifact of biased herbarium collections—and these patterns were 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), widely used analytical and predictive tools in conservation (Peterson et al. 2011). SDMs typically relate field observations of species occurrences (and sometimes absences) to environmental (usually climatic) predictors; they do so by using statistically or theoretically derived response surfaces that are supposed to represent the tolerance of species to abiotic conditions. SDMs are routinely used to study rare species, for which 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 consistently generate unrealistic results (Rocchini et al. 2011), particularly for rare species (Elith & Leathwick 2007). The effectiveness of SDMs is also hindered by our lack of knowledge about species interactions, as this limits our ability to incorporate them 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 with more complete data on the abiotic tolerances of species. Further, there is ambivalent evidence regarding the use of phylogeny as a proxy for traits (Swenson & Enquist 2009, Pavoine et al. 2013). The lack of trait data and detailed phylogenies makes it difficult to evaluate the extent to which phylogenetic information can be used as a proxy for trait variation (Díaz et al. 2013). In general, the creation of robust and practically useful models of the effects of threats on biodiversity and the impact of biodiversity loss for ecosystem functioning remains a distant possibility.

5. DEALING WITH THE SHORTFALLS

The past 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 megaprojects is the Map of Life (MoL), a web-based tool (**http://www.mol.org**) that aims to represent the distribution of every species on Earth (Jetz et al. 2012a). As of September 2015, the MoL is based on 194 data sets, 937,810 species, and 371,807,359 records. If it is ever fully realized, this and similar initiatives such as the Global Biodiversity Information Facility (**http://www.gbif.org/**), the Catalogue of Life, the Encyclopedia of Life (**http://www.eol.org**), and the Paleobiology Database (**http://paleobiodb.org**) will provide invaluable resources for the advancement of ecological research and conservation planning (Riddle et al. 2011). These global database projects are extremely important for diminishing the unknown knowns; by cataloging, organizing, and making information accessible, 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. These data must be accompanied by sound metadata covering the ancillary information associated with 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: (*a*) The extent of the knowledge gaps and their patterns within the body of attainable knowledge must be described in a tractable way, (*b*) the major biodiversity data shortfalls (and their associated biases) must be significantly reduced, and (*c*) appropriate methods must be developed to deal with or account for the intrinsic limitations of the quality, longevity, and coverage of biodiversity data (Ladle & Hortal 2013).

A necessary first step to deal with the missing information is to circumscribe and inventory the ignorance that surrounds and is included within our current knowledge, through atlases of ignorance, for example, 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 controversy about the realism of the estimates (cf. Löbl & Leschen 2014), these assessments of the magnitude of the Linnean shortfall provide baselines that allow us to plan for the extent of the task. Equivalent estimates can be developed for other shortfalls; these estimates can be used to compare estimated species geographic ranges with occurrence records (Meyer et al. 2015), for example, or to evaluate the congruence of estimated phylogenies with partial phylogenetic trees based on available data (see Antonelli et al. 2014). Most importantly, such 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. In addition, readily available instruments allow researchers to measure ecophysiological features of some invertebrates 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 may also revolutionize the volume and standardization of recordings of pollinator or disperser visits to flowering (Celep et al. 2014) or fruiting (Prasad et al. 2010) plants. DNA fingerprinting is not yet in wide use, but it promises a major breakthrough in the large-scale recording of predator-prey and plant-herbivore interactions (Kress et al. 2015). In mass samples, the gut contents of invertebrate herbivores may suffice to establish that they feed on a given plant; time-consuming observational or experimental confirmation may become unnecessary (Pinzón-Navarro et al. 2010). Another powerful approach entails using stable isotopes to trace trophic interactions or establish 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 researchers obtaining meaningless results from complex analytical techniques (e.g., Lozier et al. 2009). 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 database projects will be less useful if they do not diminish bias and increase the environmental and spatial representativeness of the available data. As in information technology, biodiversity science is particularly sensitive to the garbage in, garbage out problem, whereby biased input data (garbage in) frequently produces 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, maps of ignorance that provide information on which data are reliable or uncertain (Rocchini et al. 2011) can be used to inform the design and implementation of new surveys (Hortal & 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 by incorporating the uncertainty produced by current gaps and biases in data (Beck et al. 2012). A good starting point would be to routinely present estimates of uncertainty together with data, results, and/or model projections. This approach could be implemented with adequate visualizations and graphical representations in multimedia environments that allow simultaneous communication of results and their associated uncertainty (McInerny et al. 2014). For example, uncertainty in observed species richness patterns within a region could be depicted by blurring the colors of the map (scaled to represent the observed values) according to the level of uncertainty at 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 known relationships and/or processes—although this would exclude most of the biologically richest and most 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, outlining how robust the conclusions are given the current level of uncertainty.

Quantitative descriptions of ignorance can allow researchers to incorporate uncertainty explicitly into modeling and decision-making processes, by, for example, deploying fuzzy logic (Petrou et al. 2014). Most regression techniques allow researchers to incorporate covariates to account for measurement error, but an increasingly popular analytical strategy in ecological and evolutionary research is the switch from significance testing to probabilistic approaches and Bayesian analyses (Clark 2005, Beck et al. 2012). The utility of Bayesian statistics is currently being put to the test in SDMs (Beale & Lennon 2012; F. Rodríguez-Sánchez, D.W. Purves, A. Hampe, P. Jordano & D. Coomes, unpublished data). Here, maps of ignorance can be used as spatially explicit estimates of uncertainty for assessing model sensitivity (Rocchini et al. 2011) or for including uncertainty in models (e.g., through Bayesian techniques). Similar approaches could be used to account for the uncertainty caused by other shortfalls.

6. CONCLUSIONS

The aim of this review was to describe the key remaining areas of ignorance about biodiversity, the relationships between them, and their implications for research and conservation practice. Important first steps toward obtaining a better knowledge of biodiversity are for us to recognize how much we do not know and identify critical gaps in knowledge. This explicit concern with ignorance is not, in any sense, a criticism of the quality of the extensive work carried out by past generations of naturalists and researchers. Rather, we are now able to recognize and describe our ignorance because their work has progressively expanded and defined the envelope of our knowledge. It is because we could build on the work of many industrious researchers that we are now able to compile vast amounts of biodiversity data and envisage what is yet to be discovered. Once we have sufficiently described biodiversity knowledge and its biases and limitations, our task will be to improve the inventory of global biodiversity research in a way that maximizes coverage and most effectively deploys the very limited resources available for such work.

Whichever approach is adopted to fill in the gaps in our knowledge—targeting strategically chosen biases or continuing the scattershot approach of leaving survey decisions to chance—the shortfalls are never likely to be filled in their entirety, and they certainly will not be filled in the near future. This poses a serious dilemma for conservationists and policymakers, who cannot wait years or decades for data of sufficient quality to become available. Of course, such dilemmas are not unique to conservation; it is commonly agreed that the information revolution has made the world increasingly volatile, uncertain, complex, and ambiguous. As a consequence, the biggest challenges faced by modern societies are not problems that can be solved; instead, they take the form of dilemmas that must be managed (Johansen 2007). In this sense, we assert that, through careful analysis and inventive visualization of ignorance, the uncertainty caused by biodiversity data shortfalls can be incorporated into conservation planning and biodiversity research. This will create awareness of the quality of our current knowledge and, by extension, of the models developed with it. Even if this approach slows the pace of new analyses or conservation actions—which will not necessarily be the case—the conclusions produced by research in which uncertainty is taken into account are likely to be more robust and persistent. Likewise, the long-term success of conservation actions based upon them is likely to be greater. To quote Daniel J. Boorstin (1983), "The greatest enemy of knowledge is not ignorance, it is the illusion of knowledge."

SUMMARY POINTS

1. Ecology, evolution, and conservation science are entering the era of big data through the use of massive databases on different aspects of biodiversity. It is time to assess the quality, extent, and representativeness of the available information.
2. We identify seven main shortfalls in biodiversity data. One represents the lack of information on the total taxonomic extent of biodiversity (the Linnean shortfall). Three deal with the most important extrinsic characteristics of species: their geographic distribution, population dynamics, and evolutionary relationships (the Wallacean, Prestonian, and Darwinian shortfalls, respectively). Three others relate to ecological functioning in terms of functional traits, abiotic niches, and biotic interactions (the Raunkiæran, Hutchinsonian, and Eltonian shortfalls, respectively).
3. All knowledge shortfalls are interconnected to varying degrees, according to scale and spatial, temporal, and taxonomic coverage. The Linnean shortfall critically affects all of the others because lack of information on unknown species necessarily prevents the description of any other aspect of biodiversity. Lack of knowledge about extrinsic characteristics typically aggravates shortfalls in knowledge about ecological functioning. Shortfalls in knowledge about ecological functioning, in turn, are tightly entangled with one another.

4. Global biodiversity data are generally incomplete and suffer from many biases. Most current knowledge concerns large, conspicuous, or economically valuable taxa and comes from research in northern, temperate regions and areas that are easy or convenient to access.
5. Bias generates uncertainty in all analyses of biodiversity, compromising the generality and validity of theoretical knowledge on ecology and evolution, and the quality of conservation assessments and actions.
6. Ignorance of biodiversity must be carefully described, quantified, and mapped. This will identify knowledge gaps that require additional research effort and allow researchers to assess uncertainty in estimates and model projections.

FUTURE ISSUES

1. Estimates of the extent and variability of all aspects of biodiversity will be used as baselines to compare with current knowledge stored in large databases. These estimates can be used to identify gaps in knowledge and to direct research and exploratory works.
2. Maps of biogeographical ignorance based on the coverage of spatial and environmental gradients through time provided by distribution data—as gathered in the Global Biodiversity Information Facility or the MoL—should be used to design surveys, assess model uncertainty, and develop Bayesian SDMs.
3. Technological advances (e.g., DNA barcoding, automated species identification, artificial intelligence–assisted remote sensing) will revolutionize the collection of information for all domains of biodiversity knowledge, massively accelerating the rate of data capture.
4. Ongoing compilations of phylogenetic information, such as the SUPERSMART project, will be used to identify gaps in the known tree of life, allowing researchers to target clades in need of further study.
5. Assessments of the bias and representativeness of the information contained in long-term data stored in biodiversity databases on population dynamics; functional traits, including ecophysiological information; and biotic interactions will be developed to identify gaps in knowledge and to better target taxa, biomes, and ecosystems.
6. New ways of representing and communicating uncertainty should be developed to raise awareness of the certainties behind the uncertainty and the extent of current ignorance on biodiversity patterns and processes.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We wish to thank Ana Malhado for her help with the logistics of the Maceió meeting, Isaac Pozo for his help with the figures, and Jorrit Poelen and Jens Kattge for providing access to the most recent GloBI and TRY data. We would like to thank Kaustuv Roy for a constructive revision of this review, which helped clarify many points, particularly about the Wallacean and Raunkiæran shortfall. We would also like to thank Carsten Meyer, Paco Rodríguez-Sánchez, Holger Kreft, Robert Guralnick, Walter Jetz, and Jens Mutke for providing us with unpublished manuscripts and results, and them and Andrés Baselga, Luis Mauricio Bini, Alice Francener, Alberto Jiménez-Valverde, Rafael Loyola, Ana Malhado, Greg McInerny, Duccio Rocchini, Ana Santos, Juliana Stropp, Geiziane Tessarolo, Sara Varela, and Marten Winter for many discussions about the sources of ignorance and the shortfalls of biodiversity knowledge. This review was supported by a work meeting funded by the joint CSIC/CNPq project "Developing tools to unify range dynamics and community-level processes into a single analytical framework" (CSIC 2011BR0071). J.H. was supported by a Spanish DGCyT Ramón y Cajal grant and by Brazilian CNPq PVE grant 401471/2014-4. F.d.B. was funded by the Czech Science Foundation, grant P505/12/1296, and by Brazilian CAPES PVE, grant 88881.068053/2014-01. J.A.F.D.-F., R.J.L., and T.M.L. received CNPq Research Productivity fellowships.

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. *PNAS* 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 versus interspecific variability in plant traits. *Funct. Ecol.* 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. *PNAS* 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 α and β taxonomy in Eupelmid wasps: determinants of the probability of describing good species and synonyms. *J. Zool. Syst. Evol. Res.* 48:40–49

Beale CM, Lennon JJ. 2012. Incorporating uncertainty in predictive species distribution modelling. *Philos. Trans. R. Soc. 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 J, Kitching IJ. 2007. Estimating regional species richness of tropical insects from museum data: a comparison of a geography-based and sample-based methods. *J. Appl. Ecol.* 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. *Divers. Distrib.* 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. *Proc. Am. Philos. Soc.* 93:253–58

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

Boyce MS. 2006. Scale for resource selection functions. *Divers. Distrib.* 12:269–76

Bozinovic F, Calosi P, Spicer JI. 2011. Physiological correlates of geographic range in animals. *Annu. Rev. Ecol. Evol. Syst.* 42:155–79

Brown JH. 1984. On the relationship between abundance and distribution of species. *Am. Nat.* 124:255–79

Brown JH, Lomolino MV. 1998. *Biogeography.* Sunderland, MA: Sinauer. 2nd ed.

Bush MB, Lovejoy TE. 2007. Amazonian conservation: pushing the limits of biogeographical knowledge. *J. Biogeogr.* 34:1291–93

Cadotte M, Albert CH, Walker SC. 2013. The ecology of differences: assessing community assembly with trait and evolutionary distances. *Ecol. Lett.* 16:1234–44

Cadotte MW, Dinnage R, Tilman D. 2012. Phylogenetic diversity promotes ecosystem stability. *Ecology* 93:S223–33

Caley MJ, Fisher R, Mengersen K. 2014. Global species richness estimates have not converged. *Trends Ecol. Evol.* 29:187–88

Cardoso P, Erwin TL, Borges PA, New TR. 2011. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conserv.* 144:2647–55

Caut S, Angulo E, Courchamp F. 2009. Variation in discrimination factors (Δ^{15}N and Δ^{13}C): the effect of diet isotopic values and applications for diet reconstruction. *J. Appl. Ecol.* 46:443–53

Celep F, Atalay Z, Dikmen F, Doğan M, Classen-Bockhoff R. 2014. Flies as pollinators of melittophilous *Salvia* species (Lamiaceae). *Am. J. Bot.* 101:2148–59

Chacoff NP, Vázquez D, Lomáscolo S, Stevani E, Dorado J, Padrón B. 2012. Evaluating sampling completeness in a desert plant-pollinator network. *J. Anim. Ecol.* 81:190–200

Chapman AD. 2009. *Numbers of Living Species in Australia and the World.* Canberra, Aust.: Aust. Biol. Resour. Study. 2nd ed.

Clark JS. 2005. Why environmental scientists are becoming Bayesians. *Ecol. Lett.* 8:2–14

Clark JS, Bjørnstad ON. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology* 85:3140–50

Colwell RK, Rangel TF. 2009. Hutchinson's duality: the once and future niche. *PNAS* 106:19651–58

Cordlandwehr V, Meredith RL, Ozinga WA, Bekker RM, van Groenendael JM, Bakker JP. 2013. Do plant traits retrieved from a database accurately predict on-site measurements? *J. Ecol.* 101:662–70

Cornelissen JHC, Lavorel S, Garnier E, Diaz S, Buchmann N, et al. 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Aust. J. Bot.* 51:335–80

Costello MJ, May RM, Stork NE. 2013. Can we name Earth's species before they go extinct? *Science* 339:413–16

Costello MJ, Wilson S, Houlding B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. *Syst. Biol.* 61:871–83

Dayrat B. 2005. Towards integrative taxonomy. *Biol. J. Linn. Soc.* 85:407–15

de Bello F, Lavorel S, Díaz S, Harrington R, Cornelissen JC, et al. 2010. Towards an assessment of multiple ecosystem processes and services via functional traits. *Biodivers. Conserv.* 19:2873–93

Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, et al. 2015. Finding our way through phenotypes. *PLOS Biol.* 13:e1002033

Dennis R, Thomas C. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder's home range. *J. Insect Conserv.* 4:73–77

Diamond SE, Nichols LM, McCoy N, Hirsch C, Pelini SL, et al. 2012. A physiological trait-based approach to predicting the responses of species to experimental climate warming. *Ecology* 93:2305–12

Díaz S, Cabido M. 2001. Vive la differénce: Plant functional diversity matters to ecosystem processes. *Trends Ecol. Evol.* 16:646–55

Díaz S, Lavorel S, de Bello F, Quétier F, Grigulis K, Robson TM. 2007. Incorporating plant functional diversity effects in ecosystem service assessments. *PNAS* 104:20684–89

Díaz S, Purvis A, Cornelissen JHC, Mace GM, Donoghue MJ, et al. 2013. Functional traits, the phylogeny of function, and ecosystem service vulnerability. *Ecol. Evol.* 3:2958–75

Diniz-Filho JAF, Bastos RP, Rangel TFLVB, Bini LM, Carvalho P, Silva RJ. 2005. Macroecological correlates and spatial patterns of anuran description dates in the Brazilian Cerrado. *Glob. Ecol. Biogeogr.* 14:469–77

Diniz-Filho JAF, Bini LM, Hawkins BA. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Glob. Ecol. Biogeogr.* 12:53–64

Diniz-Filho JAF, Loyola RD, Raia P, Mooers AO, Bini LM. 2013. Darwinian shortfalls in biodiversity conservation. *Trends Ecol. Evol.* 28:689–95

Dornburg A, Beaulieu JM, Oliver JC, Near TJ. 2011. Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. *Syst. Biol.* 60:519–27

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* 22:1185–92

Dunne J, Williams R, Martinez N. 2004. Network structure and robustness of marine food webs. *Mar. Ecol. Prog. Ser.* 273:291–302

Dunning JB, Stewart DJ, Danielson BJ, Noon BR, Root TL, et al. 1995. Spatially explicit population models: current forms and future uses. *Ecol. Appl.* 5:3–11

Eastman JM, Alfaro ME, Joyce P, Hipp AL, Harmon LJ. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution* 65:3578–89

Elith J, Leathwick J. 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Divers. Distrib.* 13:265–75

Ellwood MDF, Foster WA. 2004. Doubling the estimate of invertebrate biomass in a rainforest canopy. *Nature* 429:549–51

Faith D, Collen B, Ariño A, Koleff P, Guinotte J, et al. 2013. Bridging the biodiversity data gaps: recommendations to meet users' data needs. *Biodivers. Inform.* 8:4126

Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15

Firestein S. 2012. *Ignorance: How it Drives Science.* Oxford, UK: Oxford Univ. Press

FitzJohn RG, Maddison WP, Otto SP. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am. Nat.* 160:712–26

Garcillán PP, Ezcurra E. 2011. Sampling procedures and species estimation: testing the effectiveness of herbarium data against vegetation sampling in an oceanic island. *J. Veg. Sci.* 22:273–80

Gaston KJ. 2003. *The Structure and Dynamics of Geographic Ranges*. Oxford, UK: Oxford Univ. Press

Gaston KJ, Blackburn TM. 1994. Are newly described bird species small-bodied? *Biodivers. Lett.* 2:16–20

Gaston KJ, Rodrigues ASL. 2003. Reserve selection in regions with poor biological data. *Conserv. Biol.* 17:188–95

Gibb H, Parr CL. 2013. Does structural complexity determine the morphology of assemblages? An experimental test on three continents. *PLOS ONE* 8:e64005

Gleason HA. 1926. The individualistic concept of the plant association. *Bull. Torrey Bot. Club* 53:7–26

González-Varo JP, Arroyo JM, Jordano P. 2014. Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. *Methods Ecol. Evol.* 5:806–14

Gouveia SF, Hortal J, Tejedo M, Duarte H, Cassemiro FAS, et al. 2014. Climatic niche at physiological and macroecological scales: the thermal tolerance–geographical range interface and niche dimensionality. *Glob. Ecol. Biogeogr.* 23:446–56

Hall BG. 2011. *Phylogenetic Trees Made Easy: A How-to Manual*. Sunderland, MA: Sinauer. 4th ed.

Hansen TF, Martins EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–17

Hartley S, Kunin WE. 2003. Scale dependency of rarity, extinction risk, and conservation priority. *Conserv. Biol.* 17:1559–70

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54:852–59

Hopkins GW, Freckleton RP. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Anim. Conserv.* 5:245–49

Hopkins MJG. 2007. Modelling the known and unknown plant biodiversity of the Amazon Basin. *J. Biogeogr.* 34:1400–11

Hortal J, Jiménez-Valverde A, Gómez JF, Lobo JM, Baselga A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos* 117:847–58

Hortal J, Lobo JM. 2005. An ED-based protocol for optimal sampling of biodiversity. *Biodivers. Conserv.* 14:2913–47

Hortal J, Lobo JM, Jimenez-Valverde A. 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. *Conserv. Biol.* 21:853–63

Huisman JM, Millar AJK. 2013. Australian seaweed collections: use and misuse. *Phycologia* 52:2–5

Hutchinson GE. 1978. *An Introduction to Population Biology*. New Haven, CT: Yale Univ. Press

Inchausti P, Halley J. 2001. Investigating long-term ecological variability using the Global Population Dynamics Database. *Science* 293:655–57

Ings TC, Montoya JM, Bascompte J, Blüthgen N, Brown L, et al. 2009. Ecological networks—beyond food webs. *J. Anim. Ecol.* 78:253–69

IUCN (Int. Union Conserv. Nat.). 2011. *IUCN Red List of Threatened Species*. Gland, Switz.: IUCN

Jackson ST. 2012. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. *Quaternary Sci. Rev.* 49:1–15

Jetz W, McPherson JM, Guralnick RP. 2012a. Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol. Evol.* 27:151–59

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012b. The global diversity of birds in space and time. *Nature* 491:444–48

Johansen B. 2007. *Get There Early*. Oakland, CA: Berrett-Koehler

Jorge LR, Prado PI, Almeida-Neto M, Lewinsohn TM. 2014. An integrated framework to improve the concept of resource specialisation. *Ecol. Lett.* 17:1341–50

Kadmon R, Farber O, Danin A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecol. Appl.* 14:401–13

Kattge J, Díaz S, Lavorel S, Prentice IC, Leadley P, et al. 2011a. TRY—a global database of plant traits. *Glob. Chang. Biol.* 17:2905–35

Kattge J, Ogle K, Bönisch G, Díaz S, Lavorel S, et al. 2011b. A generic structure for plant trait databases. *Methods Ecol. Evol.* 2:202–13

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, et al. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–61

Knape J, de Valpine P. 2012. Are patterns of density dependence in the Global Population Dynamics Database driven by uncertainty about population abundance? *Ecol. Lett.* 15:17–23

Kress WJ, García-Robledo C, Uriarte M, Erickson DL. 2015. DNA barcodes for ecology, evolution, and conservation. *Trends Ecol. Evol.* 30:25–35

Ladle R, Hortal J. 2013. Mapping species distributions: living with uncertainty. *Front. Biogeogr.* 5:8–9

Lambers H, Chapin FS III, Pons TL. 2008. *Plant Physiological Ecology*. New York: Springer-Verlag. 2nd ed.

Laughlin DC. 2014. The intrinsic dimensionality of plant traits and its relevance to community assembly. *J. Ecol.* 102:186–93

Laughlin DC, Laughlin DE. 2013. Advances in modeling trait-based plant community assembly. *Trends Plant Sci.* 18:584–93

Layman CA, Araujo MS, Boucek R, Hammerschlag-Peyer CM, Harrison E, et al. 2012. Applying stable isotopes to examine food-web structure: an overview of analytical tools. *Biol. Rev.* 87:545–62

Lewinsohn TM, Inácio Prado P, Jordano P, Bascompte J, M Olesen J. 2006. Structure in plant-animal interaction assemblages. *Oikos* 113:174–84

Löbl I, Leschen RAB. 2014. Misinterpreting global species numbers: examples from Coleoptera. *Syst. Entomol.* 39:2–6

Lobo JM. 2001. Decline of roller dung beetle (Scarabaeinae) populations in the Iberian peninsula during the 20th century. *Biol. Conserv.* 97:43–50

Lobo JM, Baselga A, Hortal J, Jiménez-Valverde A, Gómez JF. 2007. How does the knowledge about the spatial distribution of Iberian dung beetle species accumulate over time? *Divers. Distrib.* 13:772–80

Logan ML, Cox RM, Calsbeek R. 2014. Natural selection on thermal performance in a novel thermal environment. *PNAS* 111:14165–69

Loiselle BA, Jorgensen PM, Consiglio T, Jimenez I, Blake JG, et al. 2008. Predicting species distributions from herbarium collections: Does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35:105–16

Lomolino MV. 2004. Conservation biogeography. In *Frontiers of Biogeography: New Directions in the Geography of Nature*, ed. MV Lomolino, LR Heaney, pp. 293–96. Sunderland, MA: Sinauer

Losos JB. 2011. Seeing the forest for the trees: the limitations of phylogenies in comparative biology. *Am. Nat.* 177:709–27

Lozier JD, Aniello P, Hickerson MJ. 2009. Predicting the distribution of Sasquatch in western North America: anything goes with ecological niche modelling. *J. Biogeogr.* 36:1623–27

Lukoschek V, Scott Keogh J, Avise JC. 2012. Evaluating fossil calibrations for dating phylogenies in light of rates of molecular evolution: a comparison of three approaches. *Syst. Biol.* 61:22–43

Mace GM. 2004. The role of taxonomy in species conservation. *Philos. Trans. R. Soc. B* 359:711–19

Mace GM, Gittleman JL, Purvis A. 2003. Preserving the tree of life. *Science* 300:1707–9

MacNally R, Fleishman E, Bulluck LP, Betrus CJ. 2004. Comparative influence of spatial scale on β diversity within regional assemblages of birds and butterflies. *J. Biogeogr.* 31:917–29

Marrero P, Fregel R, Cabrera VM, Nogales M. 2009. Extraction of high-quality host DNA from feces and regurgitated seeds: a useful tool for vertebrate ecological studies. *Biol. Res.* 42:147–51

Martinez ND. 1993. Effects of resolution on food web structure. *Oikos* 66:403–12

May RM. 2010. Tropical arthropod species, more or less? *Science* 329:41–42

McCann S, Greenlees MJ, Newell D, Shine R. 2014. Rapid acclimation to cold allows the cane toad to invade montane areas within its Australian range. *Funct. Ecol.* 28:1166–74

McCarthy MA, Franklin DC, Burgman MA. 1994. The importance of demographic uncertainty: an example from the helmeted honeyeater *Lichenostomus melanops cassidix*. *Biol. Conserv.* 67:135–42

McGill BJ, Enquist BJ, Weiher E, Westoby M. 2006. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* 21:178–85

McInerny GJ, Chen M, Freeman R, Gavaghan D, Meyer M, et al. 2014. Information visualisation for science and policy: engaging users and avoiding bias. *Trends Ecol. Evol.* 29:148–57

McInerny GJ, Purves DW. 2011. Fine-scale environmental variation in species distribution modelling: regression dilution, latent variables and neighbourly advice. *Methods Ecol. Evol.* 2:248–57

McLachlan AJ, Ladle RJ. 2011. Barriers to adaptive reasoning in community ecology. *Biol. Rev.* 86:543–48

McPherson JM, Jetz W. 2007. Type and spatial structure of distribution data and the perceived determinants of geographical gradients in ecology: the species richness of African birds. *Glob. Ecol. Biogeogr.* 16:657–67

Meyer C, Kreft H, Guralnick R, Jetz W. 2015. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* 6:82221

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

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

Morlon H. 2014. Phylogenetic approaches for studying diversification. *Ecol. Lett.* 17:508–25

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

Nee S, May RM, Harvey PH. 1994. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B* 344:305–11

Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML. 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature* 345:714–16

Novotny V, Miller SE, Baje L, Balagawi S, Basset Y, et al. 2010. Guild-specific patterns of species richness and host specialization in plant-herbivore food webs from a tropical forest. *J. Anim. Ecol.* 79:1193–203

Ødegaard F. 2004. Species richness of phytophagous beetles in the tropical tree *Brosimum utile* (Moraceae): the effects of sampling strategy and the problem of tourists. *Ecol. Entomol.* 29:76–88

Pauly D, Palomares M-L. 2005. Fishing down marine food web: It is far more pervasive than we thought. *Bull. Mar. Sci.* 76:197–212

Pavoine S, Gasc A, Bonsall MB, Mason NWH. 2013. Correlations between phylogenetic and functional diversity: mathematical artefacts or true ecological and evolutionary processes? *J. Veg. Sci.* 24:781–93

Pelayo-Villamil P, Guisande C, Vari RP, Manjarrés-Hernández A, García-Roselló E, et al. 2015. Global diversity patterns of freshwater fishes—potential victims of their own success. *Divers. Distrib.* 21:345–56

Peterson AT, Soberón J, Pearson RG, Anderson RP, Martínez-Meyer E, et al. 2011. *Ecological Niches and Geographic Distributions*. Princeton, NJ: Princeton Univ. Press

Petrou ZI, Kosmidou V, Manakos I, Stathaki T, Adamo M, et al. 2014. A rule-based classification methodology to handle uncertainty in habitat mapping employing evidential reasoning and fuzzy logic. *Pattern Recognit. Lett.* 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. *Glob. Ecol. Biogeogr.* 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. *Syst. Entomol.* 35:18–32

Poelen JH, Simons JD, Mungall CJ. 2014. Global biotic interactions: an open infrastructure to share and analyze species-interaction datasets. *Ecol. Inform.* 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. *J. N. Am. Benthol. Soc.* 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. *J. Trop. Ecol.* 15:213–27

Poulin R. 2010. Network analysis shining light on parasite ecology and diversity. *Trends Parasitol.* 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. 2010. Who really ate the fruit? A novel approach to camera trapping for quantifying frugivory by ruminants. *Ecol. Res.* 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. *Mol. Phylogenet. Evol.* 61:543–83

Raffaelli D, Hall SJ. 1992. Compartments and predation in an estuarine food web. *J. Anim. Ecol.* 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. *Biol. Conserv.* 113:23–34

Rex MA, Etter RJ. 2010. *Deep-Sea Biodiversity: Pattern and Scale*. Cambridge, MA: Harvard Univ. Press

Rich TCG. 2006. Floristic changes in vascular plants in the British Isles: geographical and temporal variation in botanical activity 1836–1988. *Bot. J. Linn. Soc.* 152:303–30

Ricklefs RE. 2012. Species richness and morphological diversity of passerine birds. *PNAS* 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, UK: Wiley

Rocchini D, Hortal J, Lengyel S, Lobo JM, Jimenez-Valverde A, et al. 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. *Prog. Phys. Geogr.* 35:211–26

Rodrigues ASL, Andelman SJ, Bakarr MI, Boitani L, Brooks TM, et al. 2004. Effectiveness of the global protected area network in representing species diversity. *Nature* 428:640–43

Rodrigues ASL, Gray CL, Crowter BJ, Ewers RM, Stuart SN, et al. 2010. A global assessment of amphibian taxonomic effort and expertise. *Bioscience* 60:798–806

Rosen R. 1996. On the limitations of scientific knowledge. In *Boundaries and Barriers: On the Limits to Scientific Knowledge*, ed. JL Casti, A Karlqvist, pp. 199–214. Reading, MA: Perseus

Roskov Y, Abucay L, Orrell T, Nicolson D, Kunze T, et al. 2015. *Species 2000 & ITIS Catalogue of Life: 2015 Annual Checklist*. Leiden, Neth.: Catalogue of Life. **http://www.catalogueoflife.org/annual-checklist/2015**

Rovero F, Martin E, Rosa M, Ahumada JA, Spitale D. 2014. Estimating species richness and modelling habitat preferences of tropical forest mammals from camera trap data. *PLOS ONE* 9:e103300

Roy K, Foote M. 1997. Morphological approaches to measuring biodiversity. *Trends Ecol. Evol.* 12:277–81

Sánchez-Fernández D, Aragón P, Bilton DT, Lobo JM. 2012. Assessing the congruence of thermal niche estimations derived from distribution and physiological data. A test using diving beetles. *PLOS ONE* 7:e48163

Sansom RS, Randle E, Donoghue PC. 2015. Discriminating signal from noise in the fossil record of early vertebrates reveals cryptic evolutionary history. *Proc. R. Soc. B* 282:20142245

Sastre P, Lobo JM. 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. *Biol. Conserv.* 142:462–67

Schemske DW, Mittelbach GG, Cornell HV, Sobel JM, Roy K. 2009. Is there a latitudinal gradient in the importance of biotic interactions? *Annu. Rev. Ecol. Evol. Syst.* 40:245–69

Sechrest W, Brooks TM, da Fonseca GAB, Konstant WR, Mittermeier RA, et al. 2002. Hotspots and the conservation of evolutionary history. *PNAS* 99:2067–71

Slater GJ, Harmon LJ. 2013. Unifying fossils and phylogenies for comparative analyses of diversification and trait evolution. *Methods Ecol. Evol.* 4:699–702

Slatyer RA, Hirst M, Sexton JP. 2013. Niche breadth predicts geographical range size: a general ecological pattern. *Ecol. Lett.* 16:1104–14

Erratum

Soberón J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. *Ecol. Lett.* 10:1115–23

Soria-Auza RW, Kessler M. 2008. The influence of sampling intensity on the perception of the spatial distribution of tropical diversity and endemism: a case study of ferns from Bolivia. *Divers. Distrib.* 14:123–30

Sutherst R. 2014. Pest species distribution modelling: origins and lessons from history. *Biol. Invasions* 16:239–56

Swenson NG, Enquist BJ. 2009. Opposing assembly mechanisms in a Neotropical dry forest: implications for phylogenetic and functional community ecology. *Ecology* 90:2161–70

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. *Mol. Ecol.* 21:2045–50

Thompson JN. 1997. Conserving interaction biodiversity. In *The Ecological Basis of Conservation: Heterogeneity, Ecosystems, and Biodiversity*, ed. STA Pickett, RS Ostfeld, M Shachak, GE Likens, pp. 285–93. New York: Chapman & Hall

Thompson PL, Davies TJ, Gonzalez A. 2015. Ecosystem functions across trophic levels are linked to functional and phylogenetic diversity. *PLOS ONE* 10:e0117595

Thomson JR, MacNally R, Fleishman E, Horrocks G. 2007. Predicting bird species distributions in reconstructed landscapes. *Conserv. Biol.* 21:752–66

Traugott M, Kamenova S, Ruess L, Seeber J, Plantegenest M. 2013. Empirically characterising trophic networks: what emerging DNA-based methods, stable isotope and fatty acid analyses can offer. *Adv. Ecol. Res.* 49:177–224

Valiente-Banuet A, Aizen MA, Alcántara JM, Arroyo J, Cocucci A, et al. 2014. Beyond species loss: the extinction of ecological interactions in a changing world. *Funct. Ecol.* 29:299–307

Verdú JR, Alba-Tercedor J, Jiménez-Manrique M. 2012. Evidence of different thermoregulatory mechanisms between two sympatric *Scarabaeus* species using infrared thermography and micro-computer tomography. *PLOS ONE* 7:e33914

Villeger S, Miranda JR, Hernández DF, Mouillot D. 2010. Contrasting changes in taxonomic vs. functional diversity of tropical fish communities after habitat degradation. *Ecol. Appl.* 20:1512–22

Violle C, Navas ML, Vile D, Kazakou E, Fortunel C, et al. 2007. Let the concept of trait be functional! *Oikos* 116:882–92

Walther BA, Moore JL. 2005. The definitions of bias, precision, and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28:815–29

Warren DL, Cardillo M, Rosauer DF, Bolnick DI. 2014. Mistaking geography for biology: inferring processes from species distributions. *Trends Ecol. Evol.* 29:572–80

Whittaker RJ, Araújo MB, Jepson P, Ladle RJ, Watson JEM, Willis KJ. 2005. Conservation biogeography: assessment and prospect. *Divers. Distrib.* 11:3–23

Wisz MS, Pottier J, Kissling WD, Pellissier L, Lenoir J, et al. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol. Rev.* 88:15–30

Wolfe DA, Champ MA, Flemer DA, Mearns AJ. 1987. Long-term biological data sets: their role in research, monitoring, and management of estuarine and coastal marine systems. *Estuaries* 10:181–93

Yahara T, Donoghue M, Zardoya R, Faith DP, Cracraft J. 2010. Genetic diversity assessments in the century of genome science. *Curr. Opin. Environ. Sustain.* 2:43–49

Yodzis P. 1988. The indeterminacy of ecological interactions as perceived through perturbation experiments. *Ecology* 69:508–15