The macroecology of knowledge: Spatio-temporal patterns of name-bearing types in biodiversity science

Authors: Gabriel Nakamura[1,2*]§, Bruno Henrique Mioto Stabile[3]§, Lívia Estéfane F. Frateles[4]§, Matheus Lima Araujo[4], Emanuel Neuhaus[5], Manoela Maria Ferreira Marinho[6], Melina de Souza Leite[1], Aline Richter[7], Liuyong Ding[8], Tiago Magalhães da Silva Freitas[9], Bruno Eleres Soares[10], Weferson Júnio da Graça[4,11], José Alexandre Felizola Diniz-Filho[2,4]

Affiliations:

[1] Universidade de São Paulo, Ecology Department, São Paulo, Brazil

[2] National Institute of Science and Technology – Ecology, Evolution and Conservation Biology, INCT EECBio, Universidade Federal de Goiás, Goiânia, Brazil

[3] Graduate Program in Ecology of Inland Water Ecosystems, Departamento de Biologia, Centro de Ciências Biológicas, Universidade Estadual de Maringá, Maringá, Paraná, Brazil

[4] Graduate Program in Ecology and Evolution, Universidade Federal de Goiás, Goiânia, Brazil

[5] Vale Technological Institute (ITV), Belém, Pará, Brazil

[6] Fish Systematics and Morphology Laboratory, Federal University of Paraíba, João Pessoa, Brazil

[7] National Center for Research and Conservation of Reptiles and Amphibians (RAN/ICMBio), Goiânia, Goiás, Brazil

[8] Museum of Hydrobiological Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, 430072, China

[9] Zoology Laboratory, Campus Marajó-Breves, Federal University of Pará, Alameda IV, 3418, 68000-000 Breves, PA, Brasil

[10] Institute of Environmental Change & Society, University of Regina, Canada

[11] Núcleo de Pesquisas em Limnologia, Ictiologia e Aquicultura, Centro de Ciências Biológicas, Universidade Estadual de Maringá, Maringá, Paraná, Brazil

§co-first authors

*corresponding author: gabriel.nakamura.souza@gmail.com

Abstract: Ecological and evolutionary processes are recognized as the main factors generating and maintaining biodiversity. However, how biodiversity knowledge is collated, organized, and distributed worldwide influences our perceptions and inferences about biodiversity and the underlying processes. We demonstrated that name-bearing type specimens (NBT), the most fundamental reference for the identity of any species, of all freshwater and brackish fish species in the world are mostly housed in museums in Global North countries. The unequal distribution of NBT results from historical and socioeconomic factors and has implications for both the Global North and South countries. For the Global North, which concentrates most of NBT, we found a mismatch between NBT housed in their ichthyological collections and their native biotas. On the other hand, countries with most NBT of their native species housed elsewhere face a barrier in advancing biodiversity research due to the difficulty in accessing reference material, hampering global efforts in cataloging, reviewing, and describing new species. We advocate that if we are truly committed to advancing biodiversity research, we should pursue global initiatives to make the distribution of biological knowledge fairer among countries, which involves programs for specimen repatriation and facilitation of accessibility of NBT material to researchers from the countries in which they were collected.

Main Text: Museums can be viewed as gatekeepers of representative fragments of the world. They provide a valuable source and representation of human history, culture, knowledge, and biodiversity. Natural history museums and biological collections (hereafter NHBC) are known for maintaining an organized and curated record of extant and extinct specimens. Given this role of NHBC, numerous authors have highlighted different aspects in which they enhance our understanding of the natural world by shedding light on global change phenomena(*1*), spatial and temporal ecological processes(*2*), public health(*1*), and educational purposes(*3*). Additionally, more frequently associated with the natural sciences, they house representative specimens(*1*) essential for naming, describing, and classifying organisms through taxonomic research.

Among the specimens in NHBC, the name-bearing types (hereafter NBT) hold special importance in biodiversity studies (fig. S1). NBT specimens are pivotal in taxonomic studies as they constitute the fundamental reference upon which taxonomists rely to review described species or propose new ones(*2*, *3*). Accessibility to NBT is, therefore, crucial for taxonomic research, as assessing some taxonomically relevant features is only possible by handling the original NBT. Hence, even though NBT digitalization is undeniably important for biodiversity studies(*4*), the original material cannot be replaced by indirect data or molecular techniques for most organisms(*5*). However, as with any biological data, the representation, curation, and concentration of NBT in NHBC worldwide are not free from biases, making accessibility more laborious for some researchers than others.

Consequently, this hampers the development of a reliable effort in cataloging and organizing biological diversity knowledge. Here, we show that knowledge concentration over time mirrors historical periods of socioeconomic dynamics of countries, evidencing how the concentration of economic power in the world mediates the accumulation of biological knowledge and, consequently, might affect our perception of biological patterns. The investigation of the patterns of accumulation of biological knowledge and its influence on our perception of the natural world is what we call macroecology of knowledge.

Recently, Paleontology has drawn the scientific community's attention to the importance of NBT distribution in biodiversity studies (e.g., see the *Ubirajara* case(*6*)). Given the rarity of paleontological records that can identify a species, this unique sample often serves as the NBT for a given species. In this context, recent studies have shown that a biased concentration of fossil samplings deposited in NHBC in the Global North has profound effects on shaping the knowledge of the field, distorting our view of deep-time biodiversity(*7*). The same problem extends to extant organisms, producing significant consequences for the development of biodiversity science and with more negative effects in countries from the Global South.

Conservative estimates indicate that only 13-18% of species are known and described, and the probability of discovering new species and taxonomic uncertainties of already described species are concentrated mainly in tropical regions(*8*, *9*). However, taxonomists frequently report that the tropics are the region from which numerous NBT were collected and sent abroad, mainly to NHBC in Europe and North America, hampering the effort to catalog and describe the yet-to-be-discovered diversity(*10*, *11*). In this study, we argue that a fairer distribution of NBT is necessary if we aim to effectively fill the gaps in biological knowledge by cataloging, organizing, and describing biodiversity in the coming years. Achieving a fairer distribution of biodiversity knowledge requires understanding the current situation of NBT distribution.

We addressed this challenge by providing a comprehensive overview of the geographical and temporal (historical) distribution of freshwater and brackish fish NBT. First, we mapped, in space and time, all freshwater/brackish fish fauna NBT. Second, we classified the world regions based on the origin of the NBT they housed. Third, we evaluated the effectiveness of countries' NHBCs in representing their native freshwater fish fauna. Finally, we examined the main factors driving the distribution of NBT among countries. These assessments allowed us to delve into the multiple implications of the current distribution of fundamental biological knowledge at the species level worldwide and how it has accumulated over historical times. We selected fishes as the study model because they constitute the most species-rich group of vertebrates. Furthermore, freshwater/brackish fishes face challenges stemming from uncertainties in their evolutionary relationships among species and constant taxonomic changes, making the assessment of NBT imperative for a better understanding of the biodiversity within the group(*12*).

NBT distribution flow in space and time

We showed that the all-time distribution of NBT is strongly biased towards ichthyological collections housed in European (ECA) and North American (NA) countries, accounting for 68% of the world's total freshwater/brackish NBT. The source–housing of fish NBT among countries started in the 18th century, with ECA and NA museums acting as the primary destination for all fish NBT sourced from countries across all world regions. Our historical analysis of NBT flow also revealed the changing dominance of European countries in housing NBT between the 19th and 20th centuries, alongside the emergence of the United States as a new primary region for housing NBT. The dominance of European countries and North America in housing freshwater/brackish fish NBT peaked between 1900 and 1949, with 5051 (89% in the period) of all NBT housed within these regions. Only in the past 23 years we noticed a tendency to retain the NBT in NHBC within their source countries (Fig. 1, 2000-present). While NHBCs from the global south emerged in the last 50 years, housing essentially local NBT, global north NHBCs still keep the historical pattern of housing foreign NBT.

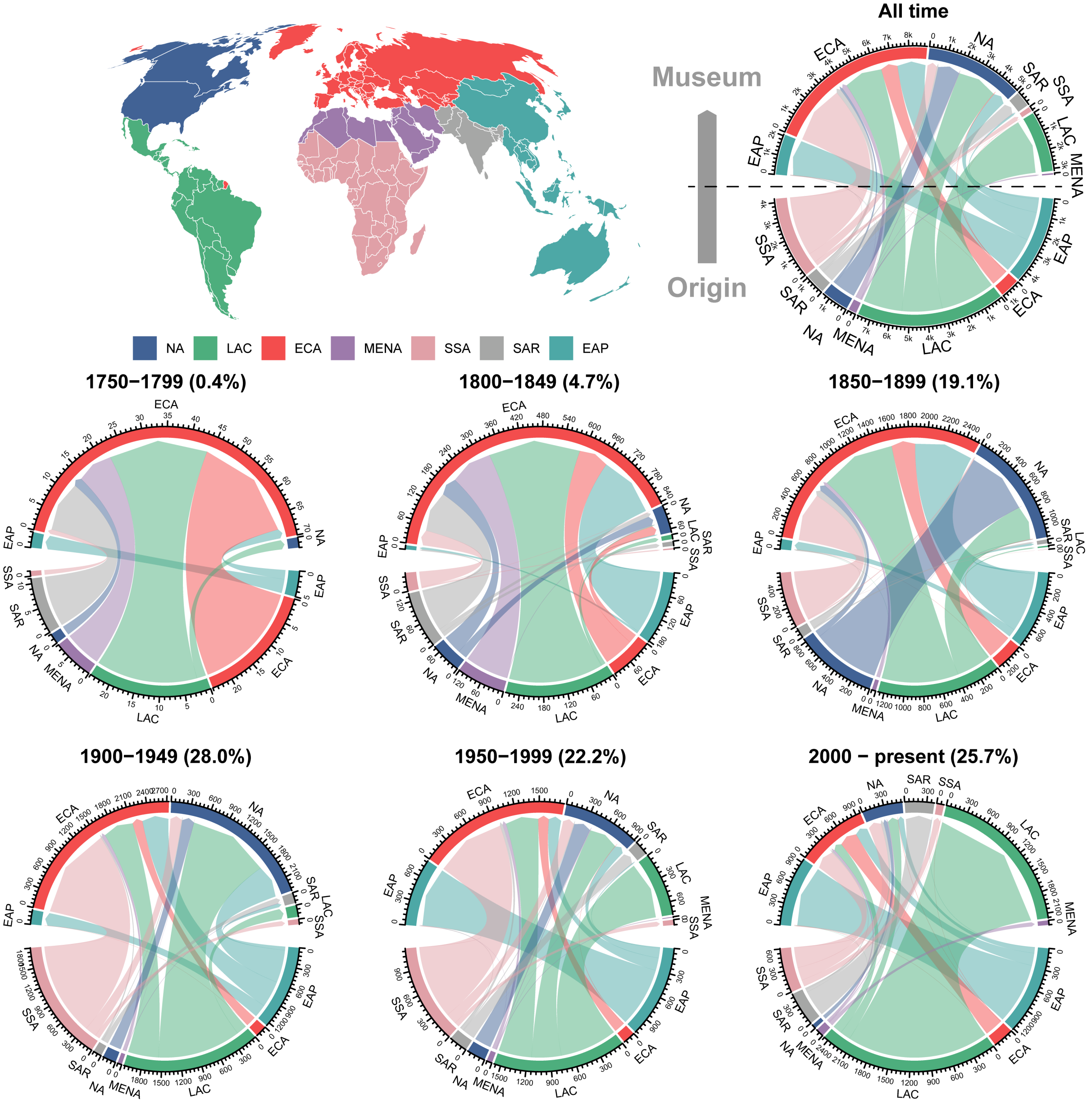


Fig. 1: NBT flow among world regions (colored according to World Bank classification as illustrated in the upper left cartogram) depicted in 50-year time intervals. The lower half of the circular plots represent the regions where the Name-Bearing Types (NBT) were collected (source regions), and the upper half is the region of the ichthyological collection in which the NBT is housed (housing regions). The numbers on the outer circles represent the total number of NBT exchanged between each region in each period. The percentage besides the year corresponds to the percentage of NBT of that period relative to the total NBT in all periods. The region acronyms are 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

Characteristics of world regions based on the source region of NBT

We identified two categories of regions regarding the sources of NBT within NHBCs. These categories were classified according to Domestic Contribution (DC) and Domestic Retention (DR) of NBT, which represent, respectively, the contribution of the local samples to all NBT in a region (DC) and the ability of a source region to retain the NBT collected within its territory (DR). ECA and NA are characterized mainly by high rates of DR (all-time 95% and 90%, respectively) but lower DC of NBT (all-time 10% and 25%, respectively), as a significant proportion of their housed NBT were sampled in other regions. Meanwhile, the all-time DC rate in LAC, MENA, SSA, EAP, and SA corresponds to 99.9%, 98%, 98%, 97%, and 86%. In contrast, the all-time DR for these regions is below 50%, except in SA (63%) (See fig. S3). In recent years, we have not found an increase in DC in these economically developed regions (ECA and NA), mainly because most of the native freshwater/brackish fish species from these regions have already been described (bottom right plot in Fig. 2).

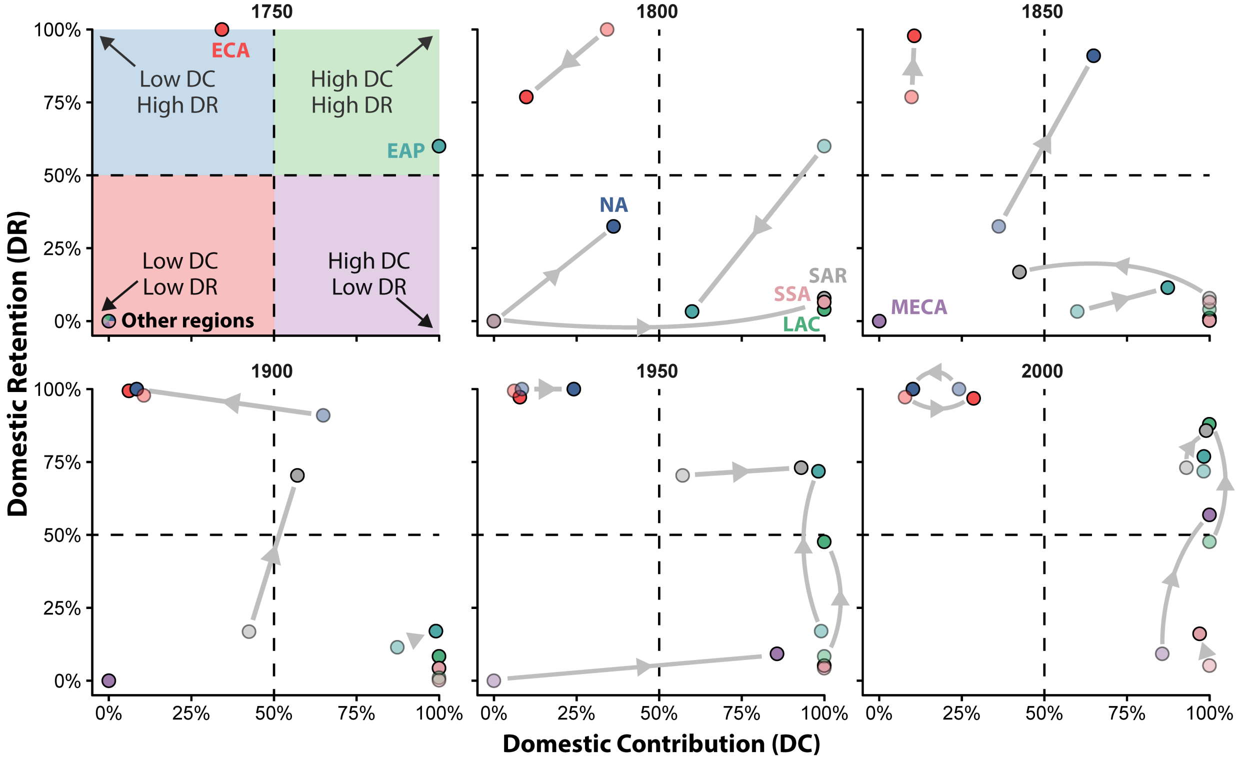


Fig. 2: Region characteristics based on the Domestic Contribution (DC) and Domestic Retention Rate (DR) of name-bearing types (NBT) depicted by 50-year time intervals (each plot). The dots and colors represent different regions, according to the World Bank (Fig. 1). Arrows indicate the direction of change regarding DC and DR from one time slice to the next. Faded dots represent the position of the region in the previous time slice.

Mismatches between NBT and native species composition

One important aspect of biodiversity knowledge distribution, regarding its uses and gains for the local scientific and general community, is whether the native biodiversity is sufficiently represented in NHBC within the country. Therefore, we computed the native turnover to express the extent to which native fish fauna of a country is underrepresented by the NBT housed there

(Fig. 3a). Conversely, NBT turnover was computed to represent how much the NBT housed within a country corresponds to non-native fish fauna (Fig. 3b). For this analysis, we used only species with one record per country (i.e., endemic species regarding country occurrence according to the Catalog of Fishes). We restricted this analysis to species with country-level endemicity to avoid possible confounding effects of large-range species in our estimates (but see fig. S3).

Countries in ECA exhibited an overall pattern of lower shortfall in representing their native fish fauna and an overrepresentation of non-native NBT compared to the native collection (Fig. 3a and b). This reinforces our findings that most Global North countries can retain their native NBT while housing foreign NBT. In contrast, most countries in the Global South presented a high shortfall of native species in their NBT collections (Fig. 3a) since most native NBTs are not housed in their ichthyological collections. They also showed low values of NBT turnover (Fig. 3b), as their NBT are primarily composed of native species. Therefore, most countries in the Global South present an overall characteristic of insufficient representation of their native NBT collections, whereas countries in the Global North present high rates of non-native NBT (Fig. 3c).

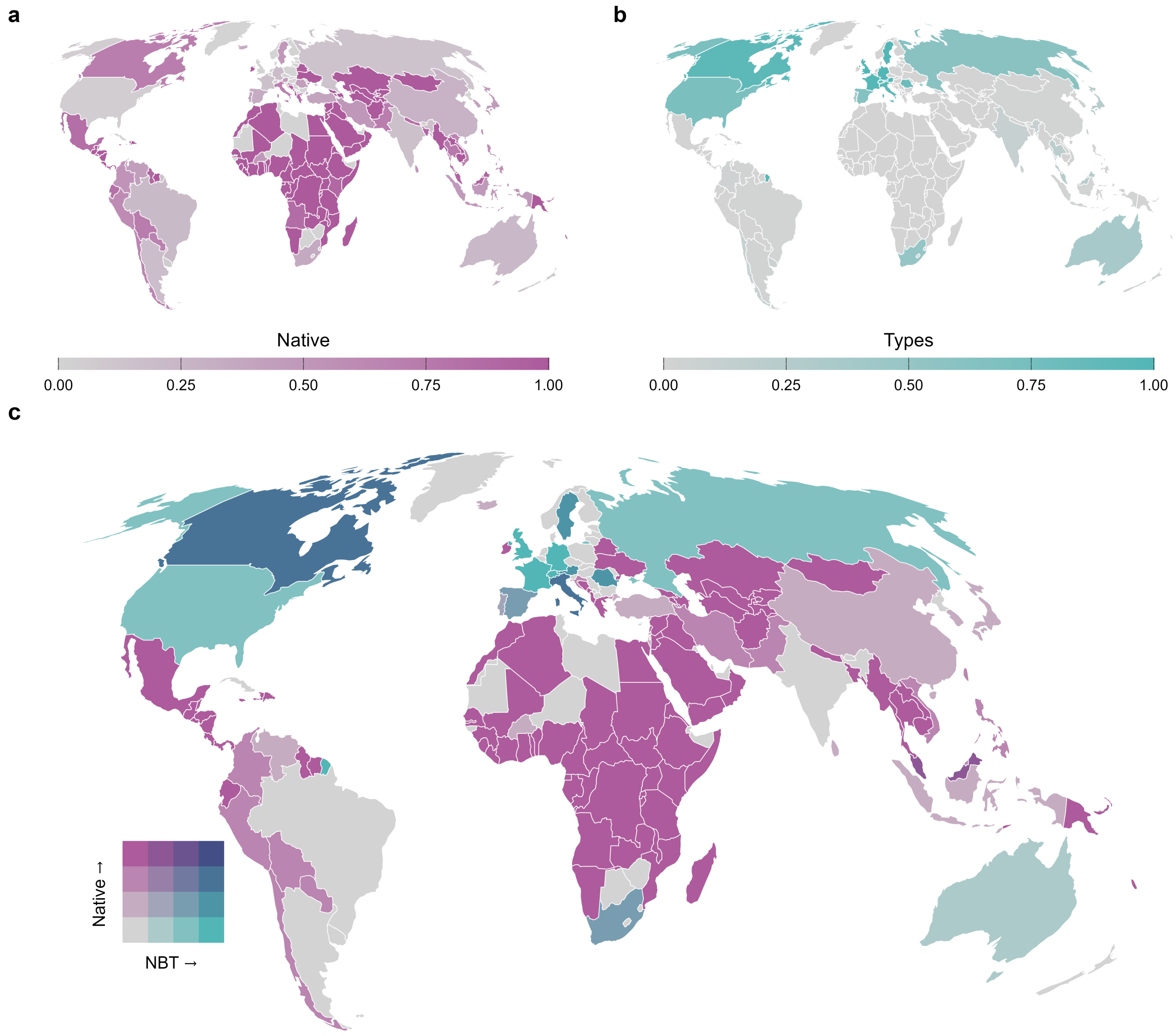


Fig. 3: World cartogram representing Native turnover (a) and Name Bearing Types (NBT) turnover (b) of species for each country. Both values are combined in a bivariate map (c), where the breaks represent 25%, 50%, and 75% of the data. Values closer to one indicate higher divergence between the two sets (native fauna and NBT composition) of a given country.

Factors explaining NBT distribution in collections around the world

Our models demonstrated that socio-economic and biological factors strongly shape NBT concentration, distribution, and the characteristics of the country´s NHBC. Specifically, the total number of NBT within the countries is driven by gross domestic product (GDP) and native species richness (Fig. 4a -b).

The characteristics of the country's biological collections regarding the source of NBT (DC and DR rates) housed in their collections are also explained by GDP. Countries with higher GDP present greater proportions of NBT sourced from other countries (low DC) combined with higher retention of NBT sampled within the country (high DR). In contrast, countries with lower GDP present a higher proportion of NBT collected within the country (high domestic contribution - DC, Figure 4e) and low domestic retention (low DR, Fig. 4f). Nonetheless, countries with higher GDP tend to have a smaller shortfall in their native NBT (low values of native turnover) in their biological collections than countries with lower GDP. Additionally, wealthier countries (high GDP) often present an overrepresentation of non-native NBT in their ichthyological collections (high values of NBT turnover) (Fig. 4d and 4f). Figure 4 represents the most important variables influencing the response variables in our model (except for the total number of NBT, where we showed the top two variables). Full results of the models are presented in Supplementary material (tables S1 – S5).

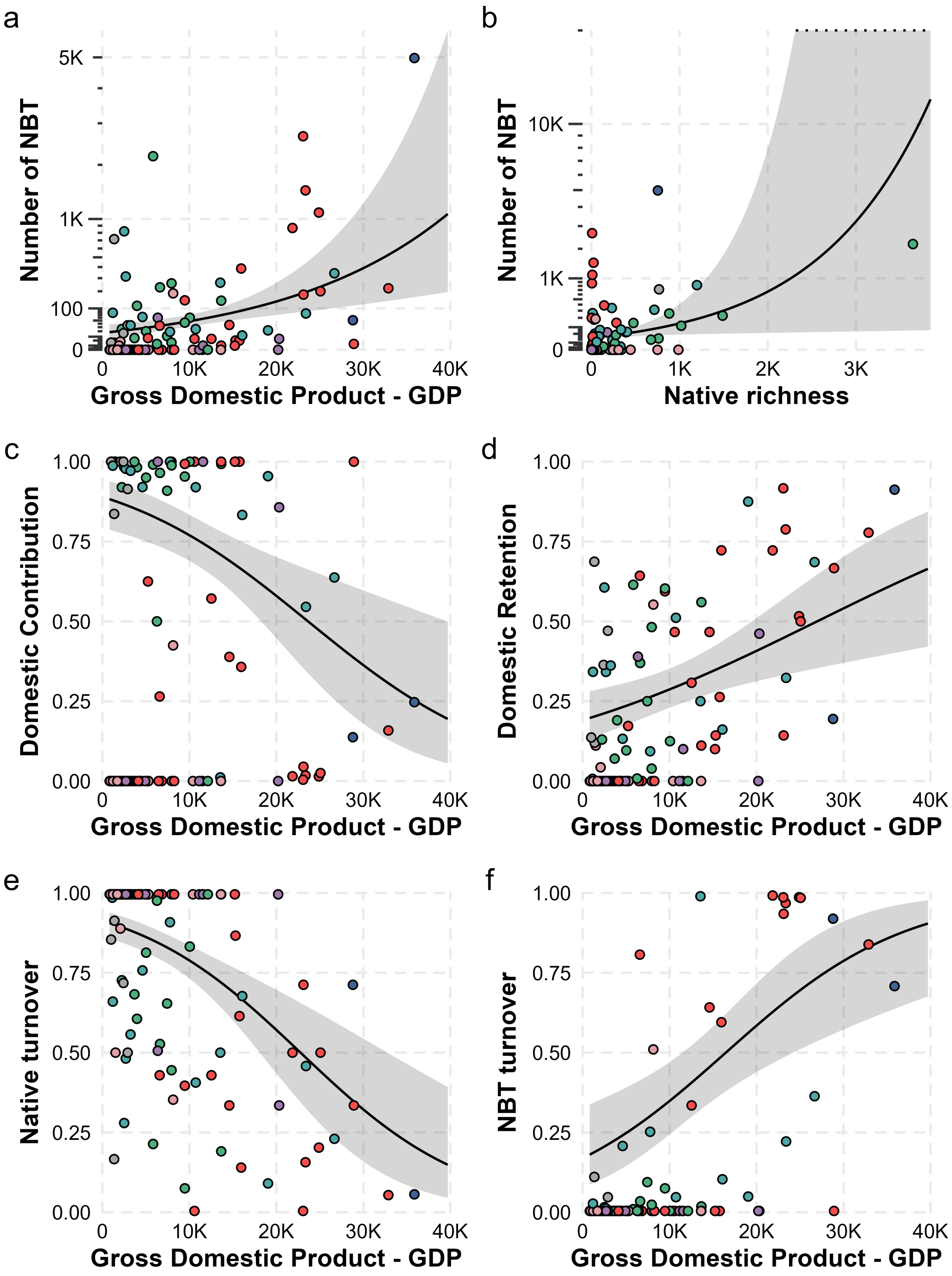


Fig. 4: Factors explaining different aspects of NBT accumulation among world countries. Relationship between the total number of NBT by country and Gross Domestic Product - GDP (a) and native species richness (b); Domestic Contribution (DC) (c) and Domestic Retention (DR)(d) and GDP; finally, the relationship between Native turnover (e) and NBT turnover(f)

with GDP. All the figures represent the marginal effects of the predictor variables over the response variables.

Discussion

We described the extent of biases in the concentration of fundamental reference for species identification for one of Earth's most diverse biological groups. By recognizing the species as the fundamental unit, we assume that institutions or countries that house most of the reference material about species also concentrate the fundamental biological knowledge. Despite critiques about the concept of species(*13*, *14*), it remains the fundamental unit of almost all investigations in biology and biodiversity conservation studies(*15*).

The arguments for a fairer and more equitable distribution of knowledge extend beyond ethical grounds. To better understand biodiversity patterns, we must address the biases in the global distribution of biodiversity knowledge(*16*). This uneven distribution of NBT has significant implications for scientific practice and our perception of biodiversity patterns.

A glimpse into the biological knowledge accumulation process

Countries with colonial histories and early concentration of economic power have accumulated NBT primarily from tropical countries in Latin America and Africa. The dynamic of Global South countries serving as providers of NBT and Global North countries as their custodians only began to change in the 21st century, with increasing retention of specimens in the countries where they were collected. For instance, the first NBT cataloged in a Brazilian museum(*17*) was almost 400 years after the country's foundation and more than 140 years after Linnaeus's binomial nomenclature system(*18*). However, since this is an ongoing accumulation process, most early discovered freshwater fish species are still housed in museums far from their source location.

We recognize that the lack of infrastructure in Global South countries and that most naturalists were not based in the colonies, leading them to send material to their home countries, contributing to the early establishment of NBT biases toward Global North. For example, the archives at the Museum of Comparative Zoology of Harvard highlight the importance of naturalistic exploration (e.g. Thayer Expedition(*19*) in Brazil) in enhancing ichthyological collection and research. However, regardless of the reason for the uneven distribution of NBTs, this disparity has significant consequences, the most direct being the difficulty researchers from Global South countries face in accessing NBTs.

Accumulating fundamental knowledge far from its origin has detrimental consequences for scientific development, as it separates the sampling location from where validation can occur more precisely. This separation impacts our understanding of biodiversity patterns by directly affecting taxonomic descriptions and reviews, ultimately hindering accurate estimates of a region’s “true” biological diversity. Estimates of species discoveries and taxonomic revisions indicate that tropical regions have higher probabilities of discovering new species in the coming years(*9*). Furthermore, a recent study on tetrapod species(*20*) (squamates, lizards, and canids) showed that taxonomic revisions are more common for species with older descriptions(*20*), as early descriptions usually lack comprehensive information. These findings – higher discovery rate in the tropics and the necessity for revising early-discovered species – along with our evidence that most of the early-discovered freshwater fish species that require revisionary studies are located far from their native range, underscore the challenges in addressing the biodiversity knowledge shortfall.

In the current NBT distribution scenario, researchers from Global South often have to travel northwards to advance their taxonomic research, or expensive research programs(*21*) must be developed and implemented to mitigate this issue. While this can be viewed as an opportunity to foster international collaboration and benefit research activities, it does not ensure symmetrical collaboration. Institutions in the Global North are likely to retain control over NBT accessibility, thereby perpetuating existing imbalances. Furthermore, while these international collaborations are positive, they require greater investment in science, which has not been the case in megadiverse countries (e.g., Brazil(*22*), Colombia(*23*)) in recent years and decades.

Implications for biodiversity research

The direct consequences of uneven distribution of fundamental biological knowledge are more pronounced on taxonomy, especially in tropical countries that harbor most of the freshwater fish diversity in the world (fig. S1 in Supplementary Material). NBT in museums and biological collections are the basic unit of investigation for taxonomists, making the accessibility of name-bearing type material crucial. Despite recent proposals advocating for the use of alternative materials, such as tissues and digitized images, to improve the accessibility of museum specimens (e.g. (*4*, *24*, *25*)), we argue that this is not a comprehensive solution to fill biodiversity gaps(*26*). Keeping the current distribution of biological knowledge also fails to contribute to a more equitable, fair, and efficient improvement of our biodiversity knowledge(*16*). Our data reinforce that addressing biodiversity challenges will require more than technocratic fixes(*16*); it necessitates overcoming long-standing and pervasive exploration practices to find solutions that rectify colonial legacies and provide a more equitable scientific assessment of this material(*27*).

Taxonomy faces inherent challenges(*28*) in identifying species and determining the best criteria to delimit them(*29*), coupled with the decline of taxonomy specialists(*30*). While funding and training for taxonomists are essential to overcome these issues(*31*), we argue that even if these challenges were solved, the problem of uneven distribution of NBT worldwide and its consequent impacts would persist.

Ethical and legal implications

Beyond its effects on our understanding of the natural world and solutions to practical problems, the concentration of NBT raises ethical and legal issues. Evidence, primarily from paleontological material, suggests instances of illegal samplings(*6*, *32*). Even when material leaves its country of origin legally, ethical concerns persist regarding this practice due to challenges in tracing the conditions under which biological material is collected and exported (e.g. parachute science(*7*)). To illustrate the neglection of this issue, the Nagoya Protocol, an important international agreement aimed at fair and equitable sharing, only encompasses the genetic component of biodiversity, ignoring the benefits associated with housing crucial reference material for local research development and educational purposes, among others. Therefore, critical documents like the Nagoya Protocol should be revised if the international community is genuinely committed to a fairer and more equitable sharing of biodiversity benefits.

Overcoming barriers to NBT distribution

Recent studies highlighted the importance of digitization of museum specimens to overcome the accessibility barriers to primary-type material. Despite the importance of those initiatives, with very successful examples in herbarium collections(*33*), not all specimens can be easily digitized(*3*). Fishes are a good example of how challenging the digitization of specimens can be since it requires high-quality and detailed images of anatomical structures in order to reach unambiguous species identification(*34*). Important diagnostic characters, such as internal structures like bones and swim bladders, cannot be captured in digitized images without high-quality 3D images. Additionally, the digitization of natural collections may replicate the inequality and the inaccessibility of data created by a colonial past that shaped these collections(*35*).

The ideal scenario would involve returning representative NBT to the NHBC of the countries where they were collected. However, this requires lengthy bilateral negotiations, even when the material was illegally taken(*6*). Repatriation often demands more than legal evidence; public pressure is also necessary(*36*). Moreover, recipient collections must have the infrastructure to maintain the biological material long-term, which may be lacking in some Global South countries. We propose a better division of labor and costs associated with a global goal of describing and protecting biodiversity(*27*, *37*). This approach would prevent delays in improving biodiversity knowledge and its benefits(*38*). An alternative would be to facilitate the accessibility of NBT by GN funding agencies covering the costs for researchers from the countries of origin to visit NBT housed in foreign collections(*37*).

Finally, protective laws and regulations can help maintain, protect, and preserve important biological samples within the country where they were collected. For instance, Brazil implemented regulations in 1990(*39*) mandating that significant biological reference materials (e.g., holotype, syntypes) be retained within the country, completely or partially in the case of paratypes. This measure has contributed to increased NBT retention in Brazil during the 21st century, as evidenced by the substantial portion of NBT from Brazil shown in Figure 1.

Conclusion

Our results, using fish 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 catalyzes developing solutions that distribute biological knowledge more equitably and fairly without mirroring our society's existing economic and social inequalities.

References and Notes

1. A. V. Suarez, N. D. Tsutsui, The Value of Museum Collections for Research and Society. *BioScience* **54**, 66 (2004).

2. R. Sluys, Attaching Names to Biological Species: The Use and Value of Type Specimens in Systematic Zoology and Natural History Collections. *Biol Theory* **16**, 49–61 (2021).

3. E. E. Gutiérrez, R. H. Pine, Specimen collection crucial to taxonomy. *Science* **355**, 1275–1275 (2017).

4. G. Nelson, S. Ellis, The history and impact of digitization and digital data mobilization on biodiversity research. *Phil. Trans. R. Soc. B* **374**, 20170391 (2019).

5. M. W. Nachman, E. J. Beckman, R. C. Bowie, C. Cicero, C. J. Conroy, R. Dudley, T. B. Hayes, M. S. Koo, E. A. Lacey, C. H. Martin, J. A. McGuire, J. L. Patton, C. L. Spencer, R. D. Tarvin, M. H. Wake, I. J. Wang, A. Achmadi, S. T. Álvarez-Castañeda, M. J. Andersen, J. Arroyave, C. C. Austin, F. K. Barker, L. N. Barrow, G. F. Barrowclough, J. Bates, A. M. Bauer, K. C. Bell, R. C. Bell, A. W. Bronson, R. M. Brown, F. T. Burbrink, K. J. Burns, C. D. Cadena, D. C. Cannatella, T. A. Castoe, P. Chakrabarty, J. P. Colella, J. A. Cook, J. L. Cracraft, D. R. Davis, A. R. D. Rabosky, G. D'Elía, J. P. Dumbacher, J. L. Dunnum, S. V. Edwards, J. A. Esselstyn, J. Faivovich, J. Fjeldså, O. A. Flores-Villela, K. Ford, J. Fuchs, M. K. Fujita, J. M. Good, E. Greenbaum, H. W. Greene, S. Hackett, A. Hamidy, J. Hanken, T. Haryoko, M. T. Hawkins, L. R. Heaney, D. M. Hillis, B. D. Hollingsworth, A. D. Hornsby, P. A. Hosner, M. Irham, S. Jansa, R. A. Jiménez, L. Joseph, J. J. Kirchman, T. J. LaDuc, A. D. Leaché, E. P. Lessa, H. López-Fernández, N. A. Mason, J. E. McCormack, C. D. McMahan, R. G. Moyle, R. A. Ojeda, L. E. Olson, C. K. Onn, L. R. Parenti, G. Parra-Olea, B. D. Patterson, G. B. Pauly, S. E. Pavan, A. T. Peterson, S. Poe, D. L. Rabosky, C. J. Raxworthy, S. Reddy, A. Rico-Guevara, A. Riyanto, L. A. Rocha, S. R. Ron, S. M. Rovito, K. C. Rowe, J. Rowley, S. Ruane, D. Salazar-Valenzuela, A. J. Shultz, B. Sidlauskas, D. S. Sikes, N. B. Simmons, M. L. J. Stiassny, J. W. Streicher, B. L. Stuart, A. P. Summers, J. Tavera, P. Teta, C. W. Thompson, R. M. Timm, O. Torres-Carvajal, G. Voelker, R. S. Voss, K. Winker, C. Witt, E. A. Wommack, R. M. Zink, Specimen collection is essential for modern science. *PLOS Biology* **21**, e3002318 (2023).

6. J. C. Cisneros, A. M. Ghilardi, N. B. Raja, P. P. Stewens, The moral and legal imperative to return illegally exported fossils. *Nat Ecol Evol* **6**, 2–3 (2021).

7. N. B. Raja, E. M. Dunne, A. Matiwane, T. M. Khan, P. S. Nätscher, A. M. Ghilardi, D. Chattopadhyay, Colonial history and global economics distort our understanding of deep-time biodiversity. *Nat Ecol Evol* **6**, 145–154 (2021).

8. B. G. Freeman, M. W. Pennell, The latitudinal taxonomy gradient. *Trends in Ecology & Evolution* **36**, 778–786 (2021).

9. M. R. Moura, W. Jetz, Shortfalls and opportunities in terrestrial vertebrate species discovery. *Nature Ecology and Evolution* **5**, 631–639 (2021).

10. A. G. Burman, THE HOMEOTYPE: A PLEA FOR RESPECTABILITY. *TAXON* **35**, 317–321 (1986).

11. M. S. Vorontsova, P. P. Lowry, S. R. Andriambololonera, L. Wilmé, A. Rasolohery, R. Govaerts, S. Z. Ficinski, A. M. Humphreys, Inequality in plant diversity knowledge and unrecorded plant extinctions: An example from the grasses of Madagascar. *Plants People Planet* **3**, 45–60 (2021).

12. R. E. Reis, J. S. Albert, F. Di Dario, M. M. Mincarone, P. Petry, L. A. Rocha, Fish biodiversity and conservation in South America. *Journal of Fish Biology* **89**, 12–47 (2016).

13. M. C. C. De Pinna, [No title found]. *Reviews in Fish Biology and Fisheries* **9**, 353–373 (1999).

14. K. De Queiroz, Species Concepts and Species Delimitation. *Systematic Biology* **56**, 879–886 (2007).

15. R. Balakrishnan, Species Concepts, Species Boundaries and Species Identification: A View from the Tropics. *Systematic Biology* **54**, 689–693 (2005).

16. M. Chapman, B. R. Goldstein, C. J. Schell, J. S. Brashares, N. H. Carter, D. Ellis-Soto, H. O. Faxon, J. E. Goldstein, B. S. Halpern, J. Longdon, K. E. A. Norman, D. O'Rourke, C. Scoville, L. Xu, C. Boettiger, Biodiversity monitoring for a just planetary future. *Science* **383**, 34–36 (2024).

17. N. Sanjad, Emílio Goeldi (1859-1917) e a Institucionalização das Ciências Naturais na Amazônia. *Rev. Bras. Inov.* **5**, 455 (2006).

18. C. V. Linné, *Species Plantarum : Exhibentes Plantas Rite Cognitas Ad Genera Relatas, Cum Diferentiis Specificis, Nominibus Trivialibus, Synonymis Selectis, Locis Natalibus, Secundum Systema Sexuale Digestas / Caroli Linnæe.* (Junk, Berlin :, 1908; http://www.biodiversitylibrary.org/bibliography/37656).

19. A. R. Wallace, 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 (1870).

20. J. J. M. Guedes, M. R. Moura, J. Alexandre F. Diniz-Filho, Species out of sight: elucidating the determinants of research effort in global reptiles. *Ecography* **2023**, e06491 (2023).

21. D. Serwadda, P. Ndebele, M. K. Grabowski, F. Bajunirwe, R. K. Wanyenze, Open data sharing and the Global South—Who benefits? *Science* **359**, 642–643 (2018).

22. G. W. Fernandes, M. M. Vale, G. E. Overbeck, M. M. C. Bustamante, C. E. V. Grelle, H. G. Bergallo, W. E. Magnusson, A. Akama, S. S. Alves, A. Amorim, J. Araújo, C. F. Barros, F. Bravo, M. J. V. Carim, R. Cerqueira, R. G. Collevatti, G. R. Colli, C. N. Da Cunha, P. S. D'Andrea, J. C. Dianese, S. Diniz, P. C. Estrela, M. R. M. Fernandes, C. S. Fontana, L. L. Giacomin, L. F. P. Gusmão, F. A. Juncá, A. C. B. Lins-e-Silva, C. R. A. S. Lopes, M. L. Lorini, L. P. De Queiroz, L. R. Malabarba, B. S. Marimon, B. H. M. Junior, M. C. M. Marques, B. M. Martinelli, M. B. Martins, H. F. De Medeiros, M. Menin, P. B. De Morais, F. H. Muniz, S. Neckel-Oliveira, J. A. De Oliveira, R. P. Oliveira, F. Pedroni, J. Penha, L. R. Podgaiski, D. J. Rodrigues, A. Scariot, L. F. Silveira, M. Silveira, W. M. Tomas, M. J. S. Vital, V. D. Pillar, Dismantling Brazil's science threatens global biodiversity heritage. *Perspectives in Ecology and Conservation* **15**, 239–243 (2017).

23. J. E. A. Huddart, A. J. Crawford, A. L. Luna-Tapia, S. Restrepo, F. Di Palma, EBP-Colombia and the bioeconomy: Genomics in the service of biodiversity conservation and sustainable development. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2115641119 (2022).

24. A. Q. Byrne, Reimagining the future of natural history museums with compassionate collection. *PLOS Biology* **21**, e3002101 (2023).

25. C. J. Raxworthy, B. T. Smith, Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution* **36**, 1049–1060 (2021).

26. E. F. A. Toussaint, C. P. D. T. Gillett, I. Löbl, The fate of natural history museums in the face of good intentions. *Biological Journal of the Linnean Society*, blae012 (2024).

27. J. S. Leal, B. Soares, A. C. S. Franco, R. G. de Sá Ferreira Lima, K. Baker, M. Griffiths, "Decolonising ecological research: a debate between global North geographers and global South field ecologists" (preprint, SocArXiv, 2022); https://doi.org/10.31235/osf.io/wbzh2.

28. H. C. J. Godfray, Challenges for taxonomy. *Nature* **417**, 17–19 (2002).

29. J. Hey, R. S. Waples, M. L. Arnold, R. K. Butlin, R. G. Harrison, Understanding and confronting species uncertainty in biology and conservation. *Trends in Ecology & Evolution* **18**, 597–603 (2003).

30. O. Paknia, H. Rajaei Sh., A. Koch, Lack of well-maintained natural history collections and taxonomists in megadiverse developing countries hampers global biodiversity exploration. *Org Divers Evol* **15**, 619–629 (2015).

31. M. R. De Carvalho, F. A. Bockmann, D. S. Amorim, C. R. F. Brandão, M. De Vivo, J. L. De Figueiredo, H. A. Britski, M. C. C. De Pinna, N. A. Menezes, F. P. L. Marques, N. Papavero, E. M. Cancello, J. V. Crisci, J. D. McEachran, R. C. Schelly, J. G. Lundberg, A. C. Gill, R. Britz, Q. D. Wheeler, M. L. J. Stiassny, L. R. Parenti, L. M. Page, W. C. Wheeler, J. Faivovich, R. P. Vari, L. Grande, C. J. Humphries, R. DeSalle, M. C. Ebach, G. J. Nelson, Taxonomic Impediment or Impediment to Taxonomy? A Commentary on Systematics and the Cybertaxonomic-Automation Paradigm. *Evol Biol* **34**, 140–143 (2007).

32. J. C. Cisneros, N. B. Raja, A. M. Ghilardi, E. M. Dunne, F. L. Pinheiro, O. R. Regalado Fernández, M. A. F. Sales, R. A. Rodríguez-de La Rosa, A. Y. Miranda-Martínez, S. González-Mora, R. A. M. Bantim, F. J. De Lima, J. D. Pardo, Digging deeper into colonial palaeontological practices in modern day Mexico and Brazil. *R. Soc. open sci.* **9**, 210898 (2022).

33. Brazilian Flora 2020: Innovation and collaboration to meet Target 1 of the Global Strategy for Plant Conservation (GSPC). *Rodriguésia* **69**, 1513–1527 (2018).

34. M. A. Kolmann, M. Kalacska, O. Lucanus, L. Sousa, D. Wainwright, J. P. Arroyo-Mora, M. C. Andrade, Hyperspectral data as a biodiversity screening tool can differentiate among diverse Neotropical fishes. *Sci Rep* **11**, 16157 (2021).

35. Kaiser, K, Heumann, I, Nadim, T, Keysar, H, Petersen, M, Korun, M, Berger, F, Promises of mass digitisation and the colonial realities of natural history collections., *Journal of Natural Science Collections*pp. 13–25.

36. M. Lenharo, M. Rodrigues, How a Brazilian dinosaur sparked a movement to decolonize fossil science. *Nature* **605**, 18–19 (2022).

37. M. E. Kemp, A. E. Boville, C. M. Carneiro, J. J. Jacisin, C. J. Law, D. T. Ledesma, A. Meza, A. Shields-Estrada, T. Xu, Looking Back for the Future: The Ecology of Terrestrial Communities Through the Lens of Conservation Paleobiology. *Annu. Rev. Ecol. Evol. Syst.* **54**, 259–282 (2023).

38. K. Nanglu, D. de Carle, T. M. Cullen, E. B. Anderson, S. Arif, R. A. Castañeda, L. M. Chang, R. E. Iwama, E. Fellin, R. C. Manglicmot, M. D. Massey, V. Astudillo-Clavijo, The nature of science: The fundamental role of natural history in ecology, evolution, conservation, and education. *Ecology and Evolution* **13**, e10621 (2023).

39. Diário oficial da união, Aprova o Regulamento sobre coleta, por estrangeiros, de dados e materiais científicos do Brasil.

40. Fricke, R, Eschmeyer, W. N., Van der Laan, R, ESCHMEYER’S CATALOG OF FISHES: GENERA, SPECIES, REFERENCES., version Eletronic version accessed 01/05/2024; ESCHMEYER'S CATALOG OF FISHES: GENERA, SPECIES, REFERENCES.

41. ICZN, International Commission of Zoological Nomenclature 4th edition (1999). http://www.iczn.org/iczn/index.jsp.

42. E. Fricke, R W. N., ESCHMEYER’S CATALOG OF FISHES: GUIDE TO FISH COLLECTIONS. (http://researcharchive.calacademy.org/research/ichthyology/catalog/collections.asp.

43. V. Arel-Bundock, N. Enevoldsen, C. Yetman, countrycode: An R package to convert country names and country codes. *JOSS* **3**, 848 (2018).

44. A. Baselga, Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography* **19**, 134–143 (2010).

45. H. Schielzeth, Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* **1**, 103–113 (2010).

46. S. Ferrari, F. Cribari-Neto, Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics* **31**, 799–815 (2004).

47. Florian Hartig, DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models, version 0.4.6; https://CRAN.R-project.org/package=DHARMa.

48. Ronald Fricke, William N. Eschmeyer, Richard van der Laan, Eschmeyer’s catalog of fishes: genera, species, references, version Eletronic version accessed 01 May 2024; http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

49. Hadley Wickham, Romain François, Lionel Henry, Kirill Muller, Davis Vaughan, dplyr: A Grammar of Data Manipulation., version R package version 1.1.4; https://dplyr.tidyverse.org/.

50. R. Abell, M. L. Thieme, C. Revenga, M. Bryer, M. Kottelat, N. Bogutskaya, B. Coad, N. Mandrak, S. C. Balderas, W. Bussing, M. L. J. Stiassny, P. Skelton, G. R. Allen, P. Unmack, A. Naseka, R. Ng, N. Sindorf, J. Robertson, E. Armijo, J. V. Higgins, T. J. Heibel, E. Wikramanayake, D. Olson, H. L. López, R. E. Reis, J. G. Lundberg, M. H. Sabaj Pérez, P. Petry, Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience* **58**, 403–414 (2008).

51. Daniel Lüdecke, sjPlot: Data Visualization for Statistics in Social Science, version R package version 2.8.16; https://CRAN.R-project.org/package=sjPlot.

Acknowledgments: We would like to thank Renata Rubia Ota for discussion in early versions of this study.

Funding: National Institute of Science and Technology – Ecology, Evolution and Conservation Biology (INCT - EECBio) grant 465610/2014-5

Author contributions: Each author’s contribution(s) to the paper should be listed [we encourage you to follow the CRediT model]. Each CRediT role should have its own line, and there should not be any punctuation in the initials.

Examples:

Conceptualization: GN and LFF

Methodology: GN, BM, LFF, MAL, MSL, LD and AR

Investigation: GN, BM, MAL, LFF, EBN and MMFM

Visualization: GN and BM

Funding acquisition: JAFDF

Project administration: JAFDF

Supervision: JAFDF, WJG and BES

Writing – original draft: GN, BM and LFF

Writing – review & editing: GN, BM, LFF, MAL, EBN, MMFM, MSL, LD, WGJ, TMSF, BES, JAFDF

Diversity, equity, ethics, and inclusion [optional]: We are a group of researchers working mainly in the Global South. Our views are mainly influenced by our work experience in ecology and taxonomy in the Global South, mainly Brazil.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: The primary data used in this paper is freely available from Eschemeyer’s Catalogue of Fishes. CAS also provided information on NBT 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). Global Biodiversity Information Facility (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 modeling approach. The supplementary material provides a detailed overview of the data used. All data and code used to reproduce the figures and models can be found at https://doi.org/10.5281/zenodo.13307474. The source code and a reproducible workflow for all analyses can be found at https://gabrielnakamura.github.io/NBT_code_data/.

Supplementary Materials

Materials and Methods

Supplementary Text

Figs. S1 to S5

Tables S1 to S5

References (*40-51*)

Science
AAAS

Supplementary Materials for

The macroecology of knowledge: Spatio-temporal patterns of name-bearing types in biodiversity science

Authors: Gabriel Nakamura[1,2*]§, Bruno Henrique Mioto Stabile[3]§, Lívia Estéfane F. Frateles[4]§, Matheus Lima Araujo[4], Emanuel Neuhaus[5], Manoela Maria Ferreira Marinho[6], Melina de Souza Leite[1], Aline Richter[7], Liuyong Ding[8], Tiago Magalhães da Silva Freitas[9], Bruno Eleres Soares[10], Weferson Júnio da Graça[4,11], José Alexandre Felizola Diniz-Filho[2,4]

Corresponding author: gabriel.nakamura.souza@gmail.com

The PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S5
References

Materials and Methods

Data acquisition

We compiled all the species names, including valid names and synonyms, of freshwater and brackish fish species using the Eschmeyer´s Catalog of Fishes (CAS)(*40*), the most updated curated database for fishes. Valid names and synonyms were included since both are necessary for taxonomic surveys. We call all these specimens as NBT (name-bearing types), following the definition of the International Code of Zoological Nomenclature(*41*).

For each NBT, we extracted the source country and the museum's name where the same NBT is housed. Based on the acronyms listed in CAS(*42*), we obtained the country where each museum/biological collection is located. If the origination country of the NBT was not specified, we checked CAS for additional information, such as coordinates or any other geographical location. All the searches in CAS were performed between March 2023 and May 2024. The final data set comprises 20,246 NBT names for freshwater/brackish fishes.

Metrics of country's NBT characterization

We mapped the country in which each NBT was collected (source country) and where it is currently housed (housing country). We characterized the flow of NBT in 50-year intervals, grouping the countries according to the World Bank classification obtained from R package countrycode(*43*) (fig. S1). We chose to use the WB classification of countries since it better represents the world's regions based on the overall infrastructure and economic characteristics that are pivotal in determining the distribution of NBT. 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 North America under other classifications).

We calculated two metrics to represent regions' characteristics based on NBT source and housing region. The first one is called **Domestic Contribution (DC)** and expresses the fraction of the total number of NBT in a region (NBT_{total}) with sampling localities (topotypes) within that region (NBT_{local}) (Equation 1). The second metric is the **Domestic Retention (DR)**, representing the proportion of NBT collected and retained ($NBT_{retained}$) in the same region where it was collected (Equation 2).

$$DC = {}^{NBT_{local}}/_{NBT_{total}} \quad \text{Equation 1}$$

$$DR = {}^{NBT_{retained}}/_{NBT_{local}} \quad \textit{Equation 2}$$

Therefore, regions with high DC are those in which most of the NBT deposited in the region were collected within the region, whereas low values of DC indicate regions in which most of the NBT were sourced from another region. Finally, DR is a proxy of the amount of NBT that a region can retain in museums and biological collections within the region. We also calculated DC and DR for each country. The results for country level were used in the modelling approach (Fig. 4c and 4d).

Congruence between native fish fauna and NBT.

Assuming that one of the roles of natural museums and ichthyological collections is to maintain a curated and organized representation of biodiversity knowledge, we wanted to explore the representativeness of NBT deposited in a country compared to the country's native species composition. It is desired that they house NBT of species found within the ecological and geographical context of which the collection is part, maximizing biodiversity's benefits in multiple aspects(*38*), such as biological research and educational purposes.

In this context, we developed two metrics slightly modifying Baselga´s turnover metric(*44*). The *Native turnover* represents the proportion of the known native species of a given country that are unrepresented within its NBT collections (i.e., native species shortfall; Equation 2). Therefore, for a given country's native composition and its NBT collection, the Native turnover will be calculated as:

$$Native\ turnover_n = {}^{Native\ out_n}\!/_{(Native\ out_n + (Native_n \cap NBT_n))}$$ (Equation 3)

where the *Native turnover* of a given country *n* is the ratio between the number of native species missing in ichthyological collections (*Native out*) and its *Native out* plus those native species housed within its NBT collection ($Native_n \cap NBT_n$). Therefore, higher values of *Native turnover* indicate a country with a high deficit in the representation of its native freshwater fish fauna in the ichthyological collections of that country, whereas low values indicate a good representation of the native fish fauna within its NBT collection.

The second metric, named *NBT turnover*, is a metric that computes the proportion in which the NBTs deposited in a given country represent non-native fish fauna (Equation 4). Therefore, for the NBT available within a country and the native species of that country, the NBT turnover is computed as follows: $NBT\ turnover_n = {}^{NBT\ non\ native\ _n}\!/_{NBT\ non\ native_n + (Native_n \cap NBT_n)}$ (Equation 4)

where the *NBT turnover* of a given country *n* is the ratio between the NBT of non-native species in the country (*NBT non* $native_n$) and the *NBT non* $native_n$ plus those native species with NBT housed in the country (Native ∩ NBT). Therefore, higher values of *NBT turnover* indicate an overrepresentation of non-native NBTs in relation to the local representation of NBT of native species. In contrast, lower values of *NBT turnover* indicate that the NBT collections within the studied country are primarily composed of native species. The number of native species for each country is represented in Figure S1 in Supplementary material.

We calculated Native turnover and NBT turnover in two ways. First, we used a dataset that considered only native species with one occurrence per country, according to the data from the Catalog of Fishes. We opted to restrict our dataset to avoid the effects of large-range species in our metrics. For example, suppose a species that occurs in multiple countries. Its NBT is housed by only one country, but this does not necessarily indicate that this country sourced the NBT from other countries, just that the country sampled the NBT first. To avoid this scenario confounding our results, we opt to a more conservative approach using only species with one occurrence per country. Second, we used the full dataset containing all native species to calculate native and NBT turnover. The result of this last analysis is shown in Supplementary material (Figure S4). Despite the differences in the datasets, the overall patterns are very similar.

Factors explaining the number of NBTs, DC, DR, native and NBT turnover by country

To explain the total number of NBTs by country, we used a Generalized Linear Model (GLM) relating the absolute number of NBT housed in each country with five explanatory variables (1) the number of native species in a country, (2) the number of samples by area in each country extracted from Global Biodiversity Information Facility (GBIF) (3), the country's Gross Domestic Product (GDP), (4) the number of ichthyological collections in each country and (5) the years after independence for the countries that have been a colony in the past. All the explanatory variables were centralized and standardized before running the model to allow comparisons between the effects of each variable in the response variables(*45*). We used a negative binomial distribution to model the residual error distribution of the total NBT. We modeled DC, DR, Native turnover, and NBT turnover using GLMs with the same explanatory variables with a beta-binomial distribution(*46*) to model the error term. To assess the validity of our models, we evaluated the homoscedasticity in the residuals against predicted values (model assessment graphics can be checked in Supplementary Material, Figures S3) and through qq-plots using a simulation approach(*47*). The predicted coefficients for all explanatory variables are provided in Tables S1 to S5 of Supplementary Material.

General overview of the freshwater/brackish fish data

We analyzed 20,246 species names of freshwater and brackish fish. According to the Catalog of Fishes(*48*) (CAS), 4,943 correspond to synonyms, and 15,303 correspond to current valid names. This information was compiled from March 2023 to May 2024; the last update in our database was made in May 2024. At that time, CAS registered 36,863 valid fish species, including freshwater, brackish, and marine species.

Finding source and housing countries

To find each register's NBT sampling locality (NBT source), we relied on description localities available in CAS and the list of countries provided in the R package *countrycode*(*43*). The descriptions of localities are not standardized in the CAS, but since the source country is usually reported, we extracted this information to map the source locality of the NBT. Given the high variability in the way the names of the countries were reported in CAS, we used regex patterns to match the source country where the NBT was collected. Housing countries were also obtained from CAS using information from museums and ichthyological collections mentioned in the main database of the Catalog of Fishes(*42*). We extracted the country of museums and ichthyological collections besides its respective World Bank region using the information in the Catalog of Fishes and the *countrycode* R package.

All data were compiled in a data frame containing information on the country locality where the NBT was collected (source) with its respective ISO 3166-1 alpha-3 country code published by the International Organization for Standardization. ISO 3166-1 alpha-3 is an administrative classification code corresponding to a three-letter country code (iso3c code in *countrycode* package) useful for standardizing countries' names. This data frame also contains the World Bank region where the country is located and the country's continent. The merging procedure was performed using *dplyr*(*49*) package functions. During the merging procedure, duplicate entries were unintentionally created since there are mentions of more than one country in some descriptions of the NBT localities from Catalog of Fishes. Some of these are "true" mentions, corresponding to NBT that the collector did not assign to a single country. Other mentions were misleading, which correspond to cases when the description mentions the name of a country but does not truly correspond to the country where the NBT was sourced (for example, river and lake with country names but located in a different country). Duplicated records corresponded to 24.9% of all data. Since some duplicated records correspond to true source locations, we performed a data mining process to assign only one source locality for each register when possible. The treatment of these duplicated records was specific for each case. When the duplicated records were in the same region (according to World Bank classification), we randomly sampled one country for the NBT source location.

The data used for all analysis contains 14,667 (72.44%) specimens as holotypes, 3,584 (17.70%) as syntypes, 1,732 (8.55%) as lectotypes, and 230 (1.13%) as neotypes. Also, we found 23 (0.11%) species with specimens as lectotype and syntypes, four (0.01%) species with specimens as syntypes and neotype, five (0.002%) species with holotype and syntype, and, finally, one (0.004%) species name with holotype and lectotype.

A glimpse into taxonomic nomenclature

Name-bearing types (NBT) 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 NBT. 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. Four distinct types of NBT are applied to fish in different cases, as shown in the fig. S1.

Richness from Catalog of Fishes

We compiled the number of native NBT for each country based on the information from CAS. We extracted the field "distribution" information from CAS for all freshwater and brackish species. The total number of species for each country is represented in fig. S2. The information on country species composition was also used in calculating Native turnover and NBT turnover presented in the main text (Fig. 3).

We decided to extract the composition of native species from the Catalog of Fishes to keep the source information consistent with the information on NBT composition in each country. This avoided possible errors due to the lack of standardization or updates in species names if the information was obtained from another database. We also extracted information on native fish distribution from FishBase, obtaining similar results regarding the richness between the two datasets (Pearson correlation of 0.89). Therefore, we decided to show only the information from the Catalog of Fishes.

Figure S1 shows that the number of native fish species is very similar to other databases that provide the richness and composition of freshwater native species, such as FEOW(*50*). We decided not to use these datasets since they were older than the data on species nomenclature we obtained from the Catalog of Fishes. Changes in nomenclature could cause mismatches between our dataset and these data sources. However, we can see that the native richness from these datasets and the one obtained from the Catalog of Fish are similar.

Overall domestic contribution and retention

We also computed the overall (considering all-time) Domestic Retention (DR) and Domestic Contribution (DC) for all WB regions, as illustrated in fig. S3. DR and DC were computed as indicated by Equations 1 and 2 in the main text.

Native and NBT turnover

We also calculated Native and NBT turnover using the full dataset of native species according to the Catalog of Fishes. Figure S4 represents the results of this analysis. Overall patterns are similar to the analysis using the dataset containing only species with one occurrence per country. The correlation between the two results, computed with the Pearson coefficient, is 0.89 for Native turnover and 0.81 for NBT turnover (fig. S4).

Ambiguities in NBT material

Here, we report NBT material with ambiguous classifications regarding the country and location where the specimen was collected. Only seven species presented more than one sample classified with Name Bearing Type (NBT) status. For example, *Chaetobranchus robustus* presents a holotype housed at the Natural History Museum of London and a syntype specimen at the Berlin Natural History Museum. Since we found only seven cases with type material in more than one country, we do not consider the distribution of different materials with type value in different countries in our analysis.

Statistical models

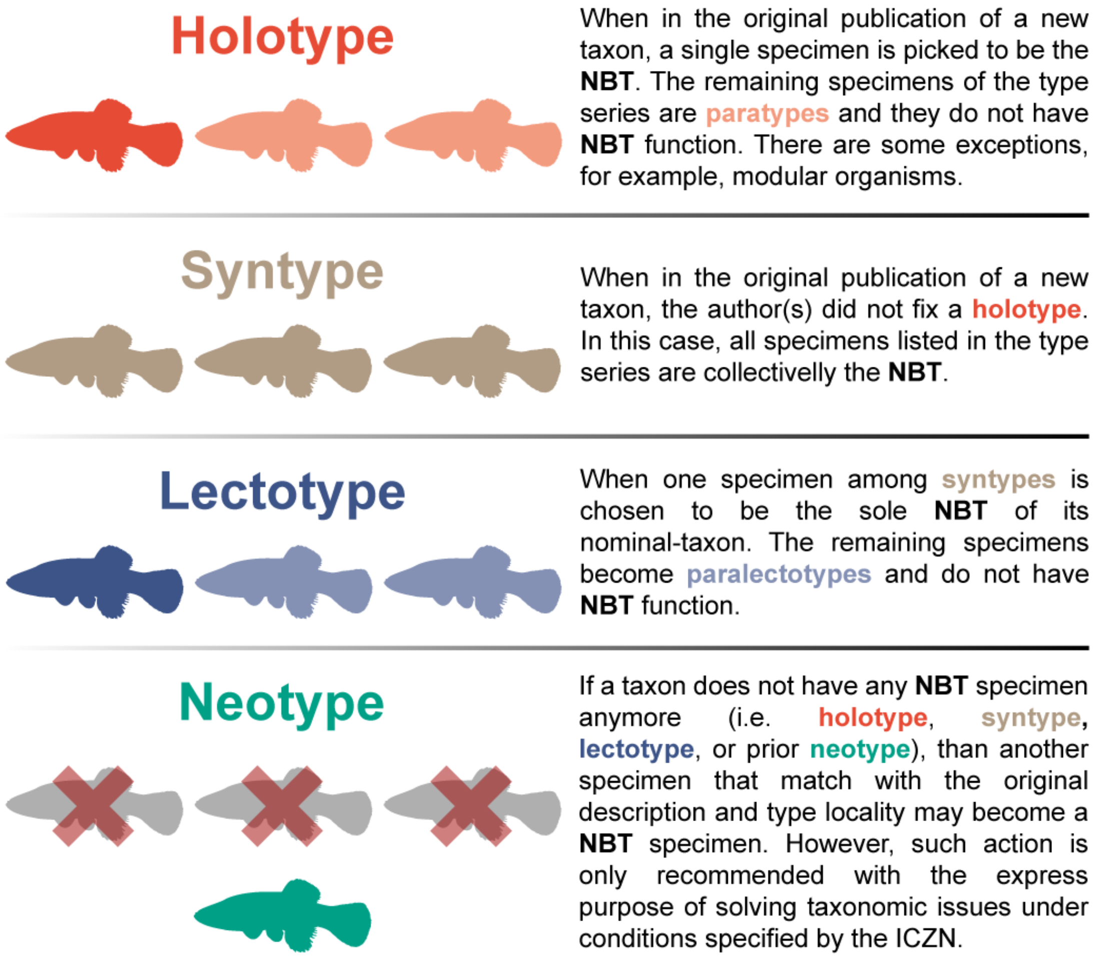
We present tables with complementary results from the models used to produce Figure 4 in the main text. All models' estimated coefficients, confidence intervals, and p-value estimates are reported below. The coefficients are all presented in the original scale of the response variable. The tables were made using the R package sjPlot(*51*).

It is worth noting that other variables also influenced the predictor variables besides GDP, which was important in explaining all the metrics. For Domestic contribution, the number of occurrence records per area, according to GBIf, had a negative influence, indicating that countries with more records have less Domestic contribution. The number of museums in a country positively influenced domestic retention. Native turnover was also negatively affected by the number of museums, which indicates that countries with more museums have more non-native NBT than countries with fewer museums.

Accessing validity of models used to explain NBT distribution

We performed residual diagnostic analysis using package DHARMa(*47*) to assess the adequacy of our models. fig. S5 corresponds to the qq-plot and residuals x predicted values relationship for each model shown in Fig. 4 of the main text. Qq-plots are used to access the overall deviation of the distribution between observed and simulated data. KS tests for the distribution correctness. Residuals x predicted plots are used to check the uniformity of residuals visually.

Fig. S1.



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

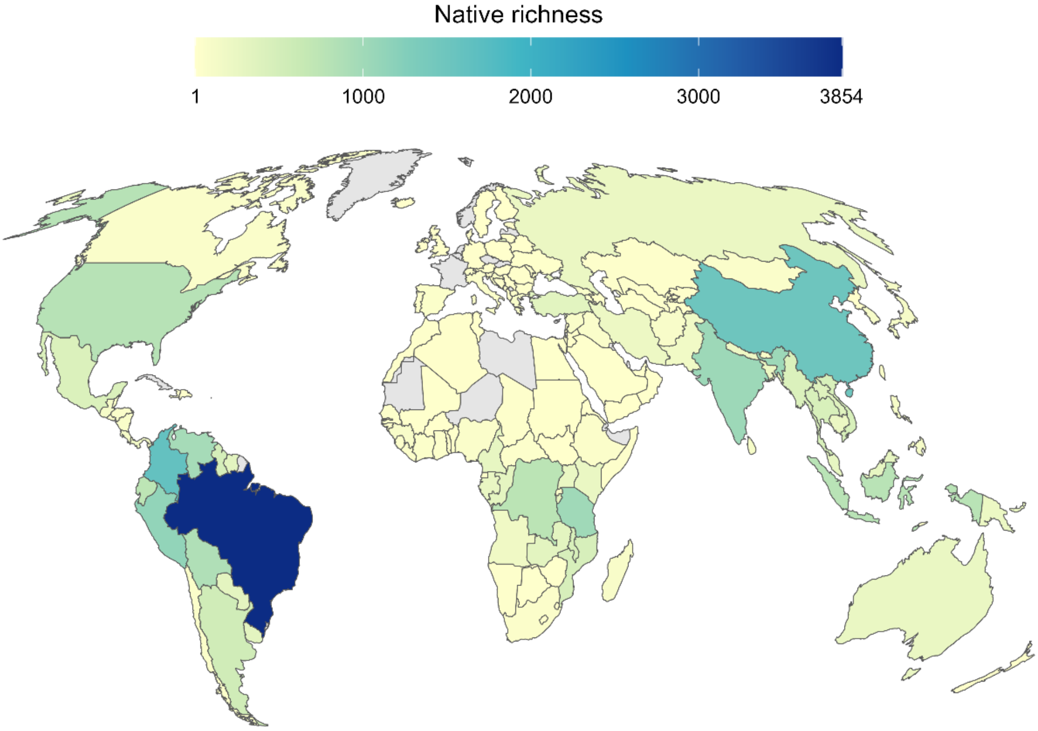
When in the original publication of a new taxon, the author(s) did not fix a **holotype**. In this case, all specimens listed in the type series are collectivelly the **NBT**.

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

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

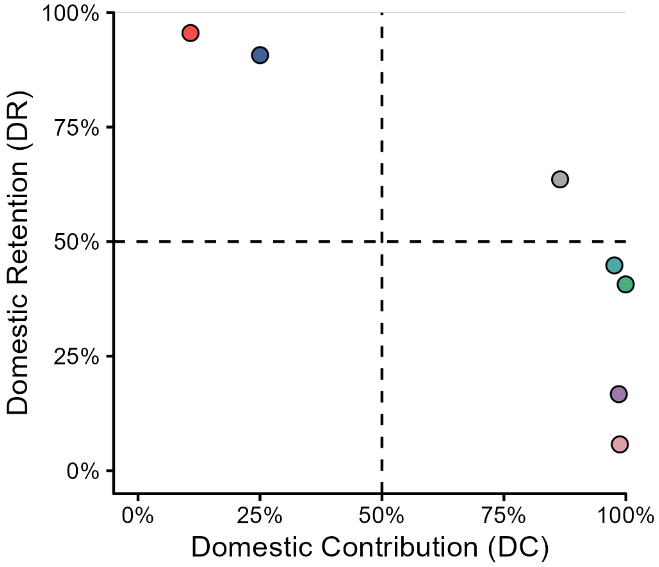
Illustration showing the specimens considered name-bearing types according to the Zoological Code of Nomenclature accompanied by a brief explanation for each category.

Fig. S2.



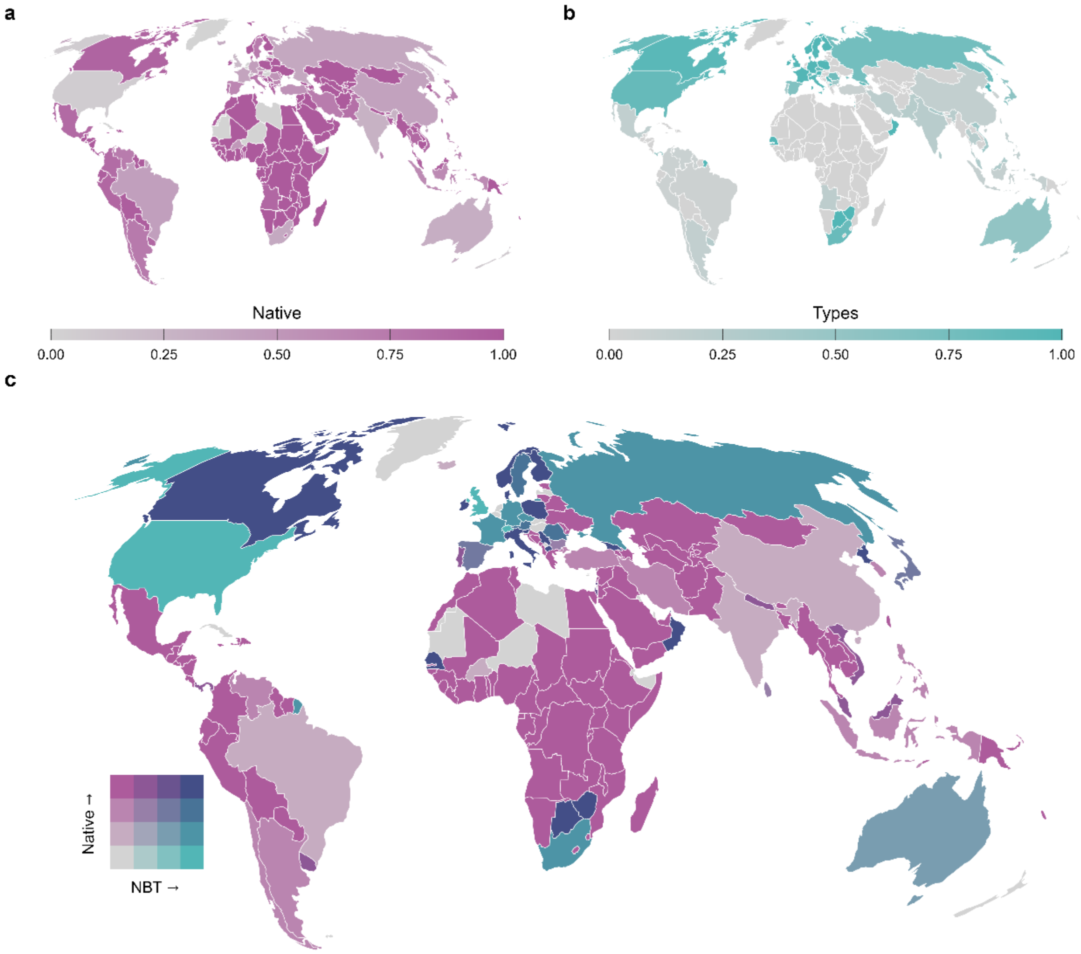
Number of native species by country according to the Catalog of Fishes. Native species were extracted from field "distribution" in the Catalog of Fishes.

Fig. S3.



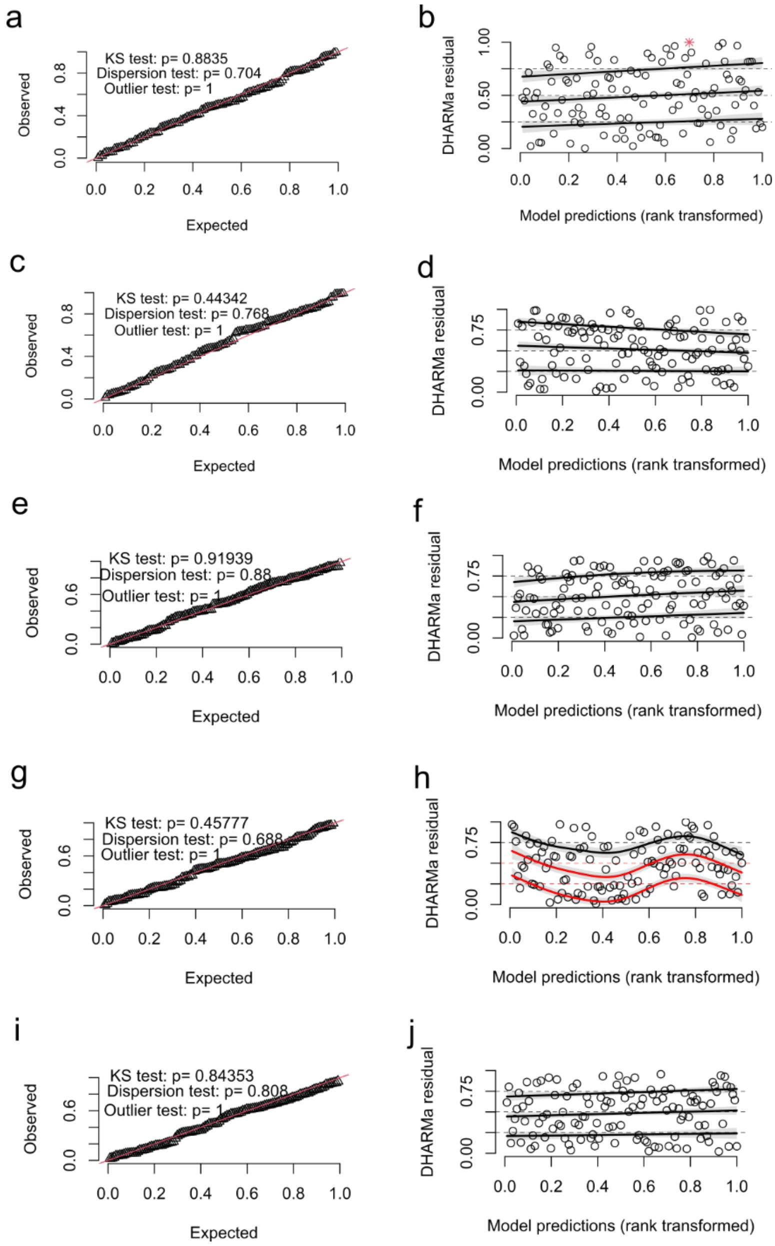
Domestic Retention (DR) and Domestic Contribution (DC) for all World Bank regions. The colors correspond to World Bank regions, as in Figures 1, 2, and 4 in the main text.

Fig. S4.



World cartogram representing Native turnover (a) and Name Bearing Types (NBT) turnover (b) of species for each country. Both values are combined in a bivariate map (c), where the breaks represent 25%, 50%, and 75% of the data. Values closer to one indicate higher divergence between the two sets (native fauna and NBT composition) of a given country.

Fig. S5.



qq-plots and DHARMa simulated residual x predicted plots from Generalized Linear Models (GLM) relating the total number of NBT per country (a, b), Domestic contribution (c, d), Domestic retention (e, f), Native turnover (g, h), and NBT turnover (i, j). The models have the same explanatory variables but different error probability distributions. Red solid lines in quantile simulated residuals x model predictions plots indicate significant deviation from expected in simulated residuals.

Table S1.
Coefficients (Estimates) estimated from a generalized linear model, with their respective confidence intervals (CI) and two-tailed p-values estimate (P-value) for the number of Name Bearing Types (NBT) in a country (dependent variables) and its relationship with number of native species (native richness), number of records per area according to Global Biodiversity Facility (records per area), number of years since the independence of a country (Years since independence), Gross Domestic Product (GDP) and the number of museums and ichthyological collections in each country (number of museums). Count component (Count component) estimates the effects of predictor variables on the non-zero values. The zero-inflated component corresponds to the estimate of absences (zero values).

	Total number of NBT		
Coefficients	*Estimates*	*CI*	*P-value*
Count Model			
Intercept	3.64	3.20 – 4.09	**<0.001**
Native richness	0.67	0.07 – 1.28	**0.030**
Gbif records per area	0.36	-0.00 – 0.73	0.052
Years since independence	-0.03	-0.41 – 0.35	0.879
GDP	0.84	0.37 – 1.31	**<0.001**
Number of museums	0.62	0.13 – 1.10	**0.013**
Dispersion parameter	0.46	0.33 – 0.65	
Zero-Inflated Model			
Intercept	-19.93	-36.70 – -3.16	**0.020**
Native richness	1.98	-1.08 – 5.04	0.205
Gbif records per area	-0.04	-1.13 – 1.05	0.941
Years since independence	1.01	-0.37 – 2.38	0.151
GDP	-0.56	-2.59 – 1.46	0.586
Number of museums	-52.35	-95.39 – -9.31	**0.017**
Observations	116		

Table S2.

Coefficients (Estimates) estimated from a generalized linear model, with their respective confidence intervals (CI) and two-tailed p-values estimate (P-value) for Domestic Contribution (dependent variable) and number of native species (native richness), number of records per area according to Global Biodiversity Facility (Gbif), number of years since the independence of a country (Years since independence), Gross Domestic Product (GDP) and the number of museums and ichthyological collections in each country (number of museums).

	Domestic Contribution (DC)		
Coefficients	*Estimates*	*CI*	*P-value*
Intercept	1.39	0.94 – 1.85	**<0.001**
Native richness	0.05	-0.30 – 0.40	0.772
Records per area	-0.41	-0.67 – -0.15	**0.002**
Years independence	0.15	-0.16 – 0.46	0.335
GDP	-0.71	-1.11 – -0.30	**0.001**
Number of museums	-0.07	-0.41 – 0.26	0.666
Observations	116		

Table S3.
Coefficients (Estimates) estimated from a generalized linear model, with their respective confidence intervals (CI) and two-tailed p-values estimate (P-value) for Domestic Retention (dependent variable) and number of native species (native richness), number of records per area according to Global Biodiversity Facility (Gbif), number of years since the independence of a country (Years since independence), Gross Domestic Product (GDP) and the number of museums and ichthyological collections in each country (number of museums).

	Domestic Retention		
Coefficients	*Estimates*	*CI*	*P-value*
Count component			
Intercept	-1.02	-1.34 – -0.70	**<0.001**
Native richness	0.08	-0.19 – 0.35	0.554
Records per area	0.06	-0.15 – 0.28	0.576
Years independence	0.02	-0.20 – 0.25	0.844
GDP	0.43	0.16 – 0.070	**0.002**
Number of museums	0.30	0.04 – 0.57	**0.024**
Zero-Inflated Model			
Intercept	-7.14	-14.16 – -0.12	**0.046**
Native richness	0.03	-1.78 – 1.85	0.971
Records per area	0.19	-1.41 – 1.80	0.813
Years independence	0.23	-0.77 – 1.22	0.656
GDP	-1.37	-3.29 – 0.55	0.163
Number of museums	-19.17	-36.83 – -1.50	**0.033**
Observations	116		

Table S4.

Coefficients (Estimates) estimated from a generalized linear model, with their respective confidence intervals (CI) and two-tailed p-values estimate (P-value) for Native turnover (dependent variable) and its relationship with number of native species (native richness), number of records per area according to Global Biodiversity Facility (Records per area), number of years since the independence of a country (Years independence), Gross Domestic Product (GDP) and the number of museums and ichthyological collections in each country (number of museums).

	Native turnover		
Coefficients	*Estimates*	*CI*	*P-value*
Intercept	1.54	1.22 – 1.85	**<0.001**
Native richness	-0.10	-0.46 – 0.26	0.585
Gbif records per area	-0.22	-0.52 – 0.08	0.151
Years since independence	-0.19	-0.49 – 0.11	0.212
GDP	-0.83	-1.16 – -0.49	**<0.001**
Number of museums	-0.62	-0.99 – -0.26	**0.001**
Observations	116		

Table S5.

Coefficients (Estimates) estimated from a generalized linear model, with their respective confidence intervals (CI) and two-tailed p-values estimate (P-value) for NBT turnover (dependent variable) and its relationship with the number of native species (native richness), number of records per area according to Global Biodiversity Facility (Records per area), number of years since the independence of a country (Years independence), Gross Domestic Product (GDP) and the number of museums and ichthyological collections in each country (Number of museums).

	Native turnover		
Coefficients	*Estimates*	*CI*	*P-value*
Intercept	1.54	1.22 – 1.85	**<0.001**
Native richness	-0.10	-0.46 – 0.26	0.585
Gbif records per area	-0.22	-0.52 – 0.08	0.151
Years since independence	-0.19	-0.49 – 0.11	0.212
GDP	-0.83	-1.16 – -0.49	**<0.001**
Number of museums	-0.62	-0.99 – -0.26	**0.001**
Observations	116		