Opinion

Bias assessments to expand research harnessing biological collections

Emily K. Meineke [1,3,5,*,@] and Barnabas H. Daru [2,4,5,*,@]

Biological collections are arguably the most important resources for investigations into the impacts of human activities on biodiversity. However, the apparent opportunities presented by museum-derived datasets have not resulted in consistent or widespread use of specimens in ecology outside phenological research and species distribution modeling. We attribute this gap between opportunity and application to biases introduced by collectors, curators, and preservation practices and an imperfect understanding of these biases and how to mitigate them. To facilitate broader use of specimen-based data, we characterize collection biases across key axes and explore interactions among them. We then present a framework for determining the bias assessments needed when extracting data from biological collections. We show that bias assessments required by particular ecological studies will depend on the response variables being measured and the predictor axes of interest. We argue that quantification of biases in specimen-derived datasets is needed to facilitate the widespread application of these data.

Biases in biological collections data

Biological collections have emerged as a unique source of long-term data revealing biotic responses to global change. Although specimens of plants, birds, mammals, fungi, fish, and insects were originally collected for taxonomic research, ecologists increasingly harvest data from specimens to describe ecological processes, such as plant–insect interactions, host–parasite interactions, food web reconstruction, microbial community shifts, and changing species distributions [1,2]. The uniqueness of biological collections comes primarily from their multidimensionality; they capture ecological processes over decades to centuries – thus harboring signatures of anthropogenic change – and additionally capture unprecedented ecological data across geographic, phylogenetic, and climatic axes (e.g., [3]). However, despite the increasing digitization and accessibility of specimens, they are not widely mined for ecological research.

Although many factors are probably at work, we attribute the gap between accessibility and application in large part to the inherent biases associated with specimen collection and preservation that can complicate inferences and, in the worst-case scenarios, lead to spurious interpretations of specimen-based data. Our definition of bias refers to the deviation of data from a true value and can manifest in myriad ways and along multiple axes (Figure 1). In ecology, data collection is typically randomized or otherwise collected in ways that are thought to be minimally biased. By contrast, specimens are nonrandomly collected over space and time, representing biased subsets of true biodiversity and underlying processes. For instance, herbarium specimens are often used to estimate the timing of plant flowering [4], but bias assessments demonstrate that collectors tend to preferentially collect plants in peak flower [3]. This bias presents a potential benefit if the goal of a study is to investigate how the timing of peak flowering changes over years, but is potentially misleading if the goal is to document the beginning or end of seasonal flowering events [5]. Therefore, any inferences about, for instance, first or last flower from specimens should be avoided or made with caution.

[1]Department of Entomology and Nematology, University of California, Davis 95616, CA, USA
[2]Department of Life Sciences, Texas A&M University–Corpus Christi, Corpus Christi, TX 78412, USA
[3]https://meinekelab.com
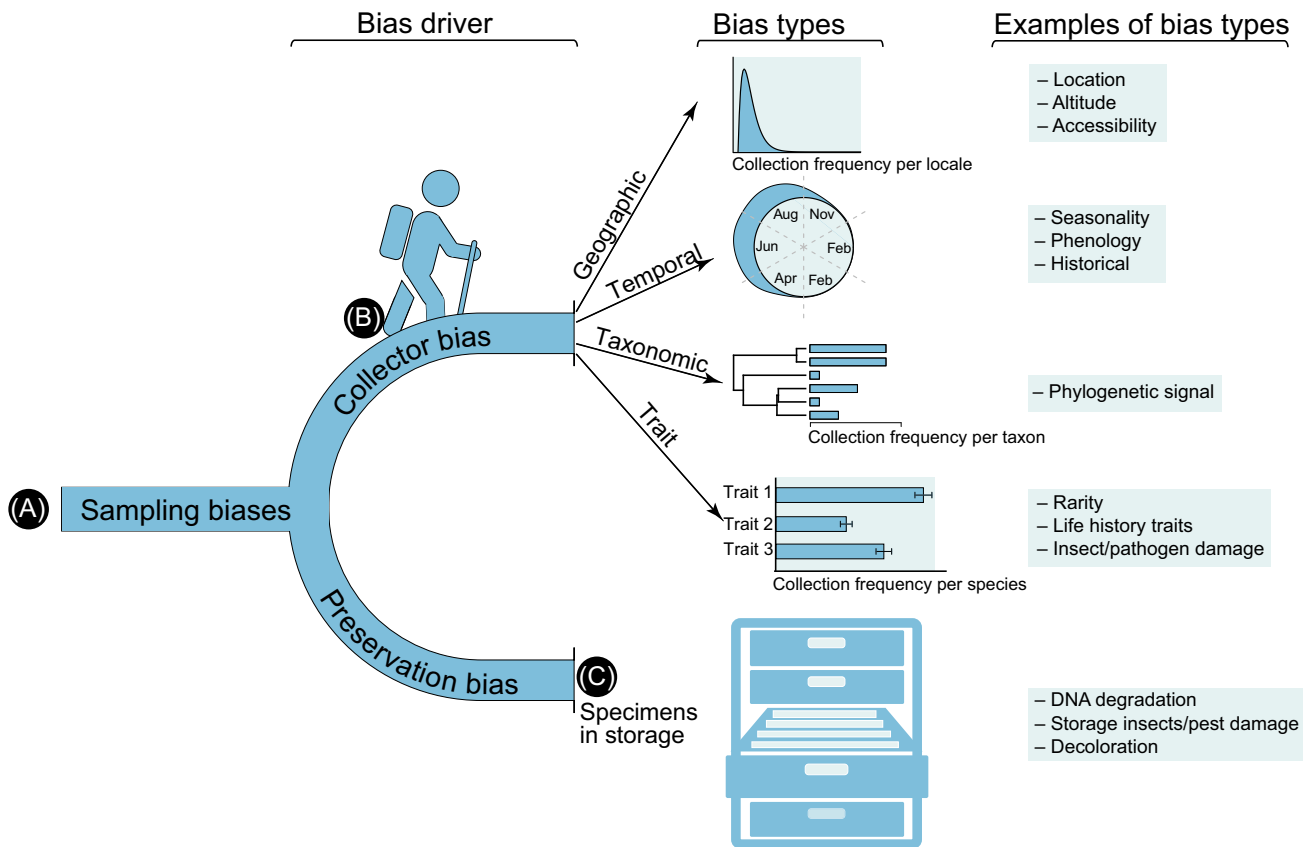[4]https://barnabasdaru.com
[5]These authors contributed equally to this work

*Correspondence:
ekmeineke@ucdavis.edu (E.K. Meineke) and barnabas.daru@tamucc.edu (B.H. Daru).
@Twitter: @emilymeineke (E.K. Meineke) and @Barnabas_Daru (B.H. Daru).

Figure 1. Categories of sampling biases represented in biological collections. (A) Biases can be broadly driven by (B) individual collectors, manifesting as geographic, temporal, taxonomic, or functional trait biases, or (C) preservation bias, in which the methods of preserving the specimens in storage can influence the utility of specimens in ecology. Examples of each type of bias are indicated.

Similar biases are not well characterized across the diverse data types and axes of interest represented by collections (Figure 1). A limited number of studies have characterized biases in aggregated datasets, such as the Global Biodiversity Information Facility [6,7] or digitized specimens from large museum collections [8]. Other studies have addressed biases within regional collections that may, or may not, represent biases in collections held in other museums (e.g., [9]). However, bias assessments are not as commonplace as expected given the potential impact of specimen-based data in ecology. Moreover, there is no consensus framework for assessing biases within datasets that are the focus of individual studies, for example, collections from particular locations, museums, or time periods, which may harbor distinct biases and thus require individualized bias assessments. Arguably the most straightforward method of bias assessment is to compare collections-based data against baseline data; for example, an assessment of biodiversity change over time in specimens versus observational data. However, for many response variables of interest, collections represent the most comprehensive datasets available, and baseline data to assess the accuracy of collections-based data do not exist. Thus, a key hurdle in the wide utilization of specimen data is to develop methods to assess and/or account for underlying biases, especially in the cases where baseline datasets are absent.

To facilitate broader application of biological collections in ecology, we synthesize existing bias assessments along five key dimensions (Figure 1). Specifically, we focus on collector- and

preservation-driven biases in biological collections from the past ~300 years of active scientific exploration. While we aim to address biases affecting diverse taxa, many examples are based on plants, which is reflective of our and much other ecological research that currently utilizes biological collections. Next, we present a paradigm to encourage ecologists to more fully harness museum specimens from the modern era to study global change for any taxonomic group at any scale.

Effects of biases on data interpretations for global change ecology

A wide range of data can be harvested from specimen labels. Perhaps the most common are species occurrences, which include the date and location where a specimen was collected. Data are additionally harvested from specimens directly, such as DNA, metabolites, endophytes, 3D morphological data, leaf herbivory, and phenology. Below, we characterize the axes along which biases in collections manifest and specific ways in which they may affect downstream ecological inferences with regard to diverse types of data (Box 1). We additionally suggest key areas where future bias assessments are needed.

Bias in geographic coverage

A bias in geographic coverage occurs when species are collected more within some regions of their ranges than in other regions. Geographic bias can limit investigations into species range shifts with climate change [10,11]. At the global scale, specimen data are likely to under-represent biodiversity in tropical areas [7]. Likewise, data accessibility and availability can introduce biases in collections in multiple dimensions. For instance, while a few countries in Europe as well as the USA, Australia, and Brazil have invested millions of dollars to digitize their holdings with the goal of making them accessible, extensive gaps remain in the vast majority of regions that harbor high concentrations of biodiversity, such as South America, Africa, India, and Southeast Asia. As a consequence, a mismatch between areas of high sampling intensity and accessibility versus areas of high species richness results.

At the regional or local extent, physical accessibility constitutes a major factor underlying the non-systematic collection of museum records. For instance, even in well-sampled regions of the world, most specimen collections occur around roads, airports, cities, harbors, and field stations [5,8,12], while private lands, deserts, Arctic regions, mountain tops, and thick rainforests are under-collected [13]. Collecting near physical structures can result in clustered geographic sampling of specimens [14], such that specimens tend to reflect the climate space of accessible areas (i.e., areas of human activities). These areas may harbor higher surface temperatures than more remote or inland areas [15,16], such that collections may represent warm, dry areas within a given region. Biases introduced by accessibility may thus reduce the value of specimens for investigations of responses to climate in natural populations and communities. At the same time, such biases may position specimens as a potential goldmine for studying the effects of urbanization over the long term [17], although they have to date been underused for this purpose.

Bias in taxonomic coverage

Taxonomic bias occurs when some, typically charismatic [18,19], organisms are collected more than others relative to their actual occurrence in nature [20]. Taxonomic bias can also manifest phylogenetically, such that closely related species tend to be collected more or less relative to chance. For instance, in South Africa, the genus *Protea* is overrepresented in collections compared with other groups, perhaps because of its horticultural appeal in the cut-flower industry [8]. These biases can impede ecological application by limiting the number of taxa to choose from when designing ecological studies. However, for most taxa, assessing what percentage of total taxon diversity is represented in collections is difficult or impossible, because our best

Box 1. Assessing trait biases across diverse taxa

We explored potential trait biases across a taxonomically diverse set of biological collections using a linear mixed-effects model of collection frequency (number of observed specimens collected per species) and functional traits innate to their ecology, morphology, or life history. These traits include geographic range size, body size, extinction risk, and 'evolutionary distinctiveness and global endangerment' (EDGE). The EDGE metric combines evolutionary distinctiveness (defined as the degree of phylogenetic isolation of a species) with extinction risk to identify threatened species whose extinction would lead to the highest losses across the Tree of Life [40]. We included taxonomic rank at the family level as a random effect in the model to account for phylogenetic nonindependence (Figure IA). Across clades, we found that species with medium to large geographic ranges were collected more frequently than narrow-ranged species. For instance, in birds, geographic range size offers a strong prediction of collection frequency (beta = 0.52, P < 0.001) for some species such as the song sparrow (*Melospiza melodia*), which is widely abundant and distributed throughout North America, represented by 28 352 specimens. This means that species with larger geographic ranges are more likely to be widely distributed across the landscape and thus more detectable, resulting in their higher collection frequencies. However, it is possible that rare species in some taxonomic groups might be overrepresented in collections not only because of their scientific value but because they are periodically assessed rather than more common species [41]. However, we show that at-risk species were collected less often for all groups (chondrichthyans: beta = −0.13; birds: beta = −0.008; mammals: beta = −0.088; reptiles: beta = −0.041; amphibians: beta = −0.081). To a greater extent, bird species with high EDGE scores were collected less often (beta = −0.074, P = 0.026) compared with other taxonomic groups. This is unsurprising given the widespread collecting restrictions on threatened species [18,42].

A common untested assertion is that collectors tend to avoid collecting very-large-sized specimens [43,44]. Consistent with this hypothesis, we found biases for the tendency to collect small- to medium-sized species in reptiles and amphibians. This could be explained by the fact that collectors prefer to collect specimens that would easily fit into a jar or portable container to keep field samples viable for transportation or due to space limitations in the collections [41]. Additional trait biases described in the literature are described in Figure IB.
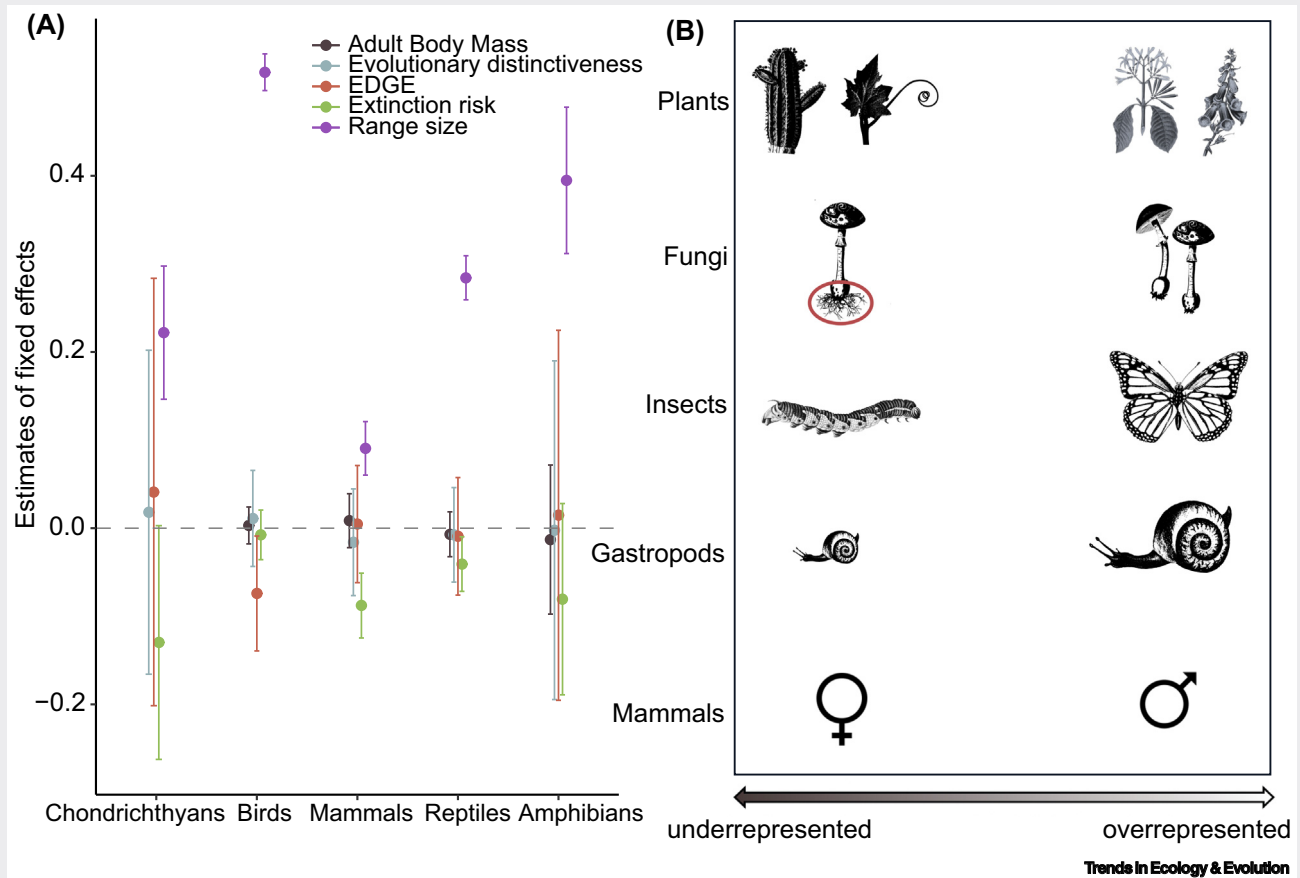
Figure I. Functional trait biases in diverse taxa. (A) The estimates with 95% confidence intervals from the explanatory models of species' collection frequency in chondrichthyans, birds, mammals, squamate reptiles, and amphibians. All explanatory variables were standardized by subtracting the mean and dividing by one standard deviation to allow comparison of effect sizes. For each taxonomic group, the number of observed specimens collected per species was modeled using linear mixed-effects models with adult body mass, geographic range size, evolutionary distinctiveness, extinction risk and evolutionary distinctiveness and global endangerment (EDGE) as fixed effects and taxonomic rank at the family level as a random effect to account for phylogenetic nonindependence. (B) From top to bottom, leafy plants with large flowers are overrepresented in herbarium collections while plants that are difficult to fit on an herbarium sheet like vines or ones that contain spines are underrepresented [45]. Fruiting bodies of fungi, often mushrooms, are more commonly collected than hyphae [46]. Anecdotally, adult insects are overrepresented compared with larvae, which, for holometabolous species, represent completely different ecosystem functions and life histories. Smaller, less charismatic snails are underrepresented [9]. For mammals, males are more commonly accessioned, although the pattern is the opposite for bats [47].

understanding of biodiversity is based on biological collections themselves. For heavily collected taxa, such as birds, taxonomic bias within the group can be assessed based on collection frequencies per genus or species. For instance, a high number of specimens collected per species would indicate taxonomic overrepresentation. For less charismatic taxonomic groups, the same methods can be applied, but the vast majority of species are represented by few and uncertain records [7], which can in turn reduce the utility for ecological applications, notably species distribution modeling [21,22].

Bias in temporal coverage

Temporal bias is the unbalanced collecting of specimens in some years or parts of the year. Within years, temporal biases can emerge as a result of targeted collection of life stages (e.g., an emphasis on collecting plants with reproductive tissues that allow identification). Among years, temporal biases can be driven by global events, such as reduced collection during World Wars I and II [7]. A major hurdle for harnessing collections to study global change is that collecting has either declined for many taxa in recent years or has been replaced by the mass production of observation-based occurrences that cannot be traced to a tangible specimen [23]. Such general patterns in collecting efforts can be understood by simply plotting the number of specimens collected within a given time period; thus, the capacity to understand these patterns grows as more collections are digitized.

Perhaps the most unique aspect of natural history collections from an ecological perspective is their long timespan, which can elucidate patterns and mechanisms underlying biological change as Earth's climate changes. However, determining whether changes in collection frequency of a given species result from collector behavior or true changes in species abundance can be challenging in some cases and impossible in others. Nevertheless, studies have now implemented statistical methods to account for temporal biases. For instance, bees are often collected using standardized methodology (e.g., sweep nets, pan traps); thus, the relative abundance of particular species can be determined from collections, with the total number of bees collected in a given year specified as weights in models [24]. However, caution must be taken to standardize by organisms collected using consistent techniques [25,26].

Bias in trait coverage

Trait bias is the disproportionate collection of species or genotypes on the basis of intrinsic life-history traits, including life cycle, sex, size, growth form, and/or rarity. Certain species are collected more than others as a result of their life-history traits. For instance, collectors may tend to avoid collecting spiny plants or those that otherwise lend themselves poorly to traditional preservation methods, (i.e., if leaves are too large for standard herbarium sheets or if fruits are large and cannot be included in pressed specimens). While such trait biases are anecdotally reported by curators and collectors, they are rarely quantified (see Box 2 for more details). Ultimately, the critical implication of collector-driven trait biases for studies of the effects of global change is that, whenever possible, conclusions must apply narrowly to the species, sexes, sizes, or other subgroups represented by the data. In some cases, this will require the revisiting of studies whose conclusions have been extrapolated, inadvertently, to encompass entire populations or species, which were not fully represented by the collections from which data were extracted. However, in some cases it may also be possible to distinguish *a priori* traits that are unlikely to be affected by collection biases. For example, proboscis lengths in bumblebee collections are now shorter than in the past, potentially in response to more generalized foraging strategies in response to wildflower decline [27] (but see [28]). Because bumblebee proboscises are not typically visible to collectors during the act of collection, it is

unlikely that this trait is subject to biases introduced by collector preferences that would negate or otherwise affect these conclusions.

Biases introduced through preservation practices

Most preservation techniques are borne out of tradition rather than best scientific practices for the preservation of specimen-based data long term (e.g., [29]) and thus can affect the amount and quality of usable ecological data derived from specimens. Preservation biases are tied to the specific preservation techniques applied, and bias assessments typically involve simply a comparison of fresh and preserved tissues (e.g., [30]). Data harvested at various timepoints after preservation are particularly useful, as they contribute additional information about how preservation biases may shift over time.

Preservation for many taxa is simply the process of drying, which can manifest biases in multiple dimensions. For instance, herbarium specimens display slightly different measured structural traits than freshly collected specimens, such as specific leaf area (SLA). However, SLAs of dried and fresh leaf tissue are highly correlated and thus herbarium specimens may be used to investigate changes in SLA over years, or other relevant axes, as long as a correction factor is applied to account for species-specific effects of preservation [31]. Isotope signatures – which can be extracted from specimens to describe an organism's trophic level and thus to reconstruct historical food webs – can be biased differentially by preservation techniques. Formalin can deplete $\delta^{13}C$, as can alcohol to a lesser extent [32]. Most effects of preservation on stable isotopes can be corrected for as they are properly characterized (e.g., [33]). Because preservation

---

Box 2. An example of bias quantification: herbivory captured in herbarium specimens

Here we describe an example of data collection from museum specimens following the steps outlined in Figure I. In this example, we demonstrate our suggested workflow for the extraction of data from museum specimens while considering biases that may affect inferences.

An application of the workflow is illustrated in Figure I. The case study comes from an assessment of insect herbivory data collected from herbarium specimens [38]. In the first step of the workflow (i.e., form hypothesis; Figure IB), the project's competing hypotheses (Figure IA) were that herbivory (second step; Figure IB) had increased during the 20th century due to warming winters in northeastern USA [48] or decreased due to widespread insect declines [49].

The third step requires determination of the data extraction methods (Figure IC). Here, the authors showed that biases in herbivory data from herbarium specimens had not been quantified in previous studies and no historical herbivory data were available for comparisons with the data collected from herbarium specimens from the focal species. Therefore, herbivory could not be compared in this study with baseline data, as indicated in the fifth step of the workflow (Figure IE). The expectation is that herbivory would be less abundant on specimens than in nature due to a bias toward the collection and accession of minimally damaged specimens. In addition, the authors acknowledged this down-bias as a limitation of the dataset, but did not expect it to affect patterns in herbivory over space and latitude, two key axes of interest. By contrast, if the goal is to determine biases *a priori* (Figure ID), biases toward minimally damaged leaves may be expected to shift with changing collection practices over time. To assess this possibility, the authors consulted with the curators at the Harvard University Herbaria (HUH). Although the curators at the HUH did not believe that collection or curation practices had changed substantially, as the institution has strong traditions, they however noted that more recent specimens tended to be collected by students rather than expert collectors. To account for this potential bias over years, the authors included 'collector' as a model covariate (Figure IF). A third concern was the potential for preservation bias introduced by chewing damage by herbivores, such as tobacco beetles, within herbaria (Figure ID). This bias would be introduced into the datasets only if it were indistinguishable from damage made by insects outdoors prior to specimen collection. However, the authors were able to distinguish between indoor and outdoor herbivory, thereby avoiding any effects of preservation bias on the data (Figure IF).

Statistical modeling indicated that herbivory had increased over time with rising winter temperatures. However, such results should be interpreted cautiously, with special attention paid to the above concern that collector and/or curator biases might have shifted over time. Consistent with the finding of increasing herbivory over time, the authors found that temperature is a key driver of herbivory by assessing herbivory across latitude, an axis across which temperature varies but collector and curator biases should not operate (Figure IF).
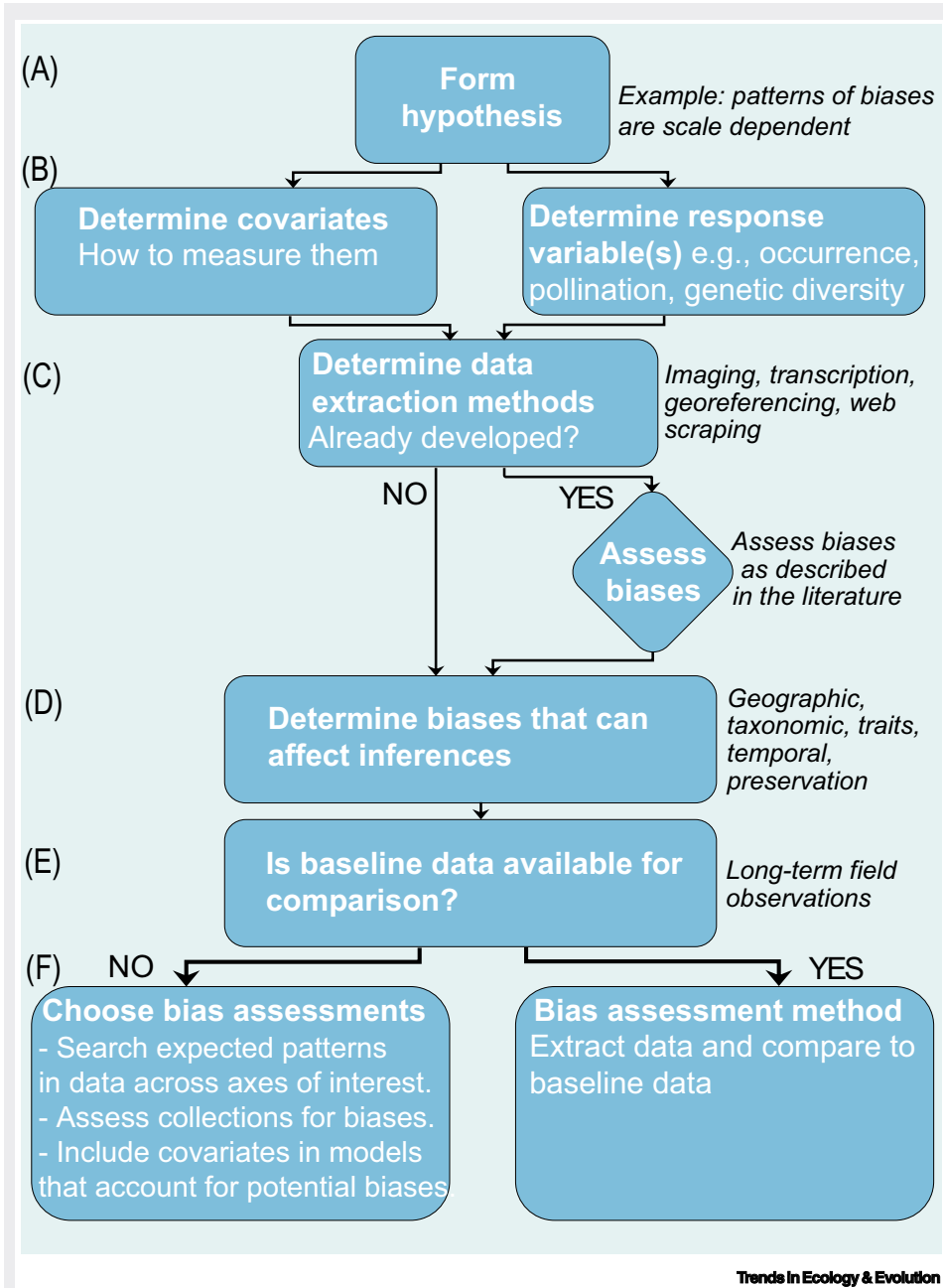
Figure I. Suggested workflow for research projects aiming to utilize museum specimen-based data. Relevant examples of each step appear on the righthand side. For (F), detailed descriptions of the specific bias assessment methods are included in the main text.

techniques differentially preserve isotopic signatures, care must be taken when comparing data from specimens that are preserved in alcohol versus formalin. Generally, if analyses include chemistry, care should be taken to determine how specimens were preserved to account for any effects of preservation bias.

Predicted associations between biases

Biases can arise across multiple axes simultaneously and thus co-occur in datasets. For instance, collectors may prefer certain traits, directly leading to taxonomic and/or phylogenetically biased data in which some clades are collected more thoroughly than others. Likewise, taxonomic and temporal biases might be tightly linked. If extinction risk or species' responses to global change are phylogenetically nonrandom, we expect some species to be collected more or less than expected over time because they are at-risk species or climate-change indicators, which could lead to biased phylogenetic patterns in the data. Biases in geographic coverage may vary with temporal bias because a lower accumulation of specimen records over time can severely compromise analyses such as distribution modeling or population demographic history [34,35]. Likewise, preservation biases may vary across a taxonomic axis because preservation techniques vary between taxa. Understanding potential links between biases may inform what bias assessments are needed in a given ecological study.

A framework and toolkit to assess biases within datasets

Here, we present a workflow that allows the evaluation of whether collections data may be appropriate for a given project or whether biases may impede hypothesis testing (see Figure I in Box 2). This workflow was implemented using an empirical example (Box 2) to determine the required *a priori* bias assessments. Here, we define various bias assessment tools in detail (Box 3).

Box 3. Computational and statistical tools to aid in bias assessments

Analysis of sampling biases can include multiple empirical challenges such as harmonizing primary data from different collectors with different collection practices or the use of computer graphics approaches to visualize patterns of sampling biases. Addressing these challenges would require access to computational tools to analyze the different types of specimen biases (Table I).

Geographic bias

The software that can be used to assess geographic bias include sampbias, phyloregion, BiotaPhy, dismo, sdm, ENMeval, biomod2, SDMtune, ssdm, esdm, ENMTML, HMMoce, and rgeolocate (Table I). Some packages are best suited for exploring biogeographic patterns with potential application to the mapping of collection intensity. Others were developed for direct assessment of geographic sampling biases in species distribution data, while several of the packages most commonly used for species distribution modeling can be extended to predict species abundances in poorly sampled regions (Table I).

Taxonomic bias

Analyses tailored exclusively for the assessment of taxonomic bias have not been implemented in software. However, several packages have been developed to standardize species names against referenced taxonomies (e.g., taxize, taxonstand, the Taxonomic Name Resolution Service) [50].

Researchers wishing to evaluate more complex hypotheses such as the tendency of closely related species to be collected similarly must often use an *ad hoc* collection of disparate phylogenetic tools such as ape, caper, and phangorn (Table I).

Temporal bias

Analysis of temporal bias can be complicated by several factors, such as time zones, leap days, and differences in date format across regions. The R package lubridate facilitates the analysis of dates, times, and time spans by providing user-friendly tools for parsing date-time data, extraction and, algebraic manipulation on date-time and time-span objects [51]. For complex analyses, such as testing whether specimen collection dates were randomly distributed against all dates spanning the entire duration of specimen collection, researchers can explore the function rayleigh. test in the R package circular, which performs a Rayleigh test of uniformity by assessing the significance of the mean resultant length (https://r-forge.r-project.org/projects/circular).

Trait bias

The software package TR8 [52], for instance, allows the retrieval of plant functional traits from freely accessible repositories including BiolFlor and LEDA Traitbase. If the goal is to contrast the abundance of a taxon if specimen collection was equal across all taxa for each trait category versus the number of observed specimens collected per taxon, the software tool cati [53] can achieve this objective.

Table I. Computational tools for assessing bias

Type of bias	Software package	Description	Refs
Geographic	sampbias	A software package for quantifying and visualizing geographic sampling biases in species distribution data	[13]
	phyloregion	Biogeographic regionalization and analyses of macroecology	[54]
	BiotaPhy	Facilitates integration, data collection, and analysis by connecting to existing data repositories such as the Open Tree of Life, iDigBio, and Lifemapper	https://biotaphy.github.io/
	dismo	Methods for species distribution modeling; this predicts the environmental similarity of any site to that of the locations of known occurrences of a species	http://cran.r-project.org/web/packages/dismo/index.html
	sdm	Develop species distribution models (SDMs) using individual and community-based approaches, generate ensembles of models, evaluate the models, and predict species' potential distributions in space and time	[55]
	ENMeval	Assessment of spatially independent evaluations and estimation of optimal model complexity for Maxent ecological niche models	[56]
	biomod2	Species distribution modeling, calibration, and evaluation, ensemble of models, ensemble forecasting, and visualization	[57]
	SDMtune	Tuning and evaluation of SDMs	[58]
	ssdm	Map species richness and endemism based on stacked SDMs	[59]
	esdm	Create and evaluate ensembles of SDM predictions; it can also be used to identify spatial uncertainties and make informed conservation and management decisions	[60]
	ENMTML	Integrated construction of ecological niche models	[61]
	HMMoce	Geolocation of archive-tagged fishes using a hidden Markov method	[62]
	rgeolocate	Taking IP addresses and geolocating them to country, city, time zone, and other geographic ranges	https://cran.r-project.org/web/packages/rgeolocate/rgeolocate.pdf
Taxonomic	taxize	Taxonomic search and retrieval from referenced taxonomies	[63]
	WorldFlora	Exaction and fuzzy matching of plant names against the World Flora Online taxonomic backbone data	[64]
	rotl	Retrieve phylogenetic trees, information about studies used to assemble the synthetic tree, and utilities to match taxonomic names to 'Open Tree identifiers'	[65]
	taxonstand	Standardize species names in vegetation databases	[66]
	ape	Analyses of phylogenetics and evolution	[67]

Table I. (continued)

Type of bias	Software package	Description	Refs
Temporal	lubridate	Analysis of dates and times	[51]
	circular	Performs circular statistics	(https://r-forge.r-project.org/projects/circular)
Trait	cati	Assessments of functional traits to detect and quantify multilevel community assembly processes	[53]
	FD	Measurement of functional diversity from multiple traits and other tools for functional ecology	[68]
	TR8	Retrieve plant species traits	[52]

Comparisons with baseline data

If baseline data are available, this is often the most informative bias assessment method. Two scenarios are possible when comparing specimen-based datasets with other datasets: (i) determine what biases might be harbored in both datasets; or (ii) in the rare cases where unbiased data are available, quantify biases in collections-based data. As an example of scenario (i), Davis *et al.* [3] contrasted data describing flowering phenology derived from herbarium specimens against comparable data documented from direct field observation in the 1800s and then again in the early 2000s. These two disparate datasets showed complementary sampling biases – the observational baseline data were available for only two timepoints and thus documented potentially spurious changes in flowering over years – but the herbarium-derived flowering data captured only peak flowering, not first-flowering events that are often of interest to ecologists studying plant phenology. In another example, Kozlov *et al.* [36] compared herbivory estimates from herbarium specimens and randomized samples and found that collectors may manifest biases in herbivory captured on herbarium specimens by collecting leaves that are minimally damaged. However, this study also represents a cautionary example of why robust sample sizes are needed to develop baseline datasets [37] and underscores the need for statistical tools that assess the sample sizes needed to achieve accurate estimates, such as power analyses and rarefaction curves.

Investigate expected patterns

This method allows the user to assess the overall data accuracy, but not necessarily the accuracy of collections-based data across the axis of interest. Nevertheless, it is a useful way to determine when to use caution and when to proceed with a certain level of confidence. For instance, based on our prior field research in seasonal climates (Box 2), we predicted that herbarium specimens collected earlier in the growing season would display lower levels of herbivory than those collected later in the year, when they have had more time to accumulate insect damage. This pattern has been present in every herbivory dataset we have collected from specimens [2,38,39] and remains a key avenue to assess data reliability.

Assess collections for expected biases *a priori* and harvest data from unbiased subsets

Biases may vary across taxa, regions, and time. For instance, in Box 2, insects in museums tend to eat and damage certain plant species and not others, creating preservation biases that vary across species. For species that are not frequently eaten in storage in museums, it is much easier to quantify herbivory from the wild before plant collection. Additionally, insect outbreaks in museums tend to be focused in certain parts of the collections and may thus tend to affect only certain taxa if the specimens are arranged phylogenetically or alphabetically by name. To avoid effects of biases across any axis described above, it may be possible to focus on taxon, time

period, or trait-based subsets that are not affected by biases that are typical within whole collections or aggregated datasets, such as the Global Biodiversity Information Facility.

Incorporate covariates in models

Covariates that account for spatial, temporal, or phylogenetic information in data can be implemented to improve explanatory power and thus eliminate the need to quantify biases, especially when baseline data are not available and thus bias quantifications are impossible (see Box 2 for an example).

Concluding remarks

Myriad investigations now provide convincing evidence that collections records can provide unique data poised to revolutionize our understanding of species' responses to global change. However, many questions remain about the biases in collections data and how they affect the accuracy of data interpretations (see Outstanding questions). Here, we outline a framework to guide the bias assessments needed to approach collections-based data of any type or scale. Bias assessments from aggregated datasets are needed, as are more bias assessments within individual studies. To fuel these efforts, we suggest investment in high-impact bias assessments from funders such as the US National Science Foundation's iDigBio or comparable organizations worldwide. Existing programs are not sufficiently targeted to the massive task of translating digitized collections into a central resource for ecology. Investment is needed to characterize biases in aggregated datasets, such as the Global Biodiversity Information Facility, and in specific subsets of collections-based data tailored for particular studies; for instance, datasets from individual museum collections or regions. These investments would promote confidence in biological collections data, and, where appropriate, reliance on other tools. Given the multidimensional scale of collection-based datasets, we believe that these investments will result in unprecedented advances in the study of biodiversity as it continues to shift over time.

Declaration of interests

No interests are declared.

References

1. Heberling, J.M. and Isaac, B.L. (2017) Herbarium specimens as exaptations: new uses for old collections. *Am. J. Bot.* 104, 963–965
2. Meineke, E.K. *et al.* (2018) The unrealized potential of herbaria for global change biology. *Ecol. Monogr.* 88, 505–525
3. Davis, C.C. *et al.* (2015) Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. *Am. J. Bot.* 102, 1599–1609
4. Willis, C.G. *et al.* (2017) Old plants, new tricks: phenological research using herbarium specimens. *Trends Ecol. Evol.* 32, 531–546
5. Daru, B.H. *et al.* (2019) Temperature controls phenology in continuously flowering *Protea* species of subtropical Africa. *Appl. Plant Sci.* 7, e01232
6. Rock, B.M. and Daru, B.H. (2021) Impediments to understanding seagrasses' response to global change. *Front. Mar. Sci.* 8, 608867
7. Meyer, C. *et al.* (2016) Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecol. Lett.* 19, 992–1006
8. Daru, B.H. *et al.* (2018) Widespread sampling biases in herbaria revealed from large-scale digitization. *New Phytol.* 217, 939–955
9. Nekola, J.C. *et al.* (2019) *Caveat consumptor notitia museo*: let the museum data user beware. *Glob. Ecol. Biogeogr.* 28, 1722–1734
10. Koh, I. *et al.* (2016) Modeling the status, trends, and impacts of wild bee abundance in the United States. *Proc. Natl. Acad. Sci. U. S. A.* 113, 140–145
11. Jamieson, M.A. *et al.* (2019) Geographic biases in bee research limits understanding of species distribution and response to anthropogenic disturbance. *Front. Ecol. Evol.* 7, 194
12. Engemann, K. *et al.* (2015) Limited sampling hampers "big data" estimation of species richness in a tropical biodiversity hotspot. *Ecol. Evol.* 5, 807–820
13. Zizka, A. *et al.* (2021) sampbias, a method for quantifying geographic sampling biases in species distribution data. *Ecography* 44, 25–32
14. Panchen, Z.A. *et al.* (2019) Patterns and biases in an Arctic herbarium specimen collection: implications for phenological research. *Appl. Plant Sci.* 7, e01229
15. Oke, T.R. (1973) City size and the urban heat island. *Atmos. Environ.* 7, 769–779
16. Kim, S.W. and Brown, R.D. (2021) Urban heat island (UHI) intensity and magnitude estimations: a systematic literature review. *Sci. Total Environ.* 779, 146389
17. Shultz, A.J. *et al.* (2020) Natural history collections are critical resources for contemporary and future studies of urban evolution. *Evol. Appl.* 14, 233–247

Outstanding questions

To what extent do museum records reflect our actual knowledge of biodiversity and shifts in biodiversity over space and time?

Do cultural forces or collector/curator proclivities introduce biases that are specific to collections, regions, time periods, and/or taxa?

If collection events are more common near human structures, are biological collections biased climatically (e.g., do they represent warmer than average regional climates)?

Within individual taxa and response variables, what are the *a priori* expectations of the resulting data that can guide a better understanding of data reliability?

To alleviate biases where possible, what time periods, taxa, and regions should be targeted in future collecting efforts?

To what extent can biases be alleviated by the development of standardized collecting protocols?

18. Donaldson, M.R. *et al*. (2016) Taxonomic bias and international biodiversity conservation research. *Facets* 1, 105–113
19. Troudet, J. *et al*. (2017) Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* 7, 9132
20. dos Santos, J.W. *et al*. (2020) Drivers of taxonomic bias in conservation research: a global analysis of terrestrial mammals. *Anim. Conserv.* 23, 679–688
21. Feng, X. *et al*. (2019) A checklist for maximizing reproducibility of ecological niche models. *Nat. Ecol. Evol.* 3, 1382–1395
22. Zurell, D. *et al*. (2020) A standard protocol for reporting species distribution models. *Ecography* 43, 1261–1277
23. Troudet, J. *et al*. (2018) The increasing disconnection of primary biodiversity data from specimens: how does it happen and how to handle it? *Syst. Biol.* 67, 1110–1119
24. Bartomeus, I. *et al*. (2013) Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc. Natl. Acad. Sci. U. S. A.* 110, 4656–4660
25. Boyle, J.H. *et al*. (2019) Monarch butterfly and milkweed declines substantially predate the use of genetically modified crops. *Proc. Natl. Acad. Sci. U. S. A.* 116, 3006–3011
26. Wepprich, T. (2019) Monarch butterfly trends are sensitive to unexamined changes in museum collections over time. *Proc. Natl. Acad. Sci. U. S. A.* 116, 13742–13744
27. Miller-Struttmann, N.E. *et al*. (2015) Functional mismatch in a bumble bee pollination mutualism under climate change. *Science* 349, 1541–5144
28. de Keyzer, C.W. *et al*. (2016) Delving deeper: questioning the decline of long-tongued bumble bees, long-tubed flowers and their mutualisms with climate change. *J. Pollinat. Ecol.* 18, 36–42
29. Simmons, J.E. (2002) *Herpetological collecting and collections management*, Society for the Study of Amphibians and Reptiles
30. Daru, B.H. *et al*. (2019) A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374, 20170395
31. Perez, T.M. *et al*. (2020) Herbarium-based measurements reliably estimate three functional traits. *Am. J. Bot.* 107, 1457–1464
32. Sarakinos, H.C. *et al*. (2002) A synthesis of tissue-preservation effects on carbon and nitrogen stable isotope signatures. *Can. J. Zool.* 80, 381–387
33. Storm-Suke, A. *et al*. (2007) Effects of formalin and ethanol preservation on otolith δ^{18}O stable isotope signatures. *Rapid Commun. Mass Spectrom.* 21, 503–508
34. Jiménez-Alfaro, B. *et al*. (2018) Modelling the distribution and compositional variation of plant communities at the continental scale. *Divers. Distrib.* 24, 978–990
35. Migliore, J. *et al*. (2018) Surviving glaciations in the Mediterranean region: an alternative to the long-term refugia hypothesis. *Bot. J. Linn. Soc.* 187, 537–549
36. Kozlov, M.V. *et al*. (2020) Biases in estimation of insect herbivory from herbarium specimens. *Sci. Rep.* 10, 12298
37. Meineke, E.K. *et al*. (2020) Response to Kozlov *et al.*: inaccurate estimation of biases in herbarium specimen data. *bioRxiv* Published online September 3, 2020. https://doi.org/10.1101/2020.09.01.278606
38. Meineke, E.K. *et al*. (2019) Herbarium specimens reveal increasing herbivory over the past century. *J. Ecol.* 107, 105–117
39. Meineke, E.K. *et al*. (2021) Phenological sensitivity to temperature mediates herbivory. *Glob. Chang. Biol.* 27, 2315–2357
40. Isaac, N.J. *et al*. (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS One* 2, e296
41. National Academies of Sciences, Engineering, and Medicine (2020) *Biological Collections: Ensuring Critical Research and Education for the 21st Century*, National Academies Press
42. Martín-López, B. *et al*. (2009) What drives policy decision-making related to species conservation? *Biol. Conserv.* 5, cox043
43. Zopfi, H.J. (1993) Ecotypic variation in *Rhinanthus alectorolophus* (Scopoli) Pollich (Scrophulariaceae) in relation to grassland management: I. Morphological delimitations and habitats of seasonal ecotypes. *Flora* 188, 15–39
44. Williams, P.H. *et al*. (2002) Data requirements and data sources for biodiversity priority area selection. *J. Biosci.* 27, 327–338
45. De Groot, S.J. (2011) Collecting and processing cacti into herbarium specimens, using ethanol and other methods. *Syst. Bot.* 36, 981–989
46. Andrew, C. *et al*. (2019) Fungarium specimens: a largely untapped source in global change biology and beyond. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 374, 20170392
47. Gower, G. *et al*. (2019) Widespread male sex bias in mammal fossil and museum collections. *Proc. Natl. Acad. Sci. U. S. A.* 116, 19019–19024
48. Intergovernmental Panel on Climate Change, ed (2014) *Climate change 2014: synthesis report*, IPCC
49. van Klink, R. *et al*. (2020) Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science* 368, 417–420
50. Boyle, B. *et al*. (2013) The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinforma.* 14, 16
51. Grolemund, G. and Wickham, H. (2011) Dates and times made easy with lubridate. *J. Stat. Softw.* Published online April 7, 2011. https://doi.org/10.18637/jss.v040.i03
52. Bocci, G. (2015) TR8: an R package for easily retrieving plant species traits. *Methods Ecol. Evol.* 6, 347–350
53. Taudiere, A. and Violle, C. (2016) cati: an R package using functional traits to detect and quantify multi-level community assembly processes. *Ecography* 39, 699–708
54. Daru, B.H. *et al*. (2020) phyloregion: R package for biogeographical regionalization and macroecology. *Methods Ecol. Evol.* 11, 1483–1491
55. Naimi, B. and Araújo, M.B. (2016) sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39, 368–375
56. Muscarella, R. *et al*. (2014) ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol. Evol.* 5, 1198–1205
57. Thuiller, W. *et al*. (2016) Package 'biomod2'. Species distribution modeling within an ensemble forecasting framework. *Ecography* 32, 369–373
58. Vignali, S. *et al*. (2020) SDMtune: an R package to tune and evaluate species distribution models. *Ecol. Evol.* 10, 11488–11506
59. Schmitt, S. *et al*. (2017) ssdm: an R package to predict distribution of species richness and composition based on stacked species distribution models. *Methods Ecol. Evol.* 8, 1795–1803
60. Woodman, S.W. *et al*. (2019) eSDM: a tool for creating and exploring ensembles of predictions from species distribution and abundance models. *Methods Ecol. Evol.* 10, 1923–1933
61. de Andrade, A.F.A. *et al*. (2020) ENMTML: an R package for a straightforward construction of complex ecological niche models. *Environ. Model. Softw.* 125, 104615
62. Braun, C.D. *et al*. (2018) HMMoce: an R package for improved geolocation of archival-tagged fishes using a hidden Markov method. *Methods Ecol. Evol.* 9, 1212–1220
63. Chamberlain, S.A. and Szöcs, E. (2013) taxize: taxonomic search and retrieval in R. *F1000Res.* 1, 291
64. Kindt, R. (2020) WorldFlora: an R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone. *Appl. Plant Sci.* 8, e11388
65. Michonneau, F. *et al*. (2016) rotl: an R package to interact with the Open Tree of Life data. *Methods Ecol. Evol.* 7, 1476–1481
66. Cayuela, L. *et al*. (2012) taxonstand: an R package for species names standardisation in vegetation databases. *Methods Ecol. Evol.* 3, 1078–1083
67. Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528
68. Laliberté, E. and Legendre, P. (2010) A distance-based framework for measuring functional diversity from multiple traits. *Ecology* 91, 299–305