Science AAAS

Supplementary Materials for

A global approach for natural history museum collections

Kirk R. Johnson, Ian F. P. Owens, the Global Collection Group

Corresponding author: Kirk R. Johnson, johnsonkr@si.edu

Science **379**, 1192 (2023)
DOI: 10.1126/science.adf6434

The PDF file includes:

The Global Collection Group members and affiliations
Supplementary Text
Figs. S1 to S3
References

Other Supplementary Material for this manuscript includes the following:

Tables S1 to S4

Supplementary Materials

A global approach for natural history museum collections in the 21st century

Kirk Johnson[1], Ian F. P. Owens[2], Alexei Abramov[3], Silvia Ametrano[4], Gretchen Baker[5], Luca Bartolozzi[6], Josh Basseches[7], Meg Beckel[8], Shannon Bennett[9], Marco Benvenuti[6], Lori Bettison-Varga[10], Tharina Bird[11,12], Dirk Brandis[13], Silvia Carrasquero[14], Hank Chaney[15], Nikita Chernetsov[16], Luis Chiappe[10], Alice Cibois[17], Alec Coles[18], Scott Cooper[19], Lynley Croswell[20], Gabor Csorba[21], Manuela da Silva[22], Aline da Silva Soares Souto[22], Bruno David[23], John Demboski[24], Tao Deng[25], Peter Denham[26], Steven Dessein[27], Richard Deverell[28], Christopher Dick[29], Ignacio Doadrio[30], Stefano Dominici[6], Steve Donnellan[31], Philip Edgar[32], Per Ericson[33], David C. Evans[7], Krista Fahy[15], Gunter A. Fischer[34], Linda Ford[35], Holger Frick[36], Matt Friedman[29], Ellen Futter[37], Lyudmila Gagarina[38], Jean-Marc Gagnon[8], Michael Galaty[39], Lawrence Gall[40], Scott Gardner[41], Dmitry Geltman[38], Gonzalo Giribet[35], Matthias Glaubrecht[42], Judy Gradwohl[43], Erin Gredell[40], Michel Guiraud[23], Doug Gurr[44], James Hanken[35], David Harris[45], Cheryl Hayashi[37], Kristofer M. Helgen[46], Dermot Henry[47], Peter M. Hollingsworth[45], Mike Hopkins[48], Marko Hyvärinen[49], Rebecca N. Johnson[1], Douglas Jones[50], Diana Jones[18], Aino Juslén[49], Gila Kahila Bar-Gal[51], Mikhail Kalyakinca[52], Alexander Kellner[53], Rahul Khot[54], Esther Kioko[55], Peter Kjærgaard[56], Michelle Koo[57], Leonard Krishtalka[58], Lars Krogmann[59], Michael Kuhlmann[60], Sebastian Kvist[33], Analía Lanteri[4], Yvonne-Marie Linton[61,1,62], Tim Littlewood[44], Hernan Lopez-Fernandez[29], Thorsten Lumbsch[63], Arnaud Maeder[17], Sekgwari Malematja[11], Joakim Malmström[33], Makoto Manabe[64], Albrecht Manegold[65], Lisa Månsson[33], Charles Marshall[66], Joel W. Martin[10], Richard McCourt[19], Kim McKay[46], Terrence Miller[26], Joseph Miller[67], Juan Carlos Monje[59], Jim Moss[39], Michael Novacek[37], Brian Oldman[31], Victor

Pacheco[68], Paulo Passos[53], Alan Paton[69], Igor Pavlinov[70], Marcelo Pelajo-Machado[22], A. Townsend Peterson[58], Dean Peterson[32], Sarah Phillips[69], Camille Pisani[71], Michelle Price[72], Nicholas D. Pyenson[1], Christiane Quaisser[73], David Reed[50], Nelson Rios[40], Adam Rountrey[29], Jeffery Saarela[8], Mark Sabaj[19], Hanieh Saeedi[74], Scott Sampson[9], Scott Schaefer[37], Nikolaj Scharff[56], Brigitta Schmid[75], Nicola Schoenenberger[72], Leif Schulman[49], Ramagwai Sebola[76], Patrick Semal[71], Cristiana Serejo[77], Emily Sessa[78], Kenichi Shinoda[64], Christian Sidor[79], Julian Siggers[63], David Skelly[40], Vincent Smith[44], George Sparks[24], Piet Stoffelen[27,80], Patricia Supply[71], Nicole Tarnowsky[78], Jordan Teisher[34], Pablo Teta[81], Barbara Theirs[78], Jim Thompson[26], Basil Thüring[36], Pablo Tubaro[81], Clare Valentine[44], Steven van der Mije[82], Edwin van Huis[82], Nurin Veis[47], Johannes Vogel[83], Katrin Vohland[75], J. Wolfgang Wägele[84], Michael Wall[43], Yuan Wenwai[85], Karsten Wesche[86,87,88], Matt Woodburn[44], Andrew Young[89], Rafael Zardoya[30]

[1]National Museum of Natural History, Smithsonian Institution, Washington, DC, USA, [2]Cornell University, Ithaca, NY, USA, [3]Zoological Institute of the Russian Academy of Sciences, Saint Petersburg, Russia, [4]Museo de La Plata, La Plata, Argentina, [5]Carnegie Museum of Natural History, Pittsburgh, PA, USA, [6]Universita degli Studi di Firenze Museo di Storia Naturale, Florence, Italy, [7]Royal Ontario Museum, Toronto, ON, Canada, [8]Canadian Museum of Nature, Ottawa, ON, Canada, [9]California Academy of Sciences, San Francisco, CA, USA, [10]Natural History Museum of Los Angeles County, Los Angeles, CA, USA, [11]Ditsong Museum of Natural History, Pretoria, South Africa, [12]Department of Zoology and Entomology, University of Pretoria, Pretoria, South Africa, [13]Zoologischen Museums Kiel, Kiel, Germany, [14]Universidad Nacional de La Plata, La Plata, Argentina, [15]Santa Barbara Museum of Natural History, Santa

Barbara, CA, USA, [16]Zoological Institute of the Russian Academy of Sciences, Saint-Petersburg, Russia, [17]Museum d'histoire naturelle de la Ville de Geneve, Geneva, Switzerland, [18]Western Australia Museum, Perth, WA, Australia, [19]Academy of Natural Sciences of Drexel University, Philadelphia, PA, USA, [20]Museums Victoria, Melbourne, VIC, USA, [21]Hungarian Natural History Museum, Budapest, Hungary, [22]Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [23]Muséum national d'Histoire naturelle, Paris, France, [24]Denver Museum of Nature and Science, Denver, CO, USA, [25]Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, China, [26]Queensland Museum South Bank, South Brisbane, QLD, Australia, [27]Meise Botanic Garden, Meise, Belgium, [28]Royal Botanic Gardens, Kew, Kew, UK, [29]University of Michigan, Ann Arbor, MI, USA, [30]Museo Nacional de Ciencias Naturales, Madrid, Spain, [31]South Australia Museum, Adelaide, SA, Australia, [32]Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand, [33]Swedish Museum of Natural History, Stockholm, Sweden, [34]Missouri Botanical Garden, Saint Louis, MO, USA, [35]Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA, [36]Naturhistorisches Museum Basel, Basel, Switzerland, [37]American Museum of Natural History, New York, NY, USA, [38]Komarov Botanical Institute of the Russian Academy of Sciences, Saint Petersburg, Russia, [39]University of Michigan Museum of Anthropological Archaeology, Ann Arbor, MI, USA, [40]Yale Peabody Museum of Natural History, New Haven, CT, USA, [41]University of Nebraska State Museum and School of Biological Sciences, Lincoln, NB, USA, [42]Centrum für Naturkunde, Universität Hamburg, Hamburg, Germany, [43]San Diego Natural History Museum, San Diego, CA, USA, [44]Natural History Museum, London, UK, [45]Royal Botanic Garden Edinburgh, Edinburgh, Scotland, [46]Australian Museum, Sydney, NSW, Australia, [47]Museums Victoria, Melbourne, VIC, Australia, [48]Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil, [49]Finnish Natural History Museum Luomus, Helsinki, Finland, [50]Florida Museum of Natural History, Gainesville, FL, USA, [51]The Hebrew University of Jerusalem, Jerusalem, Israel, [52]Zoological Museum of Moscow University, Moscow, Russia, [53]Museu Nacional - UFRJ, Rio de Janeiro, Brazil, [54]Bombay Natural History Society, Mumbai, India, [55]National Museums of Kenya, Nairobi, Kenya, [56]Natural History Museum of Denmark, Copenhagen, Denmark, [57]Museum of Vertebrate Zoology, University of California, Berkeley, CA, USA, [58]University of Kansas Biodiversity Institute, Lawrence, KS, USA, [59]Staatliches Museum fur Naturkunde Stuttgart, Stuttgart, Germany, [60]Christian Albrechts Universitat zu Kiel Zoologisches Institut und Museum, Kiel, Germany, [61]Walter Reed Biosystematics Unit, Smithsonian Institution Museum Support Center, Suitland, MD, USA, [62]Walter Reed Army Institute of Research, One Health Branch, Silver Spring, MD, USA, [63]Field Museum of Natural History, Chicago, IL, USA, [64]National Museum of Nature and Science, Tokyo, Japan, [65]State Museum of Natural History Karlsruhe, Karlsruhe, Germany, [66]Museum of Paleontology, University of California, Berkeley, CA, USA, [67]Global Biodiversity Information Facility Secretariat, Copenhagen, Denmark, [68]Universidad Nacional Mayor de San Marcos, Lima, Peru, [69]Royal Botanic Gardens, Kew, London, UK, [70]Research Zoological Museum, Lomonosov Moscow State University, Moscow, Russia, [71]Royal Belgian Institute of Natural Sciences, Brussels, Belgium, [72]Conservatoire et Jardin Botaniques de Geneve, Geneva, Switzerland, [73]Museum fur Naturkunde - Leibniz-Institut fur Evolutions - und Biodiversitatsforschung, Berlin, Germany, [74]Senckenberg Research Institute and Natural History Museum, Frankfurt, Germany, [75]Naturhistorisches Museum Wien, Vienna, Austria, [76]South African National Biodiversity Institute, Pretoria, South Africa, [77]Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, [78]New York Botanical Gardens, Bronx, NY, USA, [79]Burke Museum of Natural History and Culture, Seattle, WA, USA, [80]Agentschap

Plantentuin Meise, Meise, Belgium, [81]Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires, Argentina, [82]Naturalis Biodiversity Center, Leiden, The Netherlands, [83]Museum fur Naturkunde - Leibniz-Institut fur Evolutions- und Biodiversitatsforschung, Berlin, Germany, [84]Zoologisches Forschungsmuseum Alexander Koenig Leibniz-Institut fur Biodiversitat der Tiere, Bonn, Germany, [85]Nanjing Institute of Geology and Paleontology, Chinese Academy of Sciences, Nanjing, China, [86]Senckenberg Museum fur Naturkunde Gorlitz, Gorlitz, Germany, [87]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany, [88]Technische Universitat Dresden Internationales Hochschulinstitut Zittau, Zittau, Germany, [89]Commonwealth Scientific and Industrial Research Organisation, Canberra, ACT, Australia

*Corresponding author: johnsonkr@si.edu

1. The Global Collection

The global collection is the physical basis for our understanding of the natural world and our place in it. Natural history collections are a unique source of information that is directly relevant to issues as diverse as wildlife conservation, pandemic preparedness, food security, invasive species, rare minerals, and the bioeconomy (*1*–*8*). Advances in digital, isotopic, imaging, and genomic technologies, as well as machine learning and artificial intelligence, are transforming and amplifying how natural history collections can be used (*9*–*15*).

Despite their potential value, natural history collections are at risk. Fires recently consumed irreplaceable collections at the national museums in Rio de Janeiro in Brazil and New Delhi in India (*16*, 17). The last century of wars has shown the recurrent vulnerability of urban museums to conflict, and this threat remains for institutions in conflict zones, such as the National Herbarium of Ukraine, which was damaged by a missile strike in October 2022. Less dramatic examples of collection degradation and destruction occur because of long-term under-investment in infrastructure and expertise (*18*). All museum collections also need to confront evolving ethical standards because they hold objects that may have been collected without prior consent, without permission from source communities, or without full recognition of contributors and cultural sensitivities (*19*, *20*).

We worked with the leadership and staff of 73 of the world's largest natural history museums and herbaria from 28 countries to design and complete a simple and rapid survey of their collective holdings. Our goal was to demonstrate that it is possible to accurately calculate the size and characteristics of the global collection in advance of full digitization. We created a shared framework that defined a grid of 19 collection types by 16 geographic regions, such that any collection object would fall into only one of the resulting 304 cells. We then worked with staff from each museum to estimate the number of objects to the nearest order of magnitude in each of the cells at each of the museums (e.g., see *21*). For this effort, the term "collection unit" represents a single museum's holdings within a single cell (Supplemental Fig. 1). In addition, we censused the museum workforce that studies and cares for collections and that makes them available to the global community of users (*22*, *23*).

Our initial study does not address the hundreds of smaller museums, their collections, and their staff, which comprise the rest of the global collection; these collections are uniquely valuable because of their regional holdings and the specificity of their expertise (*24-26*). It is our intention to work with GBIF to make our survey tool available to all natural history museums, and assure the longer-term management of collections data by integrating them with visualizations within the Global Registry of Scientific Collections (GRSciColl) (*27*).

Enabling natural history museums and their collections to respond to 21st century challenges will require greater financial support and enhanced international collaboration to accelerate the availability and use of specimen-based data and to strategically plan for future collecting effort. Such support and coordination will be essential to overcome imbalances in information, expertise, and cultural differences between North America, Europe, and other regions. Also, this work needs to happen at a pace and magnitude that will meet the urgency of the Anthropocene and with the understanding that there are more species at risk of extinction than are currently known to science (in line with recommendations from Refs. *1-15*, *21-23*).

2. Case studies – the value and importance of natural history collections

Breakout box 1 – Pandemic preparedness

Recent work demonstrating the multiple zoonotic origins of COVID–19 should be unsurprising: zoonosis accounts for around 60% of known infectious diseases of humans, and 75% of all emerging infectious diseases (*28*). Natural history collections are key resources for building pandemic preparedness because of their invaluable georeferenced biological source material and associated microbes (bacteria, parasites, and viruses) (*29*). Frozen tissue collections and biorepositories hold a wealth of associated etiological agents that have not yet been realized, and these samples, along with ongoing natural history field surveillance, hold the key to global pandemic preparedness. The financial impact of the COVID-19 pandemic (estimated at $12.5 trillion through 2024 (*30*) highlights the need for stronger and more coherent pandemic preparedness. Given the similarities between SARS-CoV-2 and bat genomes, deeper investigations into potential reservoir hosts are underway across nine European institutions, focused on over 20,000 specimens in three bat families, representing >100 years of collecting. Next–generation sequencing on these specimens can detect genomic signals of novel and ancestral viruses, contributing sequences of viral species and strains to augment vaccine development, while providing a comprehensive overview of bat distribution, ecology and behavior.

Recently, CETAF (Consortium of European Taxonomic facilities) and Distributed System of Scientific Collections (DiSSCo) undertook a joint response by natural history collections to the COVID-19 pandemic that addressed: 1) animal virus carriers; 2) the construction of a knowledge base relevant for pandemics; 3) preservation of viral evidence; and

4) metadata registering practices (*29*, *31*, *32*). The taskforce demonstrated how collaboration between collection holding institutions from across the globe, and more specifically of researchers and collections experts and especially reinforced a need to share data with human disease experts in the framework of One Health to support modern pan–generic vaccine development.

Breakout box 2 – Measuring global change

On a rapidly changing planet, natural history collections provide a primary baseline for the distribution of species because museum specimens show historical (and even paleontological) occurrences prior to the onset of widespread change to species habitats, ranges, and abundances. Repeating modern day collecting and surveys in previously collected localities can demonstrate biodiversity changes from historical baselines (*33*) and monitor progress to attenuate its loss (*34*). The Natural History Museum of London's PREDICTS (Projecting Responses of Ecological Diversity in Changing Terrestrial Systems) initiative has developed a Biodiversity Intactness Index to estimate the percentage of the original number of species that remain in a modified region, their abundance in any given area, despite human impacts, which can help scientists and policy makers track biodiversity loss. Historical collections data can also provide the sole taxonomically and geographically comprehensive reference points to understand the biological impacts of large–scale natural catastrophes such as fires and floods. For example, herbarium records provided key guidance for focusing ecological restoration activities to assess the effects of the 2019–20 megafires in Australia's southeastern forest and woodland ecosystems on plant communities (*35*).

Breakout box 3 – Measuring biodiversity as a conservation tool

After more than 300 years of scientific taxonomic labor, 2.2 million species of living organisms have been formally described and named (*36*). While non–trivial, this total represents only a fraction of the estimated 15 million species on Earth (*37–40*), leaving enormous gaps in our understanding of patterns and processes that underpin the interactions between climate change and biodiversity. To truly understand the complexity of the natural world—an essential precursor to managing climate change and biodiversity loss – it is crucial to know all the species and their interactions. Next–generation sequencing and artificial intelligence platforms that enable rapid species recognition and infrastructures designed to facilitate posting and sharing this type of biodiversity big data—including whole genome sequences, e–vouchers and associated species distributions—will improve identifying and protecting these species that may not yet have been described (*41*). These new techniques, however, do not supplant the need for taxonomic experts to provide verifiable reference datasets. The rapid growth of environmental DNA analysis shows advances in imaging, machine learning, and artificial and collaborative intelligence offer great promise for assisting researchers in unpacking the taxonomic complexity of hyper–diverse and ecologically critical groups such as insects and fungi, which comprise much of the world's dark biodiversity data.

Breakout box 4 – Biodiversity data sharing

Biodiversity information is undergoing a phase of massive expansion of data resources. Chief among these data resources are primary biodiversity data (PBD), also known as "occurrence data", which document the presence of one or more individuals of a particular taxon at a particular place at a particular time. More than 2.24 billion PBD records are publicly available via GBIF. The great bulk of PBD, however, has come from recent years. Of the 2.24 x 10^9 records available via GBIF, 77.1% entries came from 2000–2022, and 87.3% entries came from 1980–2022, indicative of its growing momentum. This youthful PBD work reflects the dominant role that observational data and citizen–science initiatives play in biodiversity informatics, driving the accumulation of large numbers of occurrence records. Understanding and documenting species invasions, however, requires a more continuous record of biodiversity distributions through time. This historical record is provided by natural history museum specimens. Among the 208.7 million museum specimen records in GBIF, only 20.7% come from 2000–2022 and 40.7% from 1980–2022; indeed, 17.6% come from *before* 1950, and 4.1% come from before 1900. In addition, natural history museums have an all–taxon scope of interest, in contrast with citizen–science work, which has focused on observable and identifiable groups (e.g., birds). As such, natural history museum data have the potential to provide a crucial, longitudinal view of biodiversity distributions through time that is otherwise difficult to obtain.

Breakout box 5 –Invasive species

Natural history museum holdings, accumulated over decades and representing millennia, are irreplaceable records of extant and extinct biodiversity (*42*). The advent and wide–spread utility of DNA barcoding and metagenomics have driven a need for museum specimens because of the quality of their associated data. Large–scale global DNA barcoding initiatives, such as Fish Barcode of Life (FISH-BOL; 31,000 species) and the All Bird Barcoding Initiative (ABBI), rely heavily on museum specimens and taxonomic expertise to source the starting materials, and resolve downstream identification issues (*11*).

Once comprehensive reference libraries are established, DNA barcodes can be employed to molecularly identify species at any life stage, or from their remains. Applications include environmental monitoring and assessments (e.g., invasive species detection, biosurveillance), wildlife conservation (e.g., poaching, bushmeat, exotic products, breeding programs), and consumer fraud (e.g., validation of species sold). Recently, the Mosquito Barcoding Initiative reference library, primarily sourced from the Smithsonian's National Museum of Natural History (NMNH) and Natural History Museum, London (NHM London) collections, confirmed that recently blood–fed malaria vectors in the Sahel undertake long distance aerial migrations to avoid hostile conditions, increasing disease outbreak risk in non–malarial areas and sabotaging on–ground vector mitigation efforts (*43*). Following the unprecedented successive outbreaks of non–endemic viruses (Chikungunya in 2013 and Zika in 2015) in the Caribbean and Latin America, DNA barcodes confirmed the invasion and establishment of the mosquito *Aedes vittatus* in Cuba and the Dominican Republic —the first record of the Old World Dengue, Chikungunya, Zika and Yellow Fever virus vector in the Americas. Network analysis of reference COI barcodes from global populations confirmed multiple introductions into the Caribbean from India and Pakistan, presumably as desiccant–tolerant eggs, which may or may not have been trans–ovarially infected with the invasive Chikungunya or Zika viruses (*44*).

Breakout box 6 – Museomics

Collections now routinely store new specimen material in conditions suitable for preservation of nucleic acids (DNA and RNA). But this has not always been the case in the past, when preservation of morphological structures was the only goal. Obtaining genetic information from these archival specimens, which make up the bulk of collections, has proved challenging even for PCR–based approaches. However, this goal has become increasingly feasible with new molecular approaches complementing morphological work (*45–47*), opening opportunities to characterize the genomes of extinct species, estimate genetic variation in wild populations before human–impacts (e.g., insecticide resistance evolution in blowflies), or characterize associated microbiomes and link pathogens to their host species. Genetic barcodes linked to taxonomically verified voucher specimens are also emerging as critical reference data sets for eDNA surveys, wildlife forensics and monitoring of illegal wildlife trade. Even more recent molecular advances offer the prospect of being able to infer gene expression (*48*) or epigenetic state (*45*, *49*), providing a unique window into physiological and genetic responses to environmental change.

Breakout box 7 – Colonial heritage and mutual learning

The Museum für Naturkunde's exhibited skeleton of the sauropod dinosaur *Giraffatitan brancai* is one of many natural history objects acquired during European colonial expansion, making *Giraffatitan* both a paleontological star object and an object of cultural heritage and colonialism. Excavated between 1909 and 1913 in what is now Tanzania (former "German East Africa"), *Giraffatitan brancai* raises questions about the connection between science, museums, and colonialism, about the responsibility of museums and the future of collections. An interdisciplinary research network led by the Museum für Naturkunde in Berlin has discussed and published on these topics in collaboration with Tanzanian researchers. Three books, written in German (*50*), Kiswahili (*51*), and English (the latter forthcoming), provide broad access to the findings. The results are a milestone in collection and museum research, stimulating debate about the political framework of nature and natural history institutions, but also questioned identities and long held (natural history museum) narratives.

Looking at the past is inspiring the future. TheMuseumsLab is a platform for joint learning, exchange and continuing education on the future of museums in both Africa and Europe. It aims to strengthen knowledge and competencies, to foster new ideas and approaches as well as to establish a close and lasting network amongst future museum leaders. Several African and European museums as well as other cultural institutions are partners within the program. Together with 50 African and European fellows, over 100 lecturers, speakers and mentors from African and European arts, cultures and sciences, TheMuseumsLab aims to change not only museums but also foster societal change by addressing urgent issues such as social justice and climate change.

The annual program consists of seminar modules (online and onsite in Berlin and Cape Town) lead by prominent African and European experts, a two-week residency at a renowned European partner institution and a co-working phase. The project was developed by the Museum für Naturkunde Berlin, the German Academic Exchange Service (DAAD) and the Master's Programme in Museum Management and Communication at the University of Applied Sciences (HTW) Berlin, in close cooperation with the African consultancy group The Advisors. It is funded by three German ministries (AA, BKM, BMZ) and for further information see https://themuseumslab.org.

3. Materials, Methods & Results

Institutions included in assessment

To begin to represent the collective power of the global collection, we collated a dataset that includes 73 of the world's largest collection–based institutions, including museums, botanic gardens, research institutes, and universities (Supplementary Table 1). The participating institutions span Africa, Asia, Australasia, Europe and North and South America. Institutions were selected for potential inclusion in the assessment based on the putative size of their natural history collection in terms of the total number of specimens. Natural history collections were defined as those including the following traditional collection areas: botany (including plants, algae and fungi), entomology (including insects and spiders), vertebrate zoology, invertebrate zoology (excluding insects and spiders), paleobiology, geology, and anthropology. We included city, state, and national natural history museums, university natural history museums, and herbaria and focused on the largest institutions from a breadth of geographic regions. In general, the largest collection–based institutions exist in Europe and North America and in these regions, we targeted those collections that were larger than 10 million specimens. For other parts of the world with fewer very large institutions we sought to survey collections that were larger than 1 million specimens.

Initial information on size of collection was obtained from the Global Registry of Scientific Collections (GRSciColl https://www.gbif.org/grscicoll), Consortium of European Taxonomic facilities (CETAF www.cetaf.org), Society for the Protection of Natural History Collections (SPNHC www.spnhc.org), Integrated Digitized Biocollections (iDigBio www.idigbio.org), Global Genome Biodiversity Network (GGBN www.ggbn.org), Atlas of Living Australia (ALA https://www.ala.org.au), Index Herbariorum (sweetgum.nybg.org/science/ih), Paleobiology Database (PBDB www.paleobiodb.org) and the webpages of individual institutions.

Institutions selected on the above criteria were invited to participate in the assessment using a standard email to directors and heads of research or collections with email addresses obtained from the sources above. Institutions were approached a number of times until they either confirmed their interest in participation or declined the invitation. Institutions that confirmed interest in participating were then sent a series of data templates to complete to provide information on the specimens in their collection and the scientific staff and volunteers associated with their institution.

Institutions were classified according to the geographical region in which they are based, using the following regions: Africa, Tropical Asia, Temperate Asia, Australasia, Europe and North and South America. Russia was classified as European because the institutions included in this analysis are in the part of Russia typically considered part of Europe. In total, 116 institutions were invited to participate and 73 provided sufficient information to be included in this assessment.

Supplementary Table 1. Museums and Herbaria

[See Supplementary Table 1. Museums and Herbaria. Surveyed natural history museums and herbaria including the geographic region where they are located, their reported collection count, and (in second tab), their relative ranking.]

Collections Data Collected

Lead scientists, curators, and collection leads from a core group of museums defined 19 collection types that spanned the entirety of biological, geological, paleontological, and anthropological collections and 16 terrestrial and marine regions that covered the entirety of the Earth (Figure 1). This resulted in a 304–cell grid made up of 19 collection categories by 16 regions (9 terrestrial and 7 marine geographic regions (Supplementary Table 2) representing the initial global collection dataset.

To facilitate data collection, the collection categories are based on the traditional groupings of ‘anthropology’, ‘botany’ (includes plants, algae and fungi), ‘entomology’ (includes insects and spiders), ‘geology’, ‘paleobiology’, ‘vertebrate zoology’ and ‘invertebrate zoology’ (excludes insects and spiders).

The 19 collection categories referred to the type of specimen or object and were:

- Botany;
- Vertebrate Zoology, Fishes; Vertebrate Zoology, Amphibians; Vertebrate Zoology, Birds; Vertebrate Zoology, Mammals; Vertebrate Zoology, Reptiles;
- Entomology;
- Invertebrate Zoology, Arthropods; Invertebrate Zoology, Mollusks; Invertebrate Zoology, Other invertebrates;
- Paleobiology, Vertebrates; Paleobiology, Invertebrates; Paleobiology, Paleobotany;
- Anthropology, Cultural Collections; Anthropology, Archaeology; Anthropology, Human Biology;
- Geology, Minerals; Geology, Geology; Geology, Meteorites.

In some subsequent analyses these categories were collapsed down to Botany, Entomology Vertebrates, Invertebrates, Paleobiology, Geology and Anthropology.

The 16 geographic regions referred to the origin of the specimens were:

- Nine Terrestrial regions (*52*):

Terrestrial, Africa; Terrestrial, Antarctic; Terrestrial, Asia–temperate; Terrestrial, Asia–tropical; Terrestrial, Australasia; Terrestrial, Europe; Terrestrial, North America; Terrestrial, South America; Terrestrial, Pacific.

- Seven Marine regions based on Marineregions.org (*53*):

  Marine, Atlantic Ocean–North; Marine, Atlantic Ocean–South; Marine, Arctic; Marine, Indian Ocean; Marine, Pacific Ocean–North; Marine, Pacific Ocean–South; Marine, Southern Ocean. In some subsequent analyses these categories were reduced to Terrestrial and Marine.

A log 10 scale was used to estimate the size of collection in each of the 304 cells for each institution so that institutions could assess their collections both accurately and relatively rapidly with the available resources. Through discussion with institutions, it was established that that this was the finest scale that it was possible for all institutions to quantify the size of collection across all collection categories and geographic regions. The log 10 scale used was: 0 = 0 specimens; 1=1–10 specimens; 2=11–100 specimens; 3=101–1,000 specimens; 4=1,001–10,000 specimens; 5=10,001–100,000 specimens; 6= 100,001–1,000,000 specimens; 7= 1,000,001–10,000,000 specimens; 8 >10,000,000 specimens.

We define a single museum's collection within one cell as a collection unit. Curators and collection managers know their collections well enough that it was relatively easy to complete the surveys; even the largest museums were able to assess their collection units in a few weeks. This step allowed for a rapid assessment that was imprecise but accurate, creating heat maps that clearly show areas of collection strength and weakness.

Through our survey, 71 of the 73 museums provided total object counts or estimates, and for the remaining two an approximate value was calculated using the collection size breakdowns that they provided. Collectively, these 73 museums hold 1,147,934,687 natural history (biological, paleontological, and geological) specimens and anthropological objects. We acknowledge that this sum is overly precise, but it is based on the sum of the collection counts as provided by the institutions. The average number of objects across the 73 museums is 15.7 million. The accumulation curve for specimens and objects across participating museums shows that a large proportion of the sampled collection is held by a surprisingly small number of museums, with almost half (49%, over half a billion) in just 10 institutions and more than 90% (over 1 billion) in just 42 institutions (Supplementary Fig. 2; Supplementary Table 1). Given the shape of this curve and the sense that we included nearly all the world's very large collections in the survey, we surmise that we surveyed more than 50% and perhaps as much as 70% of all specimens and objects in the global collection. Our goal was not to discount small collections but simply to begin the global assessment with the largest collections for the sake of efficiency. Specimens counts reported by the 43 institutions that chose not to participate in the survey totaled 239,439,997 objects. Thus, if we had successfully surveyed all 116 institutions, we would have accounted for 1,387,374,684 objects.

With 71 collections and 304 cells, there are 21,584 possible collection units and the survey identified 7,445 actual collection units. Of these, 1,957 represented collection units with more than 10,000 specimens and objects (Supplementary Fig. 1A), and 242 represented collection units with more than one million specimens and objects (Supplementary Fig. 1B). Of the 71 museums, only 3 reported more than 200 collection units, 39 had between 100 and 199, and 29 had fewer than 100. We argue that this type of assessment provides a good measure of the breadth of a museum's collection. These figures also paint the first view of a large percentage of the global collection's composition. The top five collection types based on number of collection units with more than 10,000 specimens are entomology (n=248), mollusks (n=248), botany (n=232), fishes (n=182), and marine arthropods (n=157) (Supplementary Fig. 1A, Supplementary Table 2). For collection units with more than one million specimens, the hotspots are entomology (n=96), botany (n=33), invertebrate paleontology (n=26), mollusks (n=22), and marine arthropods (n=19) (Supplementary Fig. 1B). In addition to exposing the areas of most intense collecting, the heat maps also show key areas where collecting has been minimal such as the Arctic Ocean.

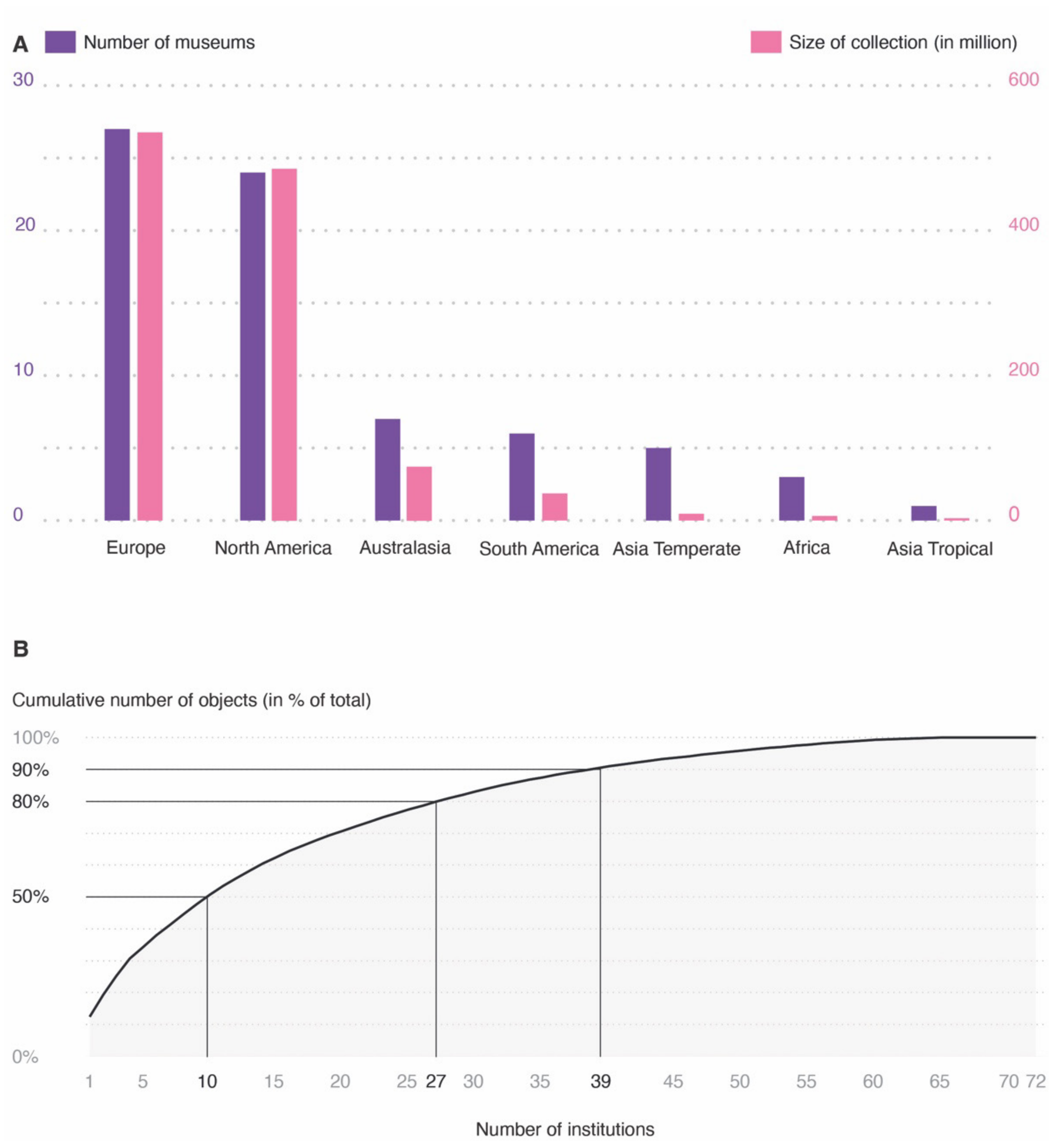
Supplementary Table 2. Collection Data.

[See Supplementary Table 2. Collection Data. Primary data showing the collection units for each of the surveyed museums and herbaria. Data are number of specimens estimated to the log base 10 for each collection cell for each museum.]

Supplementary Figure 1. Heatmaps of global collection units.

Supplementary Figure 1. Heatmaps of global collection units. Heat maps showing number of collection units in each of 304 cells based on 19 collection types and 16 geographic subdivisions of the Earth. (A) for all collection units >10k objects (n=1,957); and (B) for all collections units >1 M objects (n=242).

Supplementary Figure 2. Distribution of the global collection.



Supplementary Figure 2. Distribution of the global collection. (A) Ranked histogram of number of museums and their collection sizes, grouped by geographic region from Fig. 1, in descending order; (B) Accumulation curve for the number of specimens across 73 surveyed museums. Ten, 27, and 39 institutions make up 49% (563 million), 80% (917 million), and 90% (1.03 billion) of the total number of objects, respectively.

Collection of Workforce Data

To assess the expertise that makes up the other critical part of the global collection our data collection also captured information on roles (research staff, collections staff, and volunteers) and career demography by recording the type of role that individuals occupy.

Self–reported data on the size and composition of the scientific workforce associated with each institution was collected using the same set of 19 collection categories, recording the number of full–time equivalent staff and volunteers in each category. Staff were classified as being primarily employed to either undertake research or care for and provide access to the collection, including collection conservation, analytical and imaging facilities, digitization, databases and informatics. Volunteers were defined as non–salaried expert personnel working with the collection under the supervision of the host institution. Fractional full–time equivalent staff and volunteer numbers were included and summed.

To quantify the distribution of research expertise, for each member of research staff we also recorded expertise using the same matrix of 19 collection categories used for the data on collections. (Supplementary Table 3).

The participating institutions host a scientific expertise of 8,527 individuals, 29% are staff employed primarily to undertake research on the collection, 24% are staff employed primarily to care for, and provide access to, the collection, and 46% are volunteers who support both efforts. Approximately half the reported staff are in just 14 of the 73 institutions (Supplementary Table 3).

Among the 2,488 research staff, 23% are employed as experts in botany, 19% in paleobiology, and 14% in vertebrate zoology, 12% in invertebrate zoology, 12% in anthropology and 10% in entomology, 10% in geology (Supplementary Table 3). We note that these data were collected before the pandemic, and thus we do not fully know the impact of COVID–19 on these numbers.

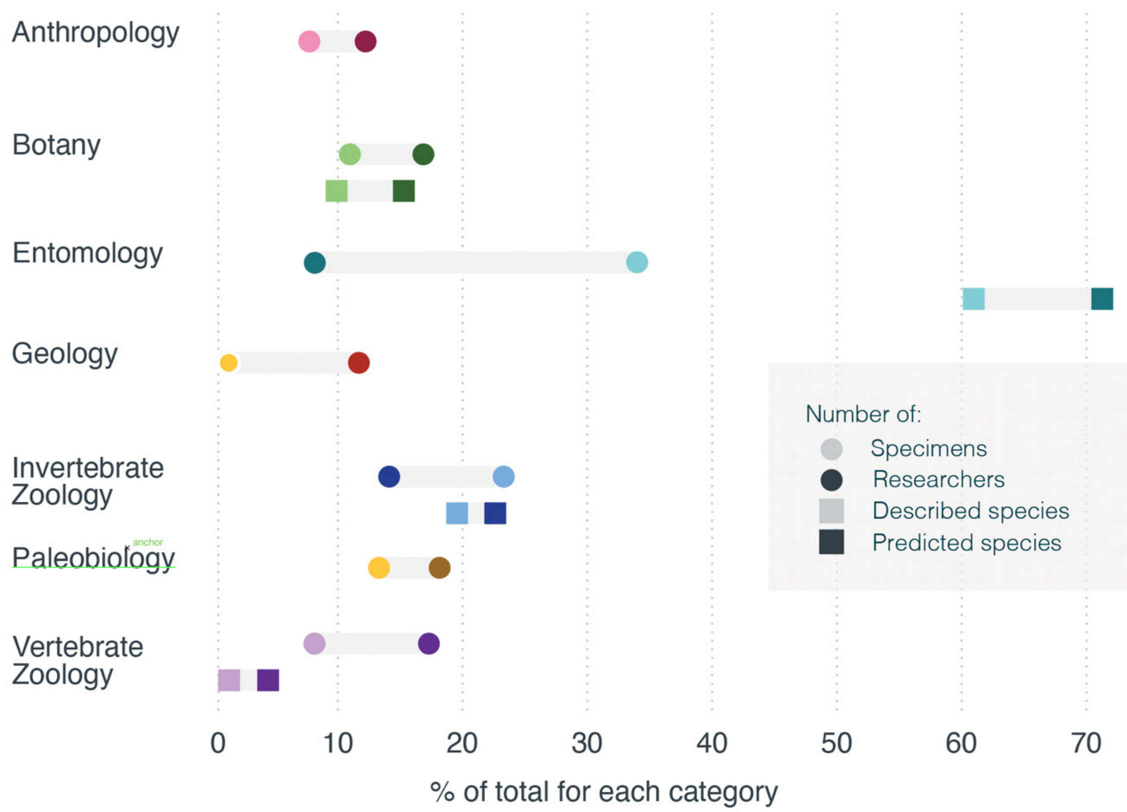
This imbalance continues with staff expertise by collection type for extant biological categories (botany, entomology, invertebrate and vertebrate zoology). When these categories are mapped by the number of described or predicted species, vertebrate zoology specimens are relatively well represented, while entomological specimens are still under collected (Supplementary Fig. 3). The focus on vertebrate specimens is likely attributable to their frequent use in public exhibits, their role as exemplars of conservation issues, and the long tradition of vertebrate taxonomy. Even though our data show that vascular plants and bryophytes are among the most well–represented groups (Supplementary Fig. 3) recent work has suggested that the number of undescribed fungal species may be much higher than previously expected (*54*), highlighting an under–collection in the botany category. While entomological specimens have broad global coverage in the global collections, however given the predicted number of species (Supplementary Fig. 3) and their importance in food security and ecosystem services these collections are still insufficient. By assessing the global collection, we conclude that increased collection of entomological specimens (and relevant expertise) highlights a future collecting need that is dire in the context of recent dramatic drops in insect biodiversity and abundance (*55*), along with their critical importance for food security and pandemic preparedness (*56*, *57*).

We hope that one of the next outcomes of this quantification of the global collection is that we may begin to assess when we have ‘enough’ to represent coverage of those groups that are being increasingly recognized as important sentinels of change (*58*, *59*).

Supplementary Table 3. Workforce

[See Supplementary Table 3. Workforce. Size and composition of museum workforce and museum volunteers for each museum.]

Supplementary Figure 3. Specimens, researchers, and species across categories



Supplementary Figure 3. Specimens, researchers, and species across categories. Comparison across collection categories between the number of specimens and anthropological objects (light circles) and the number of researchers (dark circles), and for extant biological categories (botany, entomology, invertebrate zoology, and vertebrate zoology), with the number of described (light squares) and predicted species (dark squares). Data sources: Catalogue of Life (https://www.catalogueoflife.org/) for described; and Mora et al. (*38*), for predicted. See Supplementary Table 4 on digital and genomic discoverability.

For both neo- and paleontological collection categories, the number of specimens digitally discoverable was obtained from the Global Biodiversity Informatics Facility portal (GBIF www.gbif.org). Data from GBIF was restricted to preserved specimens and fossil specimens, respectively. GBIF was used as a single source of digitally available specimens because it serves information from major disciplinary and national aggregator sites such The Paleobiology Database (PBDB www.paleobiodb.org), Integrated Digitized Biocollections (iDigBio www.idigbio.org), and Atlas of Living Australia (ALA www.ala.org.au), as well as major institutional sources. Not all data from other sources is served to the GBIF portal, however, and the degree of overlap between sources is not clear, so comparisons with the number of specimens in GBIF should be interpreted cautiously. GBIF data were accessed at the same time that the collection database was finalized.

For neontological collection categories the number of tissue and DNA specimens available was obtained from the Global Genome Biodiversity Network portal (GGBN www.ggbn.org/ggbn_portal/). GGBN was used as a single source of digitally available specimens because it serves information from major sources such as the International Barcode of Life (IBOL www.ibol.org) as well as 85 other member institutions. Again, not all data from other sources are served to the GGBN portal, and the degree of overlap between sources is not clear, so comparisons with the number of specimens in GGBN should be interpreted cautiously. GGBN data were accessed at the same time that the collection database was finalized.

Relatively few specimens in the surveyed collections are discoverable in open digital databases or genomic repositories (Supplementary Table 4). For instance, we estimate that only 16% of specimens from participating institutions are discoverable in the Global Biodiversity Information Facility (GBIF) portal https://doi.org/10.15468/dl.7gzvbw. Furthermore, about half of the specimens discoverable via GBIF are plants, and just 22% are insects, despite plants representing around 20% of all described species and insects making up more than half. When botanical collections are excluded, only about 8.5% of the estimated total specimens are discoverable by GBIF. Similarly, when considering new genomic technologies, for extant biological specimens, only around 0.2% of the estimated total number of specimens from participating institutions are genome-ready records discoverable in the Global Genome Biodiversity Network (GGBN) (chosen because these represent vouchered specimens in contrast to GenBank which does not have this requirement) with 10 animal phyla still having no genomic representation at all. These percentages should be interpreted cautiously because not all digital or genomic records are made available via the GBIF and GGBN portals, and these portals contain records from institutions outside those participating in this study. Nonetheless, taken together these comparisons show that most of the global collection currently exists as dark data that are not openly digitally discoverable to the entire international research community. Given the increasing scientific potential and societal importance of collections, illumination of these data is urgent (*14*).

Supplementary Table 4. Digital discoverability of species

[See Supplementary Table 4. Discoverability of specimen data. Comparison between the number of specimens in the collections of the participating museums and the number discoverable in digital and genomic portals. Data on digital and genomic records from the Global Biodiversity Informatics Facility (GBIF), and the Global Genome Biodiversity Network (GGBN), respectively.]

4. Data Availability and Additional Remarks

Data for the global collections survey has been published on Zenodo (https://zenodo.org/), under an open license (Creative Commons Attribution 4.0 International), and they can be persistently accessed at https://doi.org/10.5281/zenodo.6985399. Any personally identifiable information was redacted from the dataset prior to publication.

The published dataset covers the full scope of the data collected during our survey of natural history museum collections, including:

- Basic institution data for the 73 contributing institutions, including estimated total collection sizes, geographic locations (to city) and latitude/longitude, and Research Organization Registry (ROR) identifiers where available.
- Resourcing information, covering the numbers of research, collections and volunteer staff in each institution.
- Indicators of the presence and size of collections within each institution capture by a grid of 19 collection disciplines and 16 geographic regions.
- Measures of the depth and breadth of individual researcher experience across the same set of disciplines and geographic regions.

The dataset contains the data (raw and processed) collected for the survey, and specifications for the schema used to store the data. It includes:

1. A diagram of the MySQL database schema.
2. A SQL dump of the database schema, excluding the data.

3. A SQL dump of the database schema with all data. This may be imported into an instance of MySQL Server to create a full reconstruction of the database.
4. Raw data from each database table in CSV format.
5. A set of more human readable views of the data in CSV format. These largely correspond to the database tables, but foreign keys are substituted for values from the linked tables to make the data easier to read and analyse.

The global collections data may also be accessed at https://rebrand.ly/global-collections. This is a preliminary dashboard, constructed and published using Microsoft Power BI, that enables exploration of the data through a set of visualisations and filters. The dashboard consists of three pages:

Institutional profile: Enables the selection of a specific institution and provides summary information on the institution and its location, staffing, total collection size, collection breakdown and researcher expertise.

Overall heatmap: Supports an interactive exploration of the global picture, including a heatmap of collection distribution across the discipline and geographic categories, and visualisations that demonstrate the relative breadth of collections across institutions and correlations between with collection size and breadth. Various filters allow the focus to be refined to specific regions and collection sizes.

Browse: Provides some alternative methods of filtering and visualising the global dataset to look at patterns in the distribution and size of different types of collections across the global view.

Additional remarks:

The opinions or assertions contained herein are the private views of the authors, and are not to be construed as official, or as reflecting true views of the United States Department of the Army, Navy or the Department of Defense.

5. References

1. A. V. Suarez, N. D. Tsutsui, *BioScience*. **54**, 66–74 (2004).

2. K. G. Johnson *et al.*, *Bioscience*. **61**, 147–153 (2011).

3. D. P. Bebber *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22169–22171 (2010).

4. L. A. Rocha *et al.*, *Science*. **344**, 814–815 (2014).

5. M. W. Holmes *et al.*, *Mol. Ecol.* **25**, 864–881 (2016).

6. E. K. Meineke, T. J. Davies, B. H. Daru, C. C. Davis, *Philos. Trans. R. Soc. B Biol. Sci.* **374** (2019), doi:10.1098/rstb.2017.0386.

7. L. M. Page, B. J. MacFadden, J. A. Fortes, P. S. Soltis, G. Riccardi, *BioScience*. **65**, 841–842 (2015).

8. Committee on Biological Collections: Their Past, Present, and Future Contributions and Options for Sustaining Them, Board on Life Sciences, Division on Earth and Life Studies, National Academies of Sciences, Engineering, and Medicine, *Biological Collections:*

Ensuring Critical Research and Education for the 21st Century (National Academies Press, Washington, D.C., 2020; https://www.nap.edu/catalog/25592).

9. V. Blagoderov, I. J. Kitching, L. Livermore, T. J. Simonsen, V. S. Smith, *ZooKeys*. **209**, 133–146 (2012).

10. M. Staats *et al.*, *PLoS ONE*. **8** (2013), doi:10.1371/journal.pone.0069189.

11. P. D. N. Hebert *et al.*, *PLoS ONE*. **8** (2013), doi:10.1371/journal.pone.0068535.

12. M. Heerlien *et al.*, *J. Comput. Cult. Herit.* **8**, 3–3 (2015).

13. D. E. Schindel, J. A. Cook, *PLoS Biol.* **16**, 1–8 (2018).

14. D. C. Card, B. Shapiro, G. Giribet, C. Moritz, S. V. Edwards, *Annu. Rev. Genet. Vol 55*. **55**, 633–659 (2021).

15. P. S. Soltis, *Am. J. Bot.* **104**, 1281–1284 (2017).

16. K. R. Zamudio *et al.*, *Science*. **361**, 1322–1323 (2018).

17. J. Sthanapati, *Sci. Report.*, 21–24 (2016).

18. R. Dalton, *Nature*. **423**, 575–575 (2003).

19. Secretariat of the Convention on Biological Diversity United Nations Environmental Programme, *Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity* (2011).

20. F. Sarr, B. Savoy, "Rapport sur la restitution du patrimoine culturel africain. Vers une nouvelle éthique relationnelle" (2018).

21. A. H. Ariño, *Biodivers. Inform.* **7**, 81–92 (2010).

22. M. J. Costello, R. M. May, N. E. Stork, *Science*. **339**, 413–416 (2013).

23. D. P. Bebber, J. R. I. Wood, C. Barker, R. W. Scotland, *New Phytol.* **201**, 700–706 (2014).

24. A. K. Monfils *et al.*, *Bioscience*. **70**, 1045–1047 (2020).

25. M. H. Sabaj, *Copeia*. **108**, 593–669 (2020).

26. B. Thiers, "The World's Herbaria 2019: A Summary Report Based on Data from Index Herbariorum, Issue 3.0. http://sweetgum.nybg.org/science/docs/The_Worlds_Herbaria_2019.pdf" (2020).

27. M. Grosjean, M. Høfft, M. Gonzalez, T. Robertson, A. Hahn, *Biodivers. Inf. Sci. Stand.* **5** (2021), doi:10.3897/biss.5.74354.

28. U. N. E. P. and I. L. R. Institute, "Preventing the Next Pandemic: Zoonotic diseases and how to break the chain of transmission. Nairobi, Kenya." (2020), (available at https://www.unep.org/resources/report/preventing-future-zoonotic-disease-outbreaks-protecting-environment-animals-and).

29. J. A. Cook *et al.*, *BioScience*. **70**, 531–534 (2020).

30. I. M. Fund, IMF sees cost of COVID pandemic rising beyond $12.5 trillion estimate. **2022** (2022), (available at https://www.reuters.com/business/imf-sees-cost-covid-pandemic-rising-beyond-125-trillion-estimate-2022-01-20/).

31. Covid-19 Task Force: Communities taking action. *Consort. Eur. Taxon. Facil. - CETAF*, (available at https://cetaf.org/template-activities/cetaf-projects/covid19-taf-communities-taking-action/).

32. C. W. Thompson *et al.*, *mBio*. **12**, e02698-20 (2021).

33. C. Moritz *et al.*, *Science*. **322**, 261–264 (2008).

34. T. Newbold *et al.*, *Nature*. **520**, 45-+ (2015).

35. R. C. Godfree *et al.*, *Nat. Commun.* **12**, 1023 (2021).

36. Y. Wong, J. Rosindell, *Methods Ecol. Evol.* **13**, 303–313 (2022).

37. Z.-Q. Zhang, *Zootaxa*. **3703** (2013), doi:10.11646/zootaxa.3703.1.3.

38. C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, B. Worm, *PLoS Biol.* **9**, 1–8 (2011).

39. R. Fisher *et al.*, *Curr. Biol.* **25**, 500–505 (2015).

40. W. Appeltans *et al.*, *Curr. Biol.* **22**, 2189–2202 (2012).

41. A. Goodwin *et al.*, *Sci. Rep.* **11**, 13656 (2021).

42. G. H. Pyke, P. R. Ehrlich, *Biol. Rev. Camb. Philos. Soc.* **85**, 247–266 (2010).

43. D. L. Huestis *et al.*, *Nature*. **574**, 404–408 (2019).

44. B. B. Pagac *et al.*, *Acta Trop.* **213**, 105739 (2021).

45. E. E. Hahn *et al.*, *Mol. Ecol. Resour.* (2021), doi:10.1111/1755-0998.13505.

46. H. R. Kates *et al.*, *Front. Plant Sci.* **12** (2021) (available at https://www.frontiersin.org/articles/10.3389/fpls.2021.669064).

47. R. A. Folk *et al.*, *Appl. Plant Sci.* **9**, e11410 (2021).

48. K. A. Speer *et al.*, *Front. Ecol. Evol.* **10** (2022), doi:10.3389/fevo.2022.953131.

49. T. L. Rubi, L. L. Knowles, B. Dantzer, *Mol. Ecol. Resour.* **20**, 1161–1170 (2020).

50. I. Heumann, H. Stoecker, M. Tamborini, M. Vennen, *Dinosaurierfragmente - ZUR GESCHICHTE DER TENDAGURU-EXPEDITION UND IHRER OBJEKTE, 1906-2018* (2018; https://www.wallstein-verlag.de/9783835332539-dinosaurierfragmente.html).

51. I. Heumann, H. Stoecker, M. Vennen, *Vipande vya Dinosaria Historia ya Msafara wa Kpaleontolojia Kwenda Tendaguru Tanzania 1906 – 2018* (Mkuki na Nyota, Dar es Salaam, 2021; https://www.africanbookscollective.com/books/vipande-vya-dinosaria).

52. R. K. Brummitt, *World Geographical Scheme for Recording Plant Distributions* (Hunt Institute for Botanical Documentation, ed. 2, Carnegie Mellon University, 2001); http://www.tdwg.org.

53. Flanders Marine Institute Global Oceans and Seas, version 1 (2001): Available online at www.marineregions.org. https://doi.org/10.14284/542.

54. G. Nelson, S. Ellis, *Philos. Trans. R. Soc. B Biol. Sci.* **374**, 2–10 (2019).Z

55. C. L. Davis, R. P. Guralnick, E. F. Zipkin, *J. Anim. Ecol.* (2022), doi:10.1111/1365-2656.13763.

56. R. Bommarco, G. Vico, S. Hallin, *Glob. Food Secur.-Agric. Policy Econ. Environ.* **17**, 57–63 (2018).

57. A. Jankielsohn, *Adv. Entomol.* **06**, 62–73 (2018).

58. A. Y. Kawahara, L. E. Reeves, J. R. Barber, S. H. Black, *Proc. Natl. Acad. Sci.* **118**, e2002547117 (2021).

59. D. L. Wagner, E. M. Grames, M. L. Forister, M. R. Berenbaum, D. Stopak, *Proc. Natl. Acad. Sci.* **118**, e2023989118 (2021).