

Biological science  
practices**Cite this article:** Nakamura G *et al.* 2025 The hidden biodiversity knowledge split in biological collections. *Proc. R. Soc. B* **292**: 20251045.<https://doi.org/10.1098/rspb.2025.1045>

Received: 19 April 2025

Accepted: 18 September 2025

**Subject Category:**

Biological applications

**Subject Areas:**

taxonomy and systematics, ecology, biological applications

**Keywords:**

name bearers, ichthyological collections, reference material, taxonomy, colonial legacy, knowledge distribution

**Author for correspondence:**

Gabriel Nakamura

e-mail: [gabriel.nakamura.souza@gmail.com](mailto:gabriel.nakamura.souza@gmail.com)

†Joint first authors.

‡Joint senior authors.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.8063164>.

## The hidden biodiversity knowledge split in biological collections

Gabriel Nakamura<sup>1,†</sup>, Bruno Henrique Mioto Stabile<sup>2,†</sup>, Livia Estéfane Fernandes Frateles<sup>3,†</sup>, Matheus Lima Araujo<sup>4</sup>, Emanuel Bruno Neuhaus<sup>5</sup>, Manoela Maria Ferreira Marinho<sup>6</sup>, Melina de Souza Leite<sup>7</sup>, Aline Richter<sup>8</sup>, Liuyong Ding<sup>9,10</sup>, Tiago Magalhães da Silva Freitas<sup>11</sup>, Bruno Soares<sup>12</sup>, Weferson Júnio da Graça<sup>13</sup>, Mario R. Moura<sup>14,15,‡</sup> and José Alexandre Felizola Diniz-Filho<sup>16,‡</sup><sup>1</sup>Department of Genetics, Ecology and Evolution, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil<sup>2</sup>Universidade Estadual de Maringá, Maringá, Paraná, Brazil<sup>3</sup>Programa de pós-graduação em Ecologia e Evolução, Universidade Federal de Goiás, Goiânia, Goiás, Brazil<sup>4</sup>Universidade Federal de Goiás, Goiânia, Brazil<sup>5</sup>Instituto Tecnológico Vale Desenvolvimento Sustentável, Belém, Pará, Brazil<sup>6</sup>Biosciences Institute, Universidade Federal de Mato Grosso do Sul, Campo Grande, Mato Grosso do Sul, Brazil<sup>7</sup>Departament of Ecology, Universidade de Sao Paulo, São Paulo, Brazil<sup>8</sup>Centro Nacional de Conservação de Répteis e Anfíbios, Instituto Chico Mendes de Conservação da Biodiversidade, Brasília, DF, Brazil<sup>9</sup>Yunnan Key Laboratory International Rivers and Transboundary Eco-Security, and <sup>10</sup>Institute of International Rivers and Eco-security, Yunnan University, Kunming, Yunnan, People's Republic of China<sup>11</sup>Universidade Federal do Pará—Campus Breves, Breves, Brazil<sup>12</sup>Institute of Environmental Change & Soci, University of Regina, Regina, UK<sup>13</sup>Universidade Estadual de Maringá, Maringá, Paraná, Brazil<sup>14</sup>Departamento de Sistemática e Ecologia, Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil<sup>15</sup>Departamento de Biologia Animal, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil<sup>16</sup>Departamento de Ecologia, ICB, Universidade Federal de Goiás, Goiânia, Goiás, Brazil

GN, 0000-0002-5144-5312; BHMS, 0000-0002-3524-537X; LEFF, 0000-0003-0644-5425; MLA, 0000-0002-9111-725X; EBN, 0000-0002-5966-7585; MMFM, 0000-0003-4226-666X; MdSL, 0000-0003-0505-0667; AR, 0000-0001-6385-153X; LD, 0000-0002-5490-182X; TMdSF, 0000-0003-3203-3408; BS, 0000-0001-5678-0403; WJdaG, 0000-0002-8925-5629; MRM, 0000-0002-7369-7502; JAFD-F, 0000-0002-0967-9684

Ecological and evolutionary processes generate biodiversity, yet how biodiversity data are organized and shared globally can shape our understanding of these processes. We show that name-bearing type specimens—the primary reference for species identity—of all freshwater and brackish fish species are predominantly housed in Global North museums, disconnected from their countries of origin. This geographical divide creates a ‘knowledge split’ with consequences for biodiversity science, particularly in the Global South, where researchers face barriers in studying native species’ name bearers housed abroad. Meanwhile, Global North collections remain flooded with non-native name bearers. We relate this imbalance to historical and socioeconomic factors, which ultimately restrict access to critical taxonomic reference materials and hinder global species documentation. To address this disparity, we call for international initiatives to promote fairer access to biological knowledge, including specimen repatriation, improved accessibility protocols for researchers in countries where specimens originated and inclusive research partnerships.

## 1. Introduction

Museums can be viewed as gatekeepers of representative fragments of the world. They provide a valuable source and representation of human history,

**Box 1.** A glimpse into taxonomic nomenclature

Name-bearing types (or name bearers) are specimens, or sets of specimens, designated as the standard reference for applying the name they bear. Each nominal taxon has an actual or potential name bearer. The objective of this nomenclatural act provided and regulated by the International Commission on Zoological Nomenclature (ICZN) is to fix a name in an international standard of reference, bringing stability to the name usage of a nominal taxon [5]. Four distinct types of name bearer are applied to fish in different cases, as shown in the figure below.

## Holotype



When in the original publication of a new taxon, a single specimen is picked to be the **name bearer**. The remaining specimens of the type series are **paratypes** and they do not have **name bearer** function. There are some exceptions, for example, modular organisms.

## Syntype



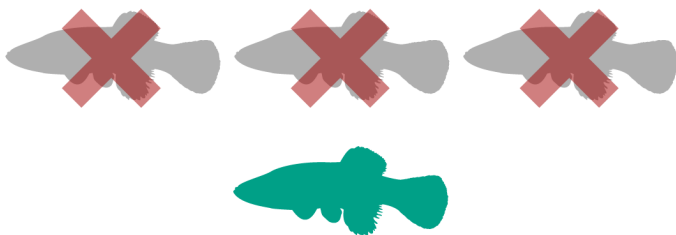
When in the original publication of a new taxon, the author(s) did not fix a **holotype**. In this case, all specimens listed in the type series are collectively the **name bearers**.

## Lectotype



When one specimen among **syntypes** is chosen to be the sole **name bearer** of its nominal taxon. The remaining specimens become **paralectotypes** and do not have **name bearer** function.

## Neotype



If a taxon does not have any **name bearer** specimen anymore (i.e. **holotype**, **syntype**, **lectotype**, or prior **neotype**), then another specimen that matches with the original description and type locality may become a **name bearer** specimen. However, such action is only recommended with the express purpose of solving taxonomic issues under conditions specified by the ICZN.

culture, knowledge and biodiversity. Natural history museums and biological collections (hereafter referred to as biological collections) are known for maintaining an organized and curated record of extant and extinct specimens. Given this role of biological collections, numerous authors have highlighted various ways in which they enhance our understanding of the natural world by shedding light on global change phenomena and spatial and temporal ecological processes [1], public health [2] and educational purposes [2]. Despite their many applications, the primary role of biological collections remains their significance in taxonomic research—particularly in a context in which approximately 80% of extant species have yet to be formally described [3,4]. The discovery of these species often depends on comparing potential new taxa with representative specimens of already-known species preserved in biological collections.

Among the specimens in biological collections, the name-bearing types (hereafter 'name bearers') hold special importance in biodiversity studies (box 1). Name bearers are pivotal in taxonomic studies, as they constitute the fundamental reference on which taxonomists rely to review described species or propose new ones [6,7]. The accessibility of name bearers is therefore crucial, as the most accurate assessment of certain taxonomically relevant features requires direct examination of the original reference specimen. While the digitization of name bearers and modern molecular techniques for species identification (e.g.

DNA barcoding) are undeniably important for biodiversity studies [8], the original material cannot be replaced solely by indirect data or molecular information for most organisms [9]. For example, around 562 tetrapod species have not been observed in over 50 years, and at least 40% are known only from the holotype [10]. The irreplaceable nature of these holotypes underscores their importance for conservation and taxonomic research, as their absence in biological collections cannot easily be compensated for through field observations.

Recently, palaeontology has drawn the scientific community's attention to the importance of name bearer distribution in biological collections (e.g. see the *Ubirajara* case [11]). Given the rarity of palaeontological records that can be used to identify a species, this unique sample often serves as the name bearer for a given species. In this context, recent studies have shown that a biased concentration of fossil samples in biological collections located in the Global North (countries with developed economies and a significant influence on global environmental governance, often located in the Northern Hemisphere) profoundly affects our view of deep-time biodiversity [12]. Beyond the impacts on our understanding of nature, this knowledge concentration also jeopardizes educational initiatives and scientific development in certain nations, as much of the research effort and public interest in biological collections depends on the presence of these materials [13]. While biases in name bearer distribution in palaeontology are receiving more attention from the scientific community and the public, similar issues extend to extant organisms, producing significant consequences for biodiversity science—especially in the context of accelerated biodiversity loss [14]—with more negative effects experienced by countries from the Global South.

Like other forms of biological data, the representation, curation and concentration of name bearers in biological collections worldwide are subject to biases, which can hinder accessibility—particularly for researchers based in countries where most of their native species are housed abroad. Taxonomists frequently report that many name bearers, particularly from the tropics, were collected and sent abroad, mainly to biological collections in Europe and North America during the late eighteenth and nineteenth centuries [15,16]. For instance, less than 18% of name-bearing reptile types are housed in Global South collections [17], despite this region harbouring 81% of reptile species [18]. Moreover, the number of plant species preserved in biological collections is twice as high in some Global North countries [19]. This colonial legacy in species name bearers can hinder efforts to accurately catalogue and organize knowledge of biological diversity [20]. Since the concentration of name bearers in certain countries reflects historical socioeconomic dynamics, a country's economic power has played—and continues to play—a crucial role in shaping the accumulation of biodiversity collections.

We argue that a fairer distribution of name bearers is necessary to close biodiversity knowledge gaps. In this study, we mapped the origin and destination of name bearers over time to identify macropolitical imprints in biodiversity knowledge accumulation. This study explores how biodiversity knowledge accumulates and shapes our perception of the natural world. We focused on freshwater and brackish fishes, the most species-rich group of vertebrates, which face challenges from uncertainties in their evolutionary relationships and ongoing taxonomic changes, making name bearer assessment crucial for improving accessibility of biodiversity knowledge [21]. We address the following questions. Where are name bearers located globally? How have countries accumulated biodiversity knowledge over time? Ultimately, we provide solutions and best practices to address these questions.

## 2. Methods

### (a) Data acquisition

We compiled all freshwater and brackish fish species names using the Eschmeyer's Catalog of Fishes (CAS) [22], the most updated curated taxonomic database for fishes. We extracted 15 303 valid names and 4943 synonyms. We included the synonyms because they also correspond to reference specimens in biological collections recognized as name-bearing types. Additionally, they are essential for conducting taxonomic reviews of species that were previously classified under these names. This resulted in 20 246 binomials that we called name bearers, following the definition of the International Code of Zoological Nomenclature [5]. For each name bearer, we extracted the country where the specimen (or specimens in case of syntype) was collected (i.e. the source country) and the museum housing the respective name bearer. Based on the abbreviations of the biological collections in CAS, we obtained the country where each museum/biological collection is located. If the source country of the name bearer was not specified, we checked CAS for additional information, such as coordinates or any other geographical location. We excluded marine species from our database due to the challenges associated with defining their distribution within political boundaries. All the searches in CAS were performed between March 2023 and May 2024.

### (b) Spatial characterization of name bearers across time

We mapped the country where each name bearer was collected (source country) and its current housing country. With this information, we characterized the flow of name bearers in 50-year intervals, grouping the countries according to the World Bank classification obtained from R package *countrycode* [23] (figure 1). We opted for the World Bank classification of countries, as it more accurately represents global regions based on overall infrastructure and economic characteristics, which we consider crucial in determining the distribution of name bearers. For example, some Latin American countries that are more similar in terms of infrastructure and economy are grouped with countries with very different economic contexts under other geographical classifications (e.g. Mexico grouped with USA and Canada under other classifications). The origin and destination of name bearers were used to derive two metrics that relate to the native composition of each country, and its name bearer composition housed in their biological collections. These two metrics are explained in the section below.

### (c) Congruence between name bearers and endemic fish composition

Biological collections should ideally house name bearers from the country's ecological and geographical context to maximize biodiversity benefits [13], such as for biological research and educational purposes. Here, we used two metrics to assess how well a country's name bearer collections reflect its endemic species composition (per country native species richness is shown in electronic supplementary material, figure S1). Endemic species composition was extracted directly from CAS. Here, we classified endemic name bearers as those species with geographic distribution (according to CAS) restricted to a single country. Further information on species composition data and analysis with all native species per country is presented in electronic supplementary material (richness and country species composition from CAS).

The first metric, endemic deficit, is the proportion of the known endemic species lacking in a country's collections (*endemic deficit* = *number of endemic name bearers outside of the country of origin* ÷ *country's endemic species richness*). High values of this metric indicate countries missing most of their endemic species' composition in their biological collections. The second metric measured the proportion of country's name bearers that belong to non-native species (*non-endemic representation* = *number of non-endemic name bearers* ÷ *country's endemic name bearer richness*). Higher values identify countries with an overrepresentation of endemic species from other countries.

We decided to present here only the analysis with endemic species, because it is a more conservative way to show the mismatch between native fish fauna and the name bearers housed in biological collections. This filter ensures that we did not underestimate the representation-based metrics, where wide-ranging species were first sampled in foreign countries. We also calculated the representation-based metrics for the full dataset (see electronic supplementary material, figure S3), but the results were qualitatively similar with only minor quantitative differences.

### (d) Determinants of per-country number of name bearers and related metrics

We investigated five factors potentially driving the per-country number of name bearers: (i) per-country number of native fish species; (ii) per-country number of preserved fish specimens extracted from Global Biodiversity Information Facility (GBIF), corrected by the respective country's area; (iii) per-country gross domestic product (GDP); (iv) per-country number of ichthyological collections; and (v) number of years since independence for formerly colonized countries (for countries that never experienced colonization the value of this metric is zero). All explanatory variables were centred and scaled (*z*-transformed) to allow direct comparisons of their effect sizes [24]. We then used a generalized linear model (GLM) with negative binomial error distribution to estimate the per-country number of name bearers.

We also checked whether these five explanatory variables affected the country-level metrics (endemic deficit and non-native representation) using GLMs with a beta-binomial error distribution [25]. To assess the validity of our models, we evaluated the homoscedasticity in the residuals against predicted values (model assessment graphics are available in electronic supplementary material, figures S3) and through qq-plots using a simulation approach [26]. The standardized effect sizes for all explanatory variables are provided in electronic supplementary material, tables S1–S5.

### (e) Data sources

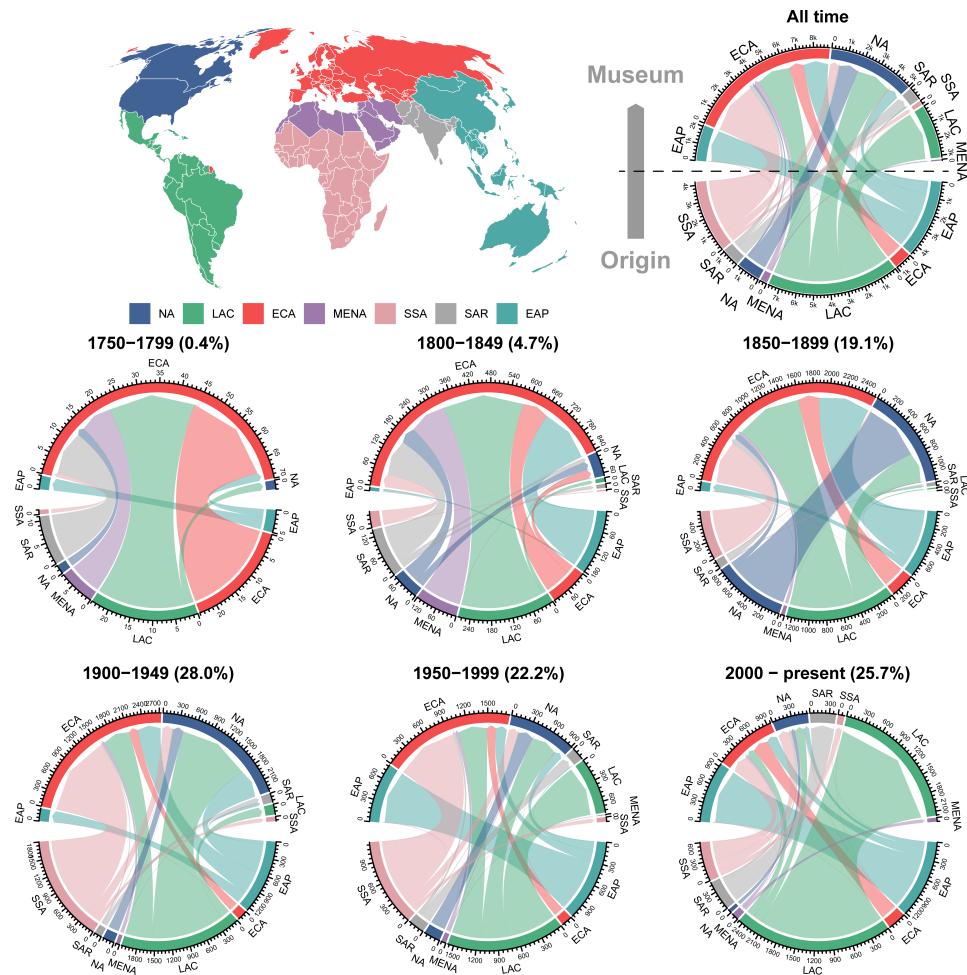
The primary data used in this article were taken from Eschemeyer's CAS [22]. CAS also provided information on name bearers sampling location, museums, native species distribution and the number of museums. Bio-Dem (<https://bio-dem.surge.sh>) provided socio-economic variables (GDP, years since independence). GBIF provided the number of occurrences per area. The number of museums, number of native species per country, GDP, years since independence and number of occurrences per area according to GBIF were used in the modelling approach.

## 3. Results

### (a) Spatial characterization of name bearers across time

Since 1758, a total of 20 246 fish binomials have been described globally. Of these binomials, 13 325 name bearers were collected in countries within the Global South (Latin America and Caribbean, Sub-Saharan Africa, Middle East and North Africa and South Asia), compared with 6921 name bearers collected in countries of the Global North (North America, Europe and Central Asia and East Asia and Pacific). Notably, although a greater number of name bearers were sampled in the Global South, these regions are home to only 4208 name bearers, whereas all Global North countries collectively house 16 038, corresponding to 79% of all freshwater name bearers.

Museums in Europe and Central Asia (ECA) and North America (NA) have historically served as the primary destination for freshwater/brackish name bearers from all continents, with their dominance in housing name bearers peaking between 1900 and 1949, when nearly 90% of name bearers ( $n = 5051$ ) were housed in these regions (figure 1). Only in the past two decades there has been a trend for countries to retain their name-bearing specimens within their own ichthyological collections (figure 1, 2000–present). Although recent patterns in name bearer retention can be related to the establishment of Global South biological collections in the last 50 years, Global North biological collections still keep the historical pattern of housing foreign name



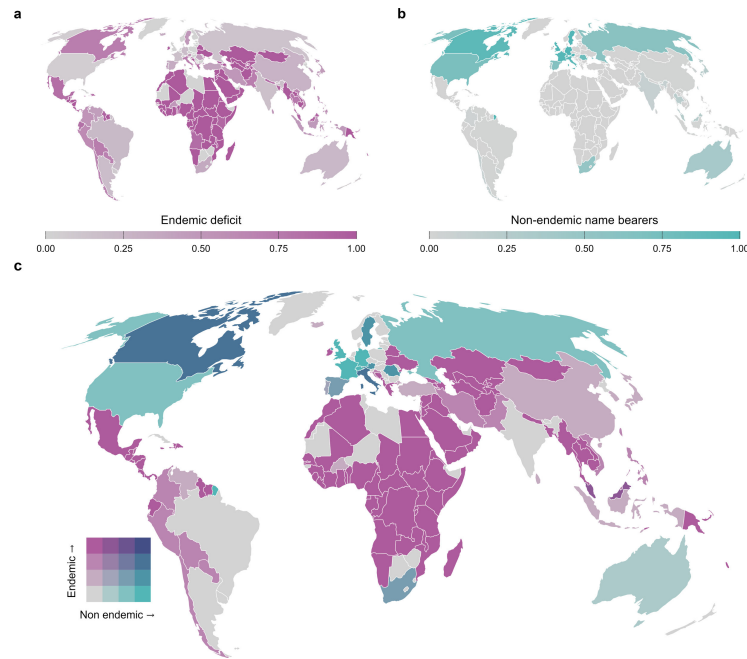
**Figure 1.** Global flow of fish name-bearing specimens among macropolitical regions in 50-year time intervals. Across all chord diagrams, macropolitical regions are coloured according to World Bank classification as illustrated in the upper left cartogram. The lower half of the circular plots represents the regions, where the name bearer was collected (source regions), and the upper half is the region of the biological collection in which the name bearer is housed (housing regions). The numbers on the outer circles represent the total number of name bearers exchanged between each region in each period. The percentage besides the year corresponds to the percentage of name bearers of that period relative to the total name bearers in all periods. Macropolitical regions: ECA, Europe and Central Asia; NA, North America; SSA, Sub-Saharan Africa; SAR, South Asian Region; MENA, Middle East and North Africa; LAC, Latin America and the Caribbean; EAP, East Asia and Pacific. The red region within LAC corresponds to French Guiana, an overseas territory of France, and is thus categorized as part of ECA in the map.

bearers. For instance, 37% of the 2466 name bearers housed in biological collections of the Global North (East Asia and Pacific, Europe and Central Asia and North America) during the 2000s were representatives of fish faunas from the Global South (Latin America and Caribbean, Middle East and North Africa, South Asia and Sub-Saharan Africa), with the opposite occurring for 0.2% (6 name bearers out of 2732) of name bearers housed in the Global South countries.

### (b) Congruence between native fish composition and name bearers

The endemic deficit was very low for some countries in the Global North (e.g. 5% for USA, 15% for France, 20% for Germany and 0% for Great Britain). On the other hand, this proportion was high for most Global South countries, with a mean of 88% of endemic deficit, with mega-diverse countries having high values, such as Ecuador (81%), Bolivia (72%) and Benin (100%). Countries in Europe and Central Asia housed the majority of name bearers representing their endemic fish fauna, while collections in South Asia and Middle East and North Africa lacked the name bearers for most of their endemic species (figure 2a).

On the other hand, non-endemic representation in Global North collections is evident across most developed countries in Europe and Central Asia, North America and Australia (figure 2b), with some countries like the USA, Great Britain, France and Germany having high values of non-endemic representation (71%, 99%, 97% and 99%, respectively). Most Global South countries also showed low non-endemic representation, as their biological collections are primarily composed of native species (figure 2b). Overall, we found an underrepresentation of native fish fauna in most biological collections across the Global South and an overrepresentation of non-endemic name bearers in Global North collections (figure 2c).



**Figure 2.** Global patterns in the representation of endemic and non-endemic fish name bearers across countries' ichthyological collections. (a) Endemic deficit, as the proportion of the known endemic species lacking in a country's collections. (b) Non-endemic name bearers, as the proportion of country's name bearers that belong to non-native species. (c) Both mapped patterns combined in a bivariate map (c), where the breaks represent data quartiles (0–25%, 25–50%, 50–75% and 75–100%). See the main text for the definition of endemic species.

### (c) Determinants of per-country number of name bearers and related metrics

Socio-economic and biological factors strongly shaped concentration, distribution and the characteristics of the per-country name bearers. Specifically, the total number of name bearers within the countries is driven by GDP and native species richness (figure 3a,b).

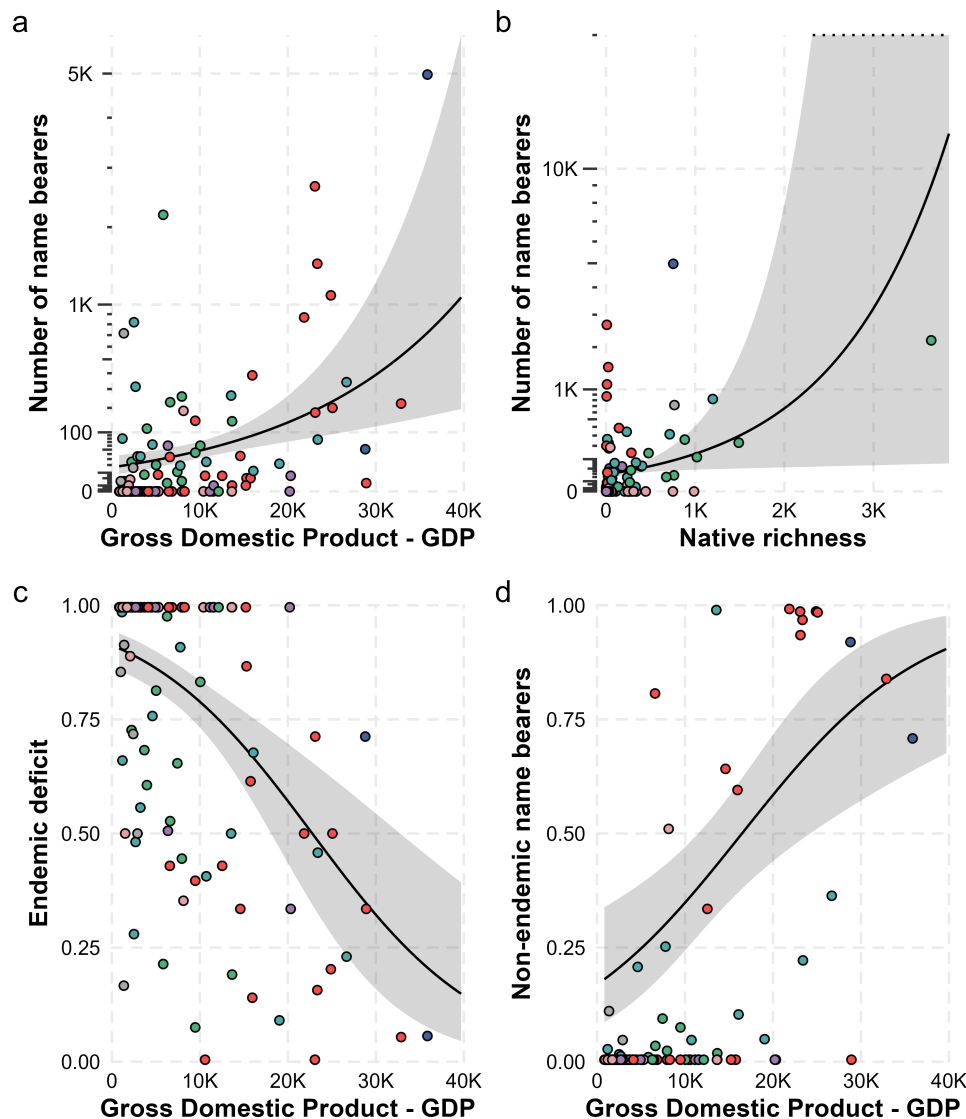
The characteristics of the country's biological collections regarding the source of the name bearer (endemic deficit and non-endemic representation) housed in their collections are also explained by GDP. Countries with higher GDP tend to have lower values of endemic deficit in their biological collections than countries with lower GDP. Additionally, wealthier countries (high GDP) often present an overrepresentation of non-endemic name bearers in their biological collections (figure 3c,d). Other factors that we initially hypothesized to be important did not appear to influence our metrics. For example, years since independence did not affect any metric, possibly indicating that, even after independence, little has been done to address other enduring aspects of colonial influence in these countries. Figure 3 represents the most important variables influencing the response variables in our model (except for the total number of name bearers, for which we show the top two variables). Full results of the models, including their parameter estimates, associated standard errors, model adequacy and other models' statistics, are presented in electronic supplementary material, tables S1–S5.

## 4. Discussion

Despite critiques on the concept of species [27,28] and the possibility of assessing biodiversity patterns using general metrics based on functional, phylogenetic and genetic data, name bearers remain the fundamental unit of nearly all investigations in biodiversity and conservation science [29]. By recognizing name bearers as the cornerstone of biodiversity knowledge, we assume that institutions or countries housing most of the name bearers also possess a disproportionate share of fundamental biological knowledge. While arguments for a more equitable distribution of knowledge could extend beyond ethical grounds and contribute to faster mitigation of biodiversity shortfalls [30], virtually no study has examined country-level asymmetries in the concentration and representation of name bearers worldwide. For the first time, we quantify such biases in reference specimens, which are crucial for the identification and description of new freshwater and brackish fish species, the most diverse vertebrate group on the Earth.

### (a) A glimpse into the biological knowledge accumulation process

Countries with colonial histories and early concentration of economic power have historically accumulated name bearers primarily from tropical countries in Latin America and Africa. This pattern, in which Global South countries serve as providers and Global North countries act as custodians of name bearers, only began to change in the twenty-first century. In recent years, there has been an increase in the retention of specimens in the countries where they were collected. Despite some recent progress, most freshwater fish species discovered in the early stages remain housed in museums located far from their original collection sites. While this situation could foster international cooperation by leveraging materials to advance ecological and



**Figure 3.** Influence of GDP on countries' knowledge representation of fish name bearers. The per-country GDP effect on the (a) total number of name bearers, (b) native species richness, (c) endemic deficit and (d) non-endemic representation. All the figures represent the marginal effects of the predictor variables on the response variables. The colours represent the geopolitical regions according to Word Bank. Colour codes for the regions are the same as presented in the inset map in figure 1.

conservation knowledge [31], it does not ensure symmetrical collaboration. Global North institutions often retain control over name bearers' accessibility, perpetuating existing imbalances and limiting access to this material to only a select few researchers (see [32] for a discussion on this topic).

Another critical aspect highlighted by our findings is the continued practice of researchers sending collected specimens back to their home countries, particularly in North America and Europe and Central Asia. Past naturalistic exploration like the Thayer Expedition [33] in Brazil advanced biological collections, but the persistence of this practice indicates that countries with a high concentration of foreign name bearers still overlook the consequences of this accumulation process for inclusivity and the development of biodiversity science. The uneven distribution of name bearers also raises both ethical and legal issues. There is evidence, particularly from palaeontological material, of illegal samplings [11,34]. Even when material is legally exported, ethical issues arise regarding the lack of transparency in how biological material is collected and exported (e.g. parachute science [12]). Protective laws and regulations can help maintain, protect and preserve important biological samples within the country where they were collected. For instance, Brazil implemented such regulations in 1990 [35] mandating that significant biological reference materials (e.g. holotype and syntypes) be retained within the country, either completely or partially in the case of paratypes. This measure has contributed to an increase in name bearer retention in Brazil during the twenty-first century, as evidenced by the substantial portion of name bearers from Brazil shown in figure 1.

### (b) Knowledge split and its consequences for biodiversity research

Regardless of the underlying causes for the biased name bearer distribution, this biodiversity knowledge split (a reference to the habitat split phenomenon [36]) poses additional challenges to Global South researchers in at least three different ways. First, there is a separation of the sampling sites from where validation can occur more precisely. Second, accessing name bearers in foreign countries incurs higher traveling expenses for researchers, particularly those from countries where science

has faced significant financial challenges (e.g. Brazil [37] and Colombia [38]) in recent years and decades. Third, relative to domestic standards, the international legislation imposes additional requirements for visiting researchers from the Global South, including more expensive visa fees and lengthier processing times [39]. The alternative is not necessarily simpler, as loans of name bearers often involve additional bureaucracy when specimens must be transported across international borders [39] and carry the risk of damage or destruction during transit [15]. These factors collectively delay the cataloguing and discovery of species, particularly in tropical countries of Global South where biodiversity is still largely unknown [4,40].

The historical growth of Global North collections in both infrastructure and personnel likely contributes to the emergence of the latitudinal taxonomy gradient (LTG) [3], where species in the tropics are less likely to be accurately recognized or classified [29,41,42]. The persistence of uncertainties in tropical species delimitation is closely tied to the unequal distribution of name bearers and preserved specimens, which are disproportionately concentrated in the Global North, far from regions where most species discoveries occur (i.e. tropical regions). The availability of a larger number of preserved specimens increases the chances of ecological, phylogenetic and taxonomic research, which in turn influences species classification and can lead to biased inferences in ecology, evolution and conservation [43–45]. Because of the LTG, tropical regions probably harbour more species than currently estimated, as many tropical species remain hidden within already described taxa. While taxonomic revisions are critical to describe these species, the fact that most of the early-discovered freshwater fishes requiring revisionary studies are located far from their native regions underscores the persistent challenges in addressing biodiversity knowledge shortfall.

### (c) Overcoming barriers to name bearer distribution

Recent proposals have advocated for the use of alternative materials, such as tissues and digitized images to improve the accessibility of museum specimens (e.g. [8,46,47]). Despite the importance of those initiatives, with very successful examples in herbarium collections [48], mitigation of biodiversity knowledge shortfalls will require more than technocratic fixes [30]. The accessibility problem to name bearer material cannot rely solely on ‘remote strategies’, because not all specimens can be easily digitized [7]. Fishes are a good example of how challenging the digitization of specimens can be as it requires high-quality and detailed images of anatomical structures to reach unambiguous species identification [49]. Important diagnostic characters, including internal structures like bones and swimbladders, cannot be captured in digitized images unless in high-quality three-dimensional images generated by, for example, CT scans. Additionally, it is possible that the digitization of natural collections replicates the inequality and the inaccessibility of data created by a colonial past that shaped these collections [50]. Finally, some successful and comprehensive examples of projects that aim to provide high-quality three-dimensional digitalization images, like *oVert* [51], do not prioritize name bearer specimens because they are not easily accessible or because important characteristics are damaged [51].

The ideal scenario would involve returning representative name bearers to the biological collections of the countries where they were collected. However, this requires lengthy bilateral negotiations, even despite instances of illegal collection [11]. It is also worth mentioning the presence of name bearers in private collections, about which we have limited information [17]. Repatriation often demands more than legal evidence; public pressure is also necessary [52]. Moreover, recipient collections must have the infrastructure to maintain the biological material long-term, which may be lacking in some Global South countries [53]. Thus, a better division of labour and costs is needed towards a global goal of describing and protecting biodiversity [54,55]. This approach would prevent delays in improving biodiversity knowledge and its benefits [13]. An alternative would be to facilitate the accessibility of name bearers by covering the costs for researchers from the countries of origin to visit name bearers housed in foreign biological collections [55]. Multilateral organizations, like the Convention on Biological Diversity, could promote research groups to secure funding that would allow researchers working on taxonomic research to visit specimens housed in different museums. We suggest that programmes for visiting research explicitly incorporate aspects of the origin of the material and of the researchers in their evaluations, giving preference for researchers working in research institutes in countries where the material was sampled.

Another important step towards a fairer and more equitable distribution of knowledge involves addressing narratives found in museums and natural history collections. Usually, these narratives highlight the achievements of Western white men, overlooking the historical exploration of the countries from which the material was taken. We agree with Ashby & Machin [56] that including stories about the troubling social histories associated with acquiring natural history specimens is crucial for biological collections and curators to reflect the societies they serve accurately.

One of the most direct benefits of retaining name bearers locally is the advantage for the research community and broader society, as it supports local scientific and educational development. The underrepresentation of name bearers in the Global South has probably decreased investments that could foster local taxonomic science [54], which is now indirectly reflected along latitudinal gaps and biases in biodiversity knowledge [41,55,56]. To illustrate the neglect of this issue, international agreements like the Nagoya Protocol, which aim to ensure fair and equitable sharing of biodiversity benefits, only address genetic resources (e.g. [57]), ignoring the importance of housing reference material for boosting scientific development locally.

## 5. Conclusion

Our results, using fishes as a model organism, reveal an uneven distribution of biological knowledge worldwide, leading to both global and local (country-level) consequences. The knowledge housed by a country often fails to represent its native freshwater and brackish fish fauna accurately. To address these biological knowledge gaps, it is crucial to facilitate access to

these materials for researchers from the countries where they were collected. Otherwise, regardless of the available technology and methods for monitoring biodiversity, our understanding of the natural world will remain inequitable. Beyond identifying the issue, we hope this study serves as a catalyst for developing solutions that promote a more equitable and fair distribution of biological knowledge, one that does not replicate the existing economic and social inequalities in our society.

**Ethics.** This work did not require ethical approval from a human subject or animal welfare committee.

**Data accessibility.** The primary data used in this article are freely available from Eschemeyer's CAS [22]. All data and code used to reproduce the figures and models are available in the Zenodo repository [58]. The source code and a reproducible workflow for all analyses can be found at GitHub [59]. The electronic supplementary material provides a detailed overview of the data used.

Supplementary material is available online [60].

**Declaration of AI use.** Grammar was thoroughly checked using specialized software and artificial intelligence tools (specifically, ChatGPT), and no other AI tools were employed.

**Authors' contributions.** G.N.: conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, validation, visualization, writing—original draft, writing—review and editing; B.H.M.S.: data curation, formal analysis, investigation, methodology, software, validation, visualization, writing—review and editing; L.E.F.F.: conceptualization, investigation, validation, writing—review and editing; M.L.A.: formal analysis, methodology, software, visualization; E.B.N.: investigation, methodology, validation, writing—review and editing; M.M.F.M.: investigation, validation, writing—review and editing; M.d.S.L.: formal analysis, validation, writing—review and editing; A.R.: methodology, writing—review and editing; L.D.: software, writing—review and editing; T.M.d.S.F.: investigation, writing—review and editing; B.S.: investigation, writing—review and editing; W.J.d.G.: investigation, writing—review and editing; M.R.M.: investigation, methodology, validation, visualization, writing—review and editing; J.A.F.D.-F.: funding acquisition, resources, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** J.A.F.D.-F. has been constantly supported by productivity grants from CNPq. L.E.F.F. is supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, no. 88887.923875/2023-00). M.L.A. thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for his scholarship (88887.903889/2023-00). E.B.N. is supported by Fundação de Ciência, Tecnologia, Inovação e Desenvolvimento Sustentável Guamá (terms ITVDS/Fundação Guamá no. 075/2023, 027/2024 and 0099/2025). M.M.F.M. was supported by INCT-Peixes (MCTIC/CNPq 405706/2022-7). A.R. is supported by a grant from the Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq (proc. 351723/2023-4). G.N. was supported by National Institutes for Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation funded by MCTIC/CNPq. (grants 465610/2014-5 and 409197/2024-6) and FAPEG (grant 201810267000023).

**Acknowledgements.** We thank M. Griffiths and Renata Rúbia Ota for critically reviewing and discussing earlier versions of this study. We also acknowledge the Catalog of Fishes team and its collaborators for publicly sharing their data, especially William Eschmeyer, Robert Fricke and Richard van der Laan.

## References

- Moura MR *et al.* 2023 Unwrapping broken tails: biological and environmental correlates of predation pressure in limbless reptiles. *J. Anim. Ecol.* **92**, 324–337. (doi:10.1111/1365-2656.13793)
- Suarez AV, Tsutsui ND. 2004 The value of museum collections for research and society. *Bioscience* **54**, 66. (doi:10.1641/0006-3568(2004)054)
- Freeman BG, Pennell MW. 2021 The latitudinal taxonomy gradient. *Trends Ecol. Evol.* **36**, 778–786. (doi:10.1016/j.tree.2021.05.003)
- Moura MR, Jetz W. 2021 Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nat. Ecol. Evol.* **5**, 631–639. (doi:10.1038/s41559-021-01411-5)
- ICZN. 1999 *International code of zoological nomenclature*, 4th edition. Singapore: International Commission on Zoological Nomenclature.
- Sluys R. 2021 Attaching names to biological species: the use and value of type specimens in systematic zoology and natural history collections. *Biol. Theory* **16**, 49–61. (doi:10.1007/s13752-020-00366-3)
- Gutiérrez EE, Pine RH. 2017 Specimen collection crucial to taxonomy. *Science* **355**, 1275–1275. (doi:10.1126/science.aan0926)
- Nelson G, Ellis S. 2019 The history and impact of digitization and digital data mobilization on biodiversity research. *Phil. Trans. R. Soc. B* **374**, 20170391. (doi:10.1098/rstb.2017.0391)
- Nachman MW *et al.* 2023 Specimen collection is essential for modern science. *PLoS Biol.* **21**, e3002318. (doi:10.1371/journal.pbio.3002318)
- Martin TE, Bennett GC, Fairbairn A, Mooers AO. 2023 'Lost' taxa and their conservation implications. *Anim. Conserv.* **26**, 14–24. (doi:10.1111/acv.12788)
- Cisneros JC, Ghilardi AM, Raja NB, Stewens PP. 2021 The moral and legal imperative to return illegally exported fossils. *Nat. Ecol. Evol.* **6**, 2–3. (doi:10.1038/s41559-021-01588-9)
- Raja NB, Dunne EM, Mawani A, Khan TM, Nätscher PS, Ghilardi AM, Chattopadhyay D. 2021 Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat. Ecol. Evol.* **6**, 145–154. (doi:10.1038/s41559-021-01608-8)
- Nanglu K *et al.* 2023 The nature of science: the fundamental role of natural history in ecology, evolution, conservation, and education. *Ecol. Evol.* **13**, e10621. (doi:10.1002/ece3.10621)
- Isbell F *et al.* 2023 Expert perspectives on global biodiversity loss and its drivers and impacts on people. *Front. Ecol. Environ.* **21**, 94–103. (doi:10.1002/fee.2536)
- Burman AG. 1986 The homeotype: a plea for respectability. *Taxon* **35**, 317–321. (doi:10.2307/1221278)
- Vorontsova MS, Lowry PP II, Andriambololona SR, Wilmé L, Rasolohery A, Govaerts R, Fincinski SZ, Humphreys AM. 2021 Inequality in plant diversity knowledge and unrecorded plant extinctions: an example from the grasses of Madagascar. *Plants People Planet* **3**, 45–60. (doi:10.1002/ppp3.10123)
- Uetz P *et al.* 2019 A global catalog of primary reptile type specimens. *Zootaxa* **4695**, 438–450. (doi:10.11646/zootaxa.4695.5.2)
- Moura MR *et al.* 2024 A phylogeny-informed characterisation of global tetrapod traits addresses data gaps and biases. *PLoS Biol.* **22**, e3002658. (doi:10.1371/journal.pbio.3002658)
- Park DS *et al.* 2023 The colonial legacy of herbaria. *Nat. Hum. Behav.* **7**, 1059–1068. (doi:10.1038/s41562-023-01616-7)
- Troutet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017 Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132. (doi:10.1038/s41598-017-09084-6)
- Reis RE, Albert JS, Di Dario F, Mincarone MM, Petry P, Rocha LA. 2016 Fish biodiversity and conservation in South America. *J. Fish Biol.* **89**, 12–47. (doi:10.1111/jfb.13016)
- Fricke R, Eschmeyer WN, Van der Laan R. *Eschmeyer's catalog of fishes*. See <https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>.
- Arel-Bundock V, Enevoldsen N, Yetman C. 2018 countrycode: an R package to convert country names and country codes. *JOSS* **3**, 848. (doi:10.21105/joss.00848)

24. Schielzeth H. 2010 Simple means to improve the interpretability of regression coefficients. *Methods Ecol. Evol.* **1**, 103–113. (doi:10.1111/j.2041-210x.2010.00012.x)
25. Ferrari S, Cribari-Neto F. 2004 Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31**, 799–815. (doi:10.1080/0266476042000214501)
26. Hartig F. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.6. See <https://CRAN.R-project.org/package=DHARMA>.
27. De Pinna MCC. 1999 No title found. *Rev. Fish Biol. Fish.* **9**, 353–373. (doi:10.1023/A:1008911414399)
28. De Queiroz K. 2007 Species concepts and species delimitation. *Syst. Biol.* **56**, 879–886. (doi:10.1080/10635150701701083)
29. Balakrishnan R. 2005 Species concepts, species boundaries and species identification: a view from the tropics. *Syst. Biol.* **54**, 689–693. (doi:10.1080/10635150590950308)
30. Chapman M *et al.* 2024 Biodiversity monitoring for a just planetary future. *Science* **383**, 34–36. (doi:10.1126/science.adh8874)
31. Meineke EK, Davies TJ, Daru BH, Davis CC. 2019 Biological collections for understanding biodiversity in the Anthropocene. *Phil. Trans. R. Soc. B* **374**, 20170386. (doi:10.1098/rstb.2017.0386)
32. Meineke EK, Daru BH. 2021 Bias assessments to expand research harnessing biological collections. *Trends Ecol. Evol.* **36**, 1071–1082. (doi:10.1016/j.tree.2021.08.003)
33. Wallace AR. 1870 Thayer expedition: scientific results of a journey in Brazil, by Louis Agassiz and his travelling companions geology and physical geography of Brazil. *Nature* **2**, 510–512. (doi:10.1038/002510a0)
34. Cisneros JC *et al.* 2022 Digging deeper into colonial palaeontological practices in modern day Mexico and Brazil. *R. Soc. Open Sci.* **9**, 210898. (doi:10.1098/rsos.210898)
35. Diário oficial da união. 1990 Aprova o Regulamento sobre coleta, por estrangeiros, de dados e materiais científicos do Brasil.
36. Becker CG, Fonseca CR, Haddad CFB, Batista RF, Prado PI. 2007 Habitat split and the global decline of amphibians. *Science* **318**, 1775–1777. (doi:10.1126/science.1149374)
37. Fernandes GW *et al.* 2017 Dismantling Brazil's science threatens global biodiversity heritage. *Perspect. Ecol. Conserv.* **15**, 239–243. (doi:10.1016/j.pecon.2017.07.004)
38. Huddart JEA, Crawford AJ, Luna-Tapia AL, Restrepo S, Di Palma F. 2022 EBP-Colombia and the bioeconomy: Genomics in the service of biodiversity conservation and sustainable development. *Proc. Natl Acad. Sci. USA* **119**, e2115641119. (doi:10.1073/pnas.2115641119)
39. Valdez J, Damasceno G, Oh RRY, Quintero Uribe LC, Barajas Barbosa MP, Amado TF, Schmidt C, Fernandez M, Sharma S. 2024 Strategies for advancing inclusive biodiversity research through equitable practices and collective responsibility. *Conserv. Biol.* **38**, e14325. (doi:10.1111/cobi.14325)
40. Moura MR, Jetz W. 2024 Reply to: estimates of the number of undescribed species should account for sampling effort. *Nat. Ecol. Evol.* **8**, 641–644. (doi:10.1038/s41559-024-02343-6)
41. Chek AA, Austin JD, Lougheed SC. 2003 Why is there a tropical–temperate disparity in the genetic diversity and taxonomy of species? *Evol. Ecol. Res* **5**, 69–77.
42. Dubey S, Shine R. 2012 Are reptile and amphibian species younger in the Northern Hemisphere than in the Southern Hemisphere? *J. Evol. Biol.* **25**, 220–226. (doi:10.1111/j.1420-9101.2011.02417.x)
43. Guedes JJM, Diniz-Filho JAF, Moura MR. 2024 Macroecological correlates of Darwinian shortfalls across terrestrial vertebrates. *Biol. Lett.* **20**, 20240216. (doi:10.1098/rsbl.2024.0216)
44. Guedes JJM, Moura MR, Jardim L, Diniz-Filho JAF. 2025 Global patterns of taxonomic uncertainty and its impacts on biodiversity research. *Syst. Biol.* f010. (doi:10.1093/sysbio/syaf010)
45. Diniz-Filho JAF *et al.* 2023 Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Front. Biogeogr.* **15**, G59566. (doi:10.21425/f5fbg59566)
46. Byrne AQ. 2023 Reimagining the future of natural history museums with compassionate collection. *PLoS Biol.* **21**, e3002101. (doi:10.1371/journal.pbio.3002101)
47. Raxworthy CJ, Smith BT. 2021 Mining museums for historical DNA: advances and challenges in museumomics. *Trends Ecol. Evol.* **36**, 1049–1060. (doi:10.1016/j.tree.2021.07.009)
48. Global Strategy for Plant Conservation (GSPC). 2018 Brazilian Flora 2020: Innovation and collaboration to meet Target 1 of the. *Rodriguesia* **69**, 1513–1527. (doi:10.1590/2175-7860201869402)
49. Kolmann MA, Kalacska M, Lucanus O, Sousa L, Wainwright D, Arroyo-Mora JP, Andrade MC. 2021 Hyperspectral data as a biodiversity screening tool can differentiate among diverse Neotropical fishes. *Sci. Rep.* **11**, 16157. (doi:10.1038/s41598-021-95713-0)
50. Kaiser K, Heumann I, Nadim T, Keysar H, Petersen M, Korun M, Berger F. 2023 Promises of mass digitisation and the colonial realities of natural history collections. *J. Nat. Sci. Collect* **11**, 13–25.
51. Blackburn DC *et al.* 2024 Increasing the impact of vertebrate scientific collections through 3D imaging: the openvertebrate (oVert) thematic collections network. *BioScience* **74**, 169–186. (doi:10.1093/biosci/biad120)
52. Lenharo M, Rodrigues M. 2022 How a Brazilian dinosaur sparked a movement to decolonize fossil science. *Nature* **605**, 18–19. (doi:10.1038/d41586-022-01093-4)
53. Paknia O, Rajaei Sh. H, Koch A. 2015 Lack of well-maintained natural history collections and taxonomists in megadiverse developing countries hampers global biodiversity exploration. *Org. Divers. Evol.* **15**, 619–629. (doi:10.1007/s13127-015-0202-1)
54. Leal JS, Soares B, Franco ACS, de Sá Ferreira Lima RG, Baker K, Griffiths M. 2022 Decolonising ecological research: a debate between Global North geographers and Global South field ecologists. *SocArXiv*. (doi:10.31235/osf.io/wbzh2)
55. Kemp ME, Boville AE, Carneiro CM, Jacisin JJ III, Law CJ, Ledesma DT, Meza A, Shields-Estrada A, Xu T. 2023 Looking back for the future: the ecology of terrestrial communities through the lens of conservation paleobiology. *Annu. Rev. Ecol. Evol. Syst.* **54**, 259–282. (doi:10.1146/annurev-ecolsys-110421-101343)
56. Ashby J, Machin R. 2021 Legacies of colonial violence in natural history collections. *J. Nat. Sci. Collect.* **8**, 44–54.
57. Kartzinel TR *et al.* 2025 Global availability of plant DNA barcodes as genomic resources to support basic and policy-relevant biodiversity research. *Mol. Ecol.* **34**, e17712. (doi:10.1111/mec.17712)
58. Bruno M, Gabriel N. 2025 GabrielNakamura/NBT\_code\_data. Zenodo. (doi:10.5281/ZENODO.17143162)
59. Nakamura G *et al.* 2025 The hidden biodiversity knowledge split in biological collections. GitHub. See [https://gabrielnakamura.github.io/NBT\\_code\\_data/](https://gabrielnakamura.github.io/NBT_code_data/).
60. Nakamura G, Stabile BHM, Frateles LEF, Araujo ML, Neuhaus EB, Marinho MMF *et al.* 2025 Supplementary material from: The hidden biodiversity knowledge split in biological collections. Figshare. (doi:10.6084/m9.figshare.c.8063164)