

# Science Review

2020



An annual compilation of  
research data uses enabled by



**GBIF**



# GBIF Science Review 2020

GBIF Secretariat

February 2021

**Citation:**

GBIF Secretariat. [2021]. GBIF Science Review 2020. <https://doi.org/10.35035/bezp-jj23>

**Licence:**

The GBIF Science Review is licensed under a Creative Commons Attribution 4.0 International [CC BY 4.0] license.

**Cover photo:**

[Pangolin](#) (*Manis pentadactyla* subsp. *pentadactyla*) by CHUNG CHANG-LIN via iNaturalist. Licensed under a Creative Commons Attribution-NonCommercial 4.0 International [CC BY-NC 4.0] license.

## Foreword

This edition of the GBIF Science Review presents a selection of some of the key research uses of GBIF-mediated data published during calendar year 2019. The COVID-19 pandemic forced the GBIF Secretariat to reassess our normal work timelines in 2020. The Review itself was not unaffected, but that has led to a happy coincidence.

As we go to press (or "press"), the Proceedings of the National Academy of Sciences has published a comprehensive analysis and review of more than 4,000 peer-reviewed studies published between 2003 and 2019 that rely on data from the GBIF network (<https://doi.org/10.1073/pnas.2018093118>). Led by J. Mason Heberling of the Carnegie Museum of Natural History assisted by Scott B. Weingart of Carnegie Mellon University and Secretariat staff, this paper explores the wider network of interdisciplinary knowledge generated through the efforts of GBIF's formal participants and data publishers in more than 130 countries.

The data uses featured in the 2020 Science Review represent part of an impressive and rapidly growing corpus of work. The data resources on which they're built supported 743 peer-reviewed publications in 2019—a rate of roughly two research articles every day. In comparison, the cumulative number of studies published between 2003 and 2009 amounted to just 157 papers.

As even this small sample of 67 summaries shows, the range of research investigations explored with the help of data from the GBIF network remains diverse and dynamic. We hope you find insight and inspiration in equal measure—and, of course, where data references #CiteTheDOI, you can also find the evidence base and sources of those studies. Happy hunting!


Joe Miller  
GBIF Executive Secretary



BALTIMORE ORIOLE (*ICTERUS GALBULA*) OBSERVED IN KENNEBUNK, ME, USA BY KEN JANES. PHOTO VIA INATURALIST (CC BY-NC 4.0)

## About the *Science Review*

The GBIF *Science Review* provides an annual survey drawn from the Secretariat's ongoing literature tracking programme, which identifies research uses and citations of biodiversity information accessed through GBIF's global infrastructure. The peer-reviewed articles summarized in the following pages offer a partial but instructive view of research investigations supported and enabled by free and open access to biodiversity data from the GBIF network.

We have labelled open-access scientific articles using the symbol . We feel this step serves those interested in reading the research at a time of changing institutional journal subscriptions.

The comprehensive list of this year's uses (as well as current and previous ones) is separate from the printed version of the *Science Review* and is available exclusively through the online literature index at <https://www.gbif.org/resource/search?contentType=literature>.

As in previous years, the categories used here are intended to help readers navigate the major subject areas of GBIF-assisted research, despite the fact that some papers may cut across multiple topics. For clarity's sake, articles appear under only one category in the *Review*. Countries assigned to authors are based on the location of the institutions identified in the author information, while funding information included for the highlighted papers draws upon the papers' acknowledgements.

Those interested in sharing research uses that may have escaped our attention can write to us at [communication@gbif.org](mailto:communication@gbif.org).

---

## CONTENTS

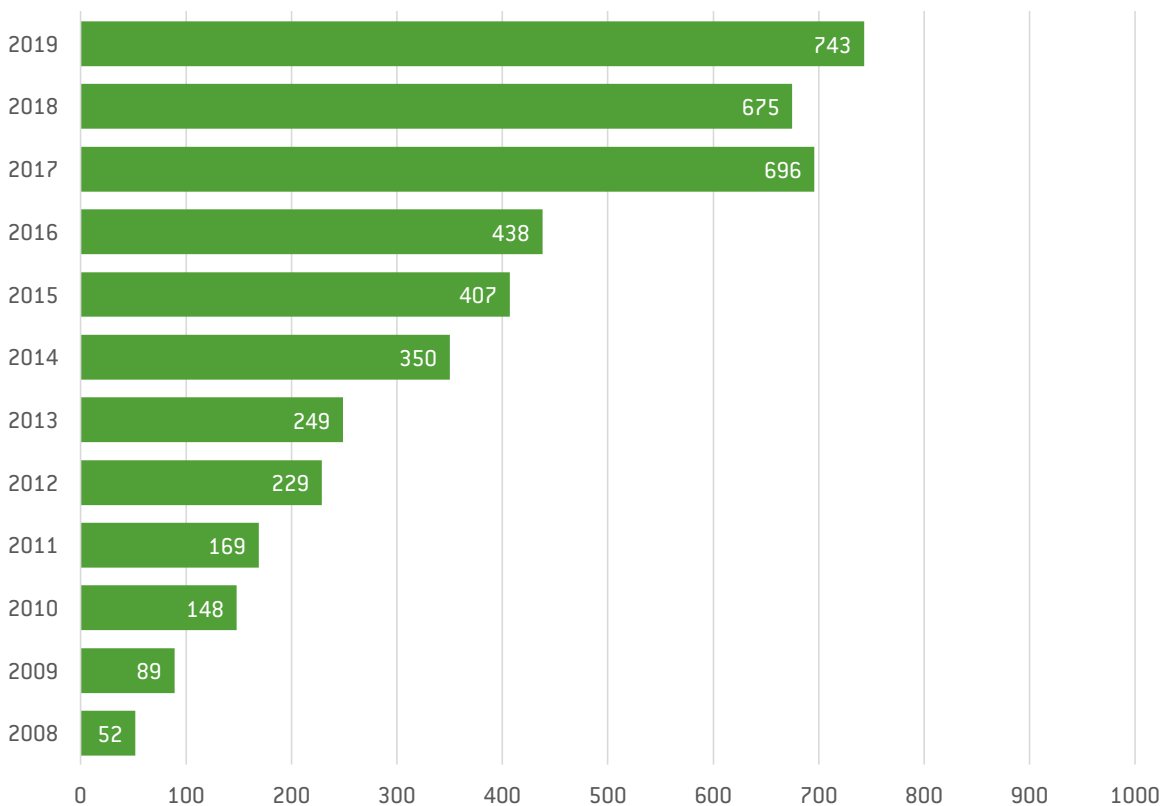
Uses and trends.....	6
Biodiversity conservation.....	8
Ecology .....	14
Environmental sciences.....	22
Evolutionary biology .....	24
Multidisciplinary sciences .....	27
Other research topics .....	34



## Uses and trends

These visualizations of the use of GBIF-mediated data serve to show not just the growing number of peer-reviewed publications appearing each year, but also the increasing geographic distribution of institutions hosting authors who apply the data in their research articles. Each trend highlights the GBIF network's widening sphere of influence and impact on scientific research related to the biological domain.

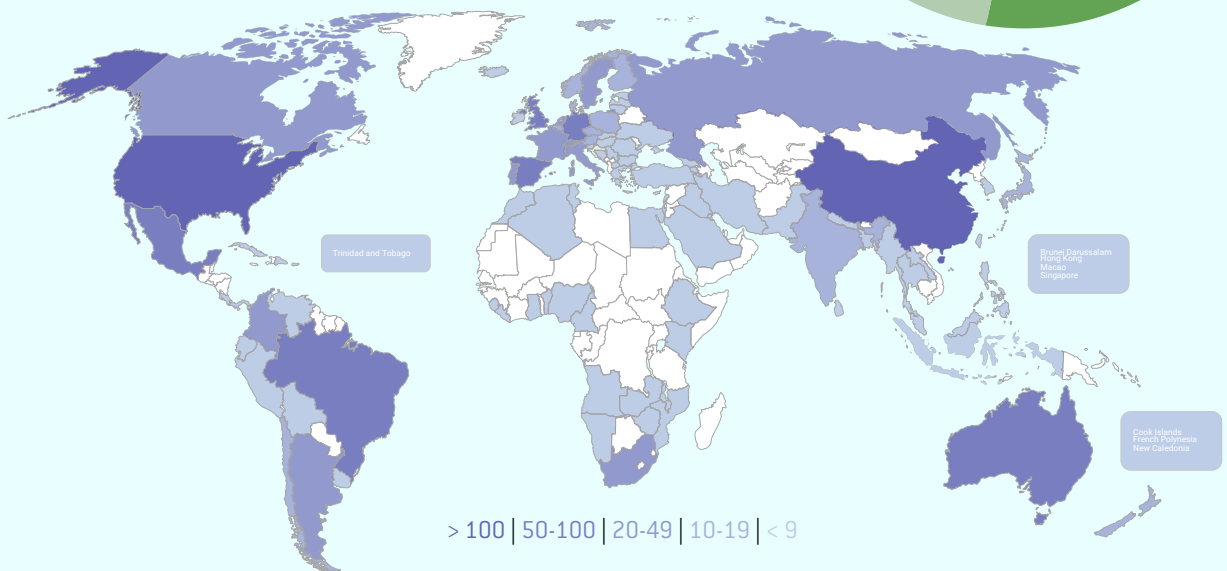
### ANNUAL NUMBER OF PEER-REVIEWED ARTICLES USING GBIF-MEDIATED DATA



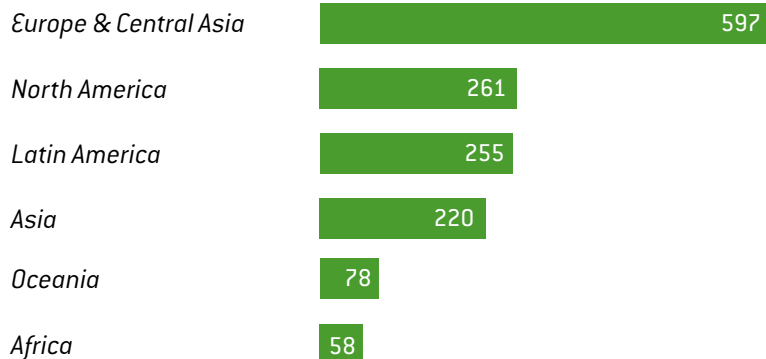
## NUMBER OF ARTICLES WITH AUTHORS BY COUNTRY/ISLAND/TERRITORY

countries, islands & territories with authors who used GBIF-mediated data in peer-reviewed papers published in 2019

# 104



## NUMBER OF ARTICLES WITH AUTHORS BY REGION



## Biodiversity Conservation



*SCINAX FUSCOVARIUS* OBSERVED IN ALTAMIRO DE MOURA PACHECO STATE PARK, BRAZIL BY WERTHER. PHOTO VIA INATURALIST (CC BY-NC 4.0)

### THE ROLE OF NATURAL HISTORY COLLECTIONS IN BIODIVERSITY KNOWLEDGE GENERATION IN ANGOLA

DATA USED: 149,701 SPECIES OCCURRENCES

 Figueira R and Lages F (2019) Museum and Herbarium Collections for Biodiversity Research in Angola. Biodiversity of Angola. Springer International Publishing 513–542. Available at: [https://doi.org/10.1007/978-3-030-03083-4\\_19](https://doi.org/10.1007/978-3-030-03083-4_19)

Author countries/areas: Angola, Portugal

Natural history collections in museums and herbaria are of great importance to biodiversity science, particularly in countries and regions with high levels of endemism and rare species. This is especially true for Angola with its unique mixture of landscapes where tropical humid climates interface with arid deserts.

Authors of this book chapter set out to strengthen knowledge of Angolan natural history collections and their role in biodiversity knowledge generation. Reviewing the history of collections in Angola, the authors provide a detailed account of 150 years of expeditions and studies in the country, while using GBIF-mediated occurrence data to summarize the current status of collections of Angolan biodiversity—held both in local institutions and in other countries.

In addition to simply preserving and documenting biodiversity, the chapter highlights how collections underpin multiple branches of scientific research and conservation science, including supporting sustainable food production and evaluating impacts of environmental changes.

### ASSESSING THE FUNCTIONAL IMPACT ON ECOSYSTEMS OF SUNDALAND FRESHWATER SPECIES LOSS

DATA USED: 893 SPECIES

Chua KWJ, Tan HH and Yeo DCJ (2019) Loss of endemic fish species drives impacts on functional richness, redundancy and vulnerability in freshwater ecoregions of Sundaland. Biological Conservation. Elsevier BV 234: 72–81. Available at: <https://doi.org/10.1016/j.biocon.2019.03.019>

Author countries/areas: Singapore

Not only a terrestrial biodiversity hotspot, the biogeographical region of Sundaland in southeast Asia is also a global hotspot for freshwater diversity. Large-scale agricultural conversion of rainforests poses a major threat to freshwater fishes and the ecosystems they help support.

By assessing functional richness, redundancy and vulnerability of Sundaic freshwater fishes,



researchers behind this study give insights into the current status of ichthyofaunal diversity and potential impacts of species loss on ecosystems.

Compiling knowledge from a wide range of sources—including GBIF—of all occurring fishes across 13 ecoregions, the authors describe functional traits for 893 species. The distribution of trait entities varies from less than 30 per cent of the total variation in trait entities in Java to more than 85 per cent in the Malay peninsula.

When modelling loss of extinction-prone species, they show a general pattern of accelerating impacts on functional diversity. With low levels of functional redundancy and thus high vulnerability to impacts of species loss, the Malay Peninsula Eastern Slope is singled out as the most urgent target for conservation efforts.

### PROTECTING FROGS OF THE FUTURE IN BRAZILIAN BIOMES

DATA USED: 500 SPECIES

Vasconcelos TS and Prado VHM (2019) Climate change and opposing spatial conservation priorities for anuran protection in the Brazilian hotspots. *Journal for Nature Conservation*. Elsevier BV 49: 118–124. Available at: <https://doi.org/10.1016/j.jnc.2019.04.003>

Author countries/areas: Brazil

Research funding: Conselho Nacional de Desenvolvimento Científico e Tecnológico

Distinguished by high levels of diversity and endemism, the Atlantic Forest and Cerrado biomes in Brazil are recognized as hotspots of biodiversity. Home to hundreds of frog species, many of which are endemic, these regions and their anuran fauna are of particular interest to conservation biologists.

Modelling the ecological niches of more than 500 frog species, authors of this study outline conservation priorities across space and time. Their results show that for both biomes in the baseline climate, total representation is achievable in just ~9 per cent of 50x50 km grid cells.

In future climates, the Atlantic Forest anurans will require an increased area in the number of grid cells for total coverage, whereas the opposite is the case for the Cerrado. In fact, fewer than a third of the baseline Cerrado cells appear to have relevance for anuran conservation across all time frames.

### IMPACTS OF GLOBAL LAND-USE AND CLIMATE CHANGE ON PLANT BIODIVERSITY

DATA USED: 52 MILLION SPECIES OCCURRENCES

Di Marco M, Harwood TD, Hoskins AJ, Ware C, Hill SLL and Ferrier S (2019) Projecting impacts of global climate and land-use scenarios on plant biodiversity using compositional-turnover modelling. *Global Change Biology*. Wiley 25(8): 2763–2778.

Available at: <https://doi.org/10.1111/gcb.14663>

Author countries/areas: Australia, Italy, UK

In response to accelerating rates of biodiversity loss, the global community is considering new, ambitious conservation targets. To inform policy decisions, fine-scale assessments predicting future impacts of threats to biodiversity are therefore crucial.

This study examined the effects of climate change and land-use since 1900 on global vascular plant biodiversity using models of compositional turnover, based on 52 million GBIF-mediated species occurrences. By applying future scenarios of different degrees of climate and land-use change, the authors used the models to predict the proportion of vascular plants expected to go extinct by 2070.

Between 1900 and 2015 plant extinctions increased by 60 per cent. When considering land-use scenarios alone, sustainable socio-economic development could slow this down moving forward. Unfortunately, under all considered climate change scenarios, rates of extinction are predicted to be 3-4 times higher than in the most pessimistic land-use scenario alone.

The combined results of the paper suggest that unless climate change is halted, efforts to introduce sustainable land-use planning will not be enough to prevent dramatic biodiversity loss.

### USING RARITY AS A SURROGATE MEASURE FOR BIODIVERSITY

DATA USED: 2,380,424 SPECIES OCCURRENCES

📄 Astudillo-Scalia Y and de Albuquerque FS (2019) Evaluating the performance of rarity as a surrogate in site prioritization for biodiversity conservation. *Global Ecology and Conservation*. Elsevier BV 18: e00639. Available at: <https://doi.org/10.1016/j.gecco.2019.e00639>

Author countries/areas: USA

In conservation planning, several metrics can be used when seeking to identify how to cover the greatest number of species in the least number of sites. Widely used, species richness, however, appears to be one of the least effective.

(continued on next page...)

This study assesses the performance of three rarity-based indices as new potential surrogates of biodiversity for solving the minimum-set coverage problem: rarity-weighted richness (RWR), index of summed rarity (ISR) and index of relative rarity (IRR). The authors tested the indices in 14 datasets—eight mediated by GBIF—spanning a broad range of taxa, spatial extent and resolution.

In all 14 datasets, rarity indices outperformed species richness and in some cases, also complementarity algorithms. The best performing index was RWR.

With few or no requirements for specialized software, computing power or programming skills, the proposed indices may represent a simple, yet reliable alternative to prohibitively long computation times of integer programming and heuristic algorithms in site prioritization for biodiversity conservation.

### GLOBAL ANALYSIS OF GAPS IN BUTTERFLY DATA

**DATA USED:** 41,077,368 SPECIES OCCURRENCES

Girardello M, Chapman A, Dennis R, Kaila L, Borges PAV and Santangeli A (2019) Gaps in butterfly inventory data: A global analysis. *Biological Conservation*. Elsevier BV 236: 289–295. Available at:

<https://doi.org/10.1016/j.biocon.2019.05.053>

**Author countries/areas:** Finland, Greece, Portugal, UK

Biodiversity data is the main pillar of conservation planning. The quality of species assessments, however, relies on the availability and standard of occurrence data. If data is scarce—the Wallacean Shortfall—results may be unreliable.

Large, colourful and with plenty of aesthetic appeal, butterflies are a remarkable exception to undersampling of insects. In this study, researchers used GBIF-mediated occurrences of six butterfly families containing more than 8,000 species to identify gaps in inventory data at the global level.

The authors found a positive trend in record accumulation in the past century—particularly since 1990—in most of the world with the exception of Central Africa. The highest levels of inventory completeness are found in Europe, North America, coastal Australia and Southern Africa.

While using regression models to identify predictors of inventory completeness, the authors found road density (i.e. accessibility) to be important, however, this was only at the finest resolution. At a large scale, completeness was better predicted by elevation and density of protected areas.

### FILLING KNOWLEDGE GAPS WITH ROADKILL DATA

**DATA USED:** 26,853,168 SPECIES OCCURRENCES

Tiedeman K, Hijmans RJ, Mandel A, Waetjen DP and Shilling F (2019) The quality and contribution of volunteer collected animal vehicle collision data in ecological research. *Ecological Indicators*. Elsevier BV 106: 105431. Available at:

<https://doi.org/10.1016/j.ecolind.2019.05.062>

**Author countries/areas:** USA

Knowledge about the distributions of species is increasingly being improved thanks to data collected by volunteers through citizen science projects. Dominated by non-experts, some projects rely on experts for validation while others verify identifications through community efforts.

This study seeks to evaluate the quality of a volunteer-based animal-vehicle collision programme—the California Roadkill Observation System (CROS)—and its contribution to knowledge about species ranges. Assessing more than 35,000 observations of more than 400 species, the authors found that species identification accuracy was high for both professional (99 per cent) and non-professional users (95 per cent).

The ranges derived from the CROS dataset overlapped to a large extent with GBIF-mediated occurrences. However, for 139 species the roadkill data expanded ranges by a mean of 3,735 km<sup>2</sup>. For some ten per cent of species, this would increase their environmental niche—0.13 degrees C for mammals, suggesting a clear significance of the dataset, that if published to GBIF could fill gaps in knowledge.

### ASSESSING SAMPLING EFFORT IN THE CAATINGA USING IGNORANCE SCORES

**DATA USED:** 1,218,897 SPECIES OCCURRENCES

Correia RA, Ruete A, Stropp J, Malhado ACM, dos Santos JW, Lessa T, Alves JA and Ladle RJ (2019) Using ignorance scores to explore biodiversity recording effort for multiple taxa in the Caatinga. *Ecological Indicators*. Elsevier BV 106: 105539. Available at:

<https://doi.org/10.1016/j.ecolind.2019.105539>

**Author countries/areas:** Brazil, Iceland, Portugal, Sweden

Making informed decisions in conservation planning requires good knowledge about species distributions. Gaps and biases in spatial data, however, may skew results and affect our ability to accurately predict the distribution of species.

This study explores the use of ‘ignorance scores’ to evaluate sampling effort and bias—in a case study of species present in the Caatinga semi-arid tropical

forest in Brazil. Downloading all GBIF-mediated occurrences in the region, researchers organized the data into taxonomic reference groups of similar collection methods and calculated scores for each 10x10 km cell.

Their results showed a staggering taxonomic bias with high ignorance scores across the vast majority of cells for all groups, except plants—suggesting a preference towards plant recording in the region. For amphibians, the percentage of grid cells with no records was more than 99 per cent.

Exploring reasons for recording biases, the authors point to road density as a main factor, while population density and distance to nearest university are secondary predictors.

### TARGETING THE KNOWN UNKNOWNNS OF THE BOTANICAL WORLD

**DATA USED:** 214,964,486 SPECIES OCCURRENCES

Cornwell WK, Pearse WD, Dalrymple RL and Zanne AE (2019) What we (don't) know about global plant diversity. *Ecography*. Wiley 42(11): 1819–1831. Available at: <https://doi.org/10.1111/ecog.04481>

**Author countries/areas:** Australia, USA

This study by an Australian and U.S. team compares the coverages and overlaps between the most comprehensive sources of geographic, genetic and trait-based botanical data to provide an actionable view of the current presence and absence of knowledge about the world's plants.

Their analysis draws on nearly 215 million plant records from GBIF.org and compares their coverage with data aggregated in GenBank and TRY, placing each of the 350,699 species names in The Plant List into one of three groups: 1) broadly covered species (17.7 per cent) where at least some knowledge of their locations, genes and traits exists; 2) patchily covered species (55.6 per cent) missing from at least one data source; or 3) species with no information other than their names (26.7 per cent).

While the GBIF network provides the most complete species coverage (73.3 per cent), the results outline the shapes of some well-known but previously unquantified shortfalls in botanical knowledge. By providing clear targets for data collection and specimen digitization, the authors present reasons to be optimistic about their stated hope of "turn[ing] what appears to be an insurmountable task into a manageable checklist of gaps to be filled."

### PUTTING A PRICE ON NATURE'S CONTRIBUTION TO PEOPLE

Ⓐ Hipólito J, Sousa B dos SB, Borges RC, Brito RM de, Jaffé R, Dias S, Imperatriz Fonseca VL and Giannini TC (2019) Valuing nature's contribution to people: The pollination services provided by two protected areas in Brazil. *Global Ecology and Conservation*. Elsevier BV 20: e00782. Available at: <https://doi.org/10.1016/j.gecco.2019.e00782>

**Author countries/areas:** Brazil

**Research funding:** Conselho Nacional de Desenvolvimento Científico e Tecnológico

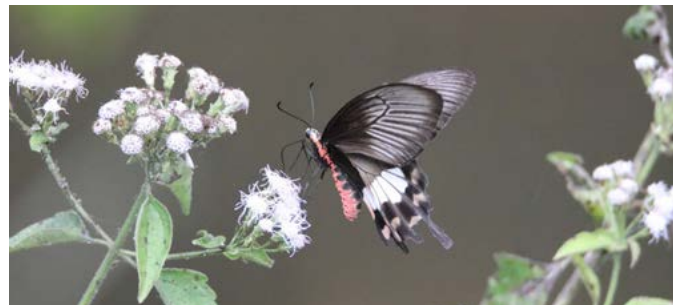
Directly contributing to human well-being, pollination is one of the most studied ecosystem services, the effect of which can be measured in order to incentivize preservation.

Focused on two protected areas in Brazil—Serra da Bocaina in the Amazon Forest and Mata do Jambreiro in the Atlantic Forest—this study calculates both monetary and non-monetary values of pollination services provided by bees in the surroundings.

Using data on land cover and use, the authors estimated the size of agricultural areas and their dependence on animal pollination. Derived from GBIF-mediated occurrences, they compiled a list of bee species and estimated their maximum foraging distance.

In Serra da Bocaina, they identified 17 agricultural crops with varying levels of pollination dependence, the annual value of which is estimated to \$564,500. For Mata do Jambreiro, 16 crops are cultivated, of which half depend on pollination—estimated at a value of \$246,000 per year.

The non-monetary valuation of pollination services provided by the two protected areas identified important areas with higher potential for nesting and food resources in Canaã dos Carajás and Parauapebas for Bocaina—and, especially, Nova Lima for Jambreiro.



COMMON ROSE [*PACHLIOPTA ARISTOLOCHIAE*] OBSERVED IN THAILAND BY T.V.P. PHOTO VIA INATURALIST (CC BY-NC-ND 4.0)

## HOW LAND USE AFFECTS BIODIVERSITY IN DRYLANDS AND MEDITERRANEAN ENVIRONMENTS

**a** García-Vega D and Newbold T (2019) Assessing the effects of land use on biodiversity in the world's drylands and Mediterranean environments. *Biodiversity and Conservation*. Springer Science and Business Media LLC 29(2): 393–408. Available at: <https://doi.org/10.1007/s10531-019-01888-4>

**Author countries/areas:** UK

**Research funding:** Natural Environment Research Council

Human land uses, such as agriculture, urbanization and plantation forests, are expanding worldwide, destroying habitats and leading to local species extinctions. The effects of land use on biodiversity, however, vary across ecosystem types, as a crop field may reduce biodiversity in tropical rainforests but may increase it in a desert.

This study focuses on drylands and Mediterranean ecosystems, estimated to have experienced the largest land-use derived declines in biodiversity to date. Assembling a dataset of biodiversity surveys in these biomes, the authors create a model comparing species richness, abundance, diversity, endemism and compositional turnover across different types of land use.

To assess effects on endemism and composition, the authors used GBIF-mediated occurrence data to define species range occupancy, which was also modelled as a function of land use.

The overall results showed strong negative responses to land use in Mediterranean environments. Relative to undisturbed natural habitat, all measured parameters showed decline across all land uses. For drylands, however, the results showed a strong turnover in species composition, but no significant response in any other measures.

## CLIMATE CHANGE THREATENS THE MOST BIODIVERSE REGIONS OF MEXICO

**DATA USED:** 1,927,034 SPECIES OCCURRENCES

Esperon-Rodriguez M, Beaumont LJ, Lenoir J, Baumgartner JB, McGowan J, Correa-Metrio A and Camac JS (2019) Climate change threatens the most biodiverse regions of Mexico. *Biological Conservation*. Elsevier BV 240: 108215. Available at: <https://doi.org/10.1016/j.biocon.2019.108215>

**Author countries/areas:** Australia, France, Mexico

Climate change poses a major threat to global biodiversity, however, identifying and conserving potential stable climate refugia may help limit biodiversity losses.

Harbouring nearly 10 per cent of the world's biodiversity, Mexico is a megadiverse country. In this study, researchers focused on five bioclimatic variables to assess the vulnerability to climate change of 40 protected areas (PA) in Mexico. Using GBIF-mediated occurrences in the same areas, the authors analysed species richness and composition under current and future climates.

Their results showed temperature increases for all PAs, and for 31 PAs—habitat to more than 22,000 species—the climate will change completely by 2050. The remaining nine PAs are predicted to retain about half their baseline climatic space, and may to some degree constitute potential refugia.

Species composition varied among PAs and the most vulnerable appeared to cluster broadly in the central mountainous regions and the tropical south-eastern part of Mexico. Considering all factors, the study points to los Tuxtlas as the most vulnerable protected area, and Tiburon Ballena as the least vulnerable.

## ESTIMATING RANGE SIZES USING OPEN ACCESS BIODIVERSITY DATA

**a** Smith JA, Benson AL, Chen Y, Yamada SA and Mims MC (2020) The power, potential, and pitfalls of open access biodiversity data in range size assessments: Lessons from the fishes. *Ecological Indicators*. Elsevier BV 110: 105896. Available at: <https://doi.org/10.1016/j.ecolind.2019.105896>

**Author countries/areas:** USA

**Research funding:** Natural Environment Research Council

When prioritizing conservation efforts, the range size of a species is often used as a measure of rarity and to determine its intrinsic risk of extinction. How ranges are defined or estimated, however, may affect how rare or common a species is seen to be, and whether conservation efforts are justified.

Stream-dwelling organisms, such as freshwater fishes constrained to river networks, may present a particular challenge, and in this study, researchers compared different methods of range size estimation for freshwater species.

Initially selecting a wide taxonomic and geographic representation of species, the authors gathered GBIF-mediated occurrences of 128 fish species in the conterminous United States and calculated range sizes based on minimum convex polygons.

They then compared the GBIF-based results to range sizes derived from NatureServe maps—considered the best available estimates of current distributions—

finding strong correlations between the approaches with no detectable taxonomic bias.

Importantly, both analytical approaches consistently identified the rarest and perhaps most vulnerable species, suggesting an important role of open data in filling fundamental knowledge gaps, especially for poorly understood taxa.



*PELEGRINA AENEOLA* OBSERVED IN GOAT MARSH, WASHINGTON, USA BY CORNDOG. PHOTO VIA INATURALIST CC BY-NC 4.0.

### PLANT INVENTORY COMPLETENESS AND COLLECTION PATTERNS IN TROPICAL EAST AFRICA

DATA USED: 32,536,833 SPECIES OCCURRENCES

Wang S, Zhou Y, Musili PM, Mwachala G, Hu G and Wang Q (2020) Inventory incompleteness and collecting priority on the plant diversity in tropical East Africa. *Biological Conservation*. Elsevier BV 241: 108313. Available at: <https://doi.org/10.1016/j.biocon.2019.108313>

Author countries/areas: China, Kenya

Research funding: National Natural Science Foundation of China

Primary biodiversity data is crucial to understanding biogeographic patterns of species richness and for being able to predict impacts of climate change or anthropogenic activity on biodiversity.

Tropical East Africa has a very rich flora with important hotspots. In this study, researchers use plant data from GBIF and other sources to assess the completeness and collection patterns in the region.

Authors of the study find extremely uneven collection densities across the region—with very low density in northern Kenya and southern Tanzania. By comparing richness observed in occurrence data with predicted richness modelled using Maxent, the authors reveal that collection effort in half of East Africa is indeed incomplete.

Species richness was the most important factor in explaining both collection density (positive correlation) and inventory incompleteness (negative correlation) across the region. The authors highlight the importance of digitization and prioritizing future collection efforts, especially in low-density areas of Kenya and Tanzania

### USEFULNESS OF GBIF-MEDIATED OCCURRENCE DATA IN SPIDER RED LISTING

DATA USED: 4,528,821 SPECIES OCCURRENCES

a Shirey V, Seppälä S, Branco V and Cardoso P (2019) Current GBIF occurrence data demonstrates both promise and limitations for potential red listing of spiders. *Biodiversity Data Journal*. Pensoft Publishers 7. Available at: <https://doi.org/10.3897/bdj.7.e47369>

Author countries/areas: Finland, USA

Assessing the extinction risk status of a species, and prioritizing conservation efforts, require knowledge on species populations and trends. With such data often lacking, GBIF-mediated occurrence data can aid in documenting species ranges and changes over time.

In this study, a group of researchers—including one of the 2020 GBIF Young Researchers Award winners—assessed the usefulness of GBIF-mediated occurrences in deriving IUCN Red List classifications based on Extent of Occurrence (EOO) calculations in a largely under-sampled taxon—spiders (Araneae)—by comparing hypothetical Red List classifications derived from literature, GBIF data, and both combined.

The analysis showed that while GBIF data alone was not sufficient for 88.7 per cent of the study species, it enabled assessment of some species (3.3 per cent) not classifiable based on literature alone. In a few species, the addition of GBIF data changed the hypothetical classification—in one species, *Myrmarachne bicolor*, from Vulnerable to Least Concern.

The results show the potential of using GBIF-mediated data in extinction risk assessments, while highlighting the importance of promoting increased digitization and mobilization of data to GBIF.



## Ecology



PRETZEL SLIME MOLD (*HEMITRICHIA SERPULA*) BY ALEXANDER SHIROKIKH VIA INATURALIST. PHOTO LICENSED UNDER CC BY-NC 4.0.

### COORDINATECLEANER: FAST AND STANDARDIZED CLEANING OF SPECIES OCCURRENCE DATA

**DATA USED:** 116M SPECIES OCCURRENCES

Zizka A, Silvestro D, Andermann T, Azevedo J, Duarte Ritter C, Edler D, Farooq H, Herdean A, Ariza M, Scharn R, Svanteson S, Wengström N, Zizka V and Antonelli A (2019) CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*. Wiley. Available at: <https://doi.org/10.1111/2041-210X.13152>

**Author countries/areas:** Germany, Norway, Sweden, Switzerland

With more than two scientific articles relying on species occurrence data from GBIF.org published every day, ensuring the data quality of this indispensable resource is critical. While small datasets may be inspected visually, this is impractical—if not impossible—for studies using thousands or millions of records.

Faced with this challenge, a team of researchers including the [2nd prize winner of the 2016 Ebbe Nielsen Challenge](#) developed [CoordinateCleaner](#)—a new open source R library for standardized cleaning of species occurrence records. Based on a reference database of coordinates indicating typical geolocation problems, e.g. country centroids and capitals, as

well as coordinates of biodiversity institutions (e.g. museums), CoordinateCleaner quickly flags records with potential issues. The software also employs novel algorithms to spot datasets including records with systemized coordinate conversion and rasterized sampling bias.

In an example using all GBIF-mediated records of flowering plants (~91 million records) and the [Paleobiology Database](#) (19,000 records), the authors demonstrate how CoordinateCleaner flags 3.6 per cent of GBIF-mediated records and 6.3 per cent of PBDB records as having potential geolocation issues. At the contributing dataset level, the example flagged four per cent of datasets for potential coordinate conversion bias and 18.5 per cent of datasets for potential rounding or rasterization. While the output should not be taken as immediate grounds for exclusion, it provides researchers with a more objective means of flagging datasets for manual case-by-case validation.

CoordinateCleaner provides a comprehensive set of routines that can help researchers perform fast, systematic and reproducible cleaning of occurrence data for use in ecological, biogeographic and paleontological studies.

## DIVERSITY OF SLIME MOULDS—MUCH MORE THAN MEETS THE EYE

Shchepin ON, Schnittler M, Erastova DA, Prikhodko IS, Borg Dahl M, Azarov DV, Chernyaeva EN and Novozhilov YK (2019) Community of dark-spored myxomycetes in ground litter and soil of taiga forest (Nizhne-Svirskiy Reserve, Russia) revealed by DNA metabarcoding. *Fungal Ecology*. Elsevier BV 39: 80–93. Available at: <https://doi.org/10.1016/j.funeco.2018.11.006>

**Author countries/areas:** Germany, New Zealand, Russia

Myxomycetes—or slime moulds—are fascinating lifeforms with a complex life cycle involving a single-cell amoeba-like stage followed by multinuclear aggregates, culminating in a fruiting body-like structure with airborne spores restarting the cycle.

In a study of so-called dark-spored slime moulds, researchers used DNA metabarcoding to assess the diversity of myxomycetes in the lowland pine forests of Nizhne-Svirskiy Reserve, Russia. Sampling both ground litter and soil in separate study plots representing four different stages after a forest fire, followed by DNA extraction, amplification and sequencing, the authors produced more than 300,000 sequence reads based on myxomycete-specific primers. These sequences clustered to fewer than 200 operational taxonomic units (OTU), of which the authors identified 15 species and nine genera.

Comparing identified taxa against GBIF-mediated occurrences, the authors found four species new to the Leningrad region. While the myxomycete composition shifted significantly with the fire history of a plot, the authors were unable to find a clear chronosequence, suggesting that factors shaping assemblages are not simply a linear product of time passed since a fire.

The study identifies DNA metabarcoding as a useful tool in describing slime mould diversity, as invisible amoeba-stage populations are much more widespread than the recordings of visible fruiting bodies.

## THE GLOBAL DISTRIBUTION AND DRIVERS OF POLYPLOIDY IN PLANTS

**DATA USED:** 113,504,266 SPECIES OCCURRENCES

Rice A, Šmarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J and Mayrose I (2019) The global biogeography of polyploid plants. *Nature Ecology & Evolution*. Springer Science and Business Media LLC 3(2): 265–273.

Available at: <https://doi.org/10.1038/s41559-018-0787-9>

**Author countries/areas:** Czechia, Israel

Polyploidy—having more than two copies of each chromosome—is often lethal in animals, but not an uncommon phenomenon in plants. In fact, all angiosperm species are believed to have undergone one or more polyploidization events in their lifetime. Allopolyploidy—when the genomes of two different species merge and hybrids arise—can result in higher capacity for adaptation, but what are the drivers and how are the polyploid plants distributed globally?

Reconstructing a complete angiosperm phylogeny while gathering data on chromosome numbers and GBIF-mediated occurrences, this study maps the global distribution of polyploid frequency among nearly 135,000 plant species. The resulting maps show a clear latitudinal pattern with polyploidy increasing towards the poles. At the biome-level, most polyploid plants are found in tundras. Most ecoregions show the same patterns as their biomes—with some notable polyploid-rich tropical ecoregions, such as the Andes and Hawaii.

Exploring possible environmental drivers, the authors show that polyploidy is correlated with temperature, prevailing in cooler climates. The effect, however, appears predominantly mediated by the indirect impact of temperature on the floral composition of ecoregions.



A HEXAPLOID, CALIFORNIA REDWOOD (*SEQUOIA SEMPERVIRENS*) HAS SIX COPIES OF EACH CHROMOSOME. PHOTO BY VENETIAN3 (CC BY-NC 4.0).

## THE ROLE OF HORTICULTURE IN FACILITATING PLANT INVASIONS

Guo W, van Kleunen M, Pierce S, Dawson W, Essl F, Kreft H, Maurel N, Pergl J, Seebens H, Weigelt P and Pyšek P (2019) Domestic gardens play a dominant role in selecting alien species with adaptive strategies that facilitate naturalization. *Global Ecology and Biogeography*. Wiley 28(5): 628–639. Available at: <https://doi.org/10.1111/geb.12882>

**Author countries/areas:** China, Czechia, Denmark, Germany  
**Research funding:** Austrian Science Fund, Deutsche Forschungsgemeinschaft, Akademie Věd České Republiky, Grantová Agentura České Republiky

More than 13,000 naturalized alien species of vascular plants have been recorded worldwide, and horticulture is the major pathway of plant invasions. Botanic gardens play a vital role in conserving plant diversity and preventing extinction, but they also cultivate many species with invasion potential.

Examining the role of horticulture in plant invasions, this study rigorously examines the direct and indirect relationships between species traits, horticultural use, native range size and naturalization success. Using a dataset—downloaded from GBIF and other sources—of nearly 4,000 species with defined adaptive strategies, the authors reveal that species grown in botanic or domestic gardens are significantly more likely to become naturalized than those not cultivated.

The study finds that naturalization is linked to particular adaptation strategies among plant species, but that species using these strategies are also more likely to be cultivated. In other words—horticulture is not just a major pathway for invasions, it also tends to select alien species with the highest invasive potential.

## TERRESTRIAL HOTSPOTS OF EVOLUTIONARY DIVERSITY POINT TO NEW PRIORITIES FOR CONSERVATION

**DATA USED:** OCCURRENCES OF 6,483 GENERA

Daru BH, le Roux PC, Gopalraj J, Park DS, Holt BG and Greve M (2019) Spatial overlaps between the global protected areas network and terrestrial hotspots of evolutionary diversity. *Global Ecology and Biogeography*. Wiley 28(6): 757–766. Available at: <https://doi.org/10.1111/geb.12888>

**Author countries/areas:** Denmark, South Africa, UK, USA

Identifying biodiversity hotspots is a common approach for prioritizing conservation. Most such designations, however, rely on species-level metrics and largely ignore phylogenetics. Although not necessarily more informative than species-level hotspots, phylogenetic diversity hotspots can represent evolutionary history and, potentially, adaptive capacity.

Integrating data on phylogeny and geographical distribution of amphibians, mammals, birds and angiosperms mediated by GBIF, this study identifies 29 clusters of phylogenetic diversity worldwide. While a large proportion of these overlap with species-level hotspots, the study finds novel hotspots of phylogenetic diversity in Central Chile, Honshu (Japan), New Caledonia, the Appalachian Mountains and parts of Texas.

Across all taxonomic groups, the most phylogenetically diverse concentrations were found in tropical regions, and overall, fewer than 10 per cent of phylogenetic diversity hotspots are designated as protected areas.



COMMON STAR-OF-BETHEHEM (*ORNITHOGALUM UMBELLATUM*) OBSERVED IN ATLANTA, USA BY MRSDYNE. PHOTO VIA INATURALIST (CC BY-NC 4.0)





VELVET FOOT (*FLAMMULINA VELUTIPES*) OBSERVED IN KURSK OBLAST, RUSSIA BY RYZHKOV OLEG. PHOTO VIA INATURALIST (CC BY-NC 4.0)

## GLOBAL ANALYSIS OF EVOLUTIONARY HISTORY OF MUSHROOM-FORMING FUNGI

DATA USED: 5,884,445 SPECIES OCCURRENCES

**a** Varga T, Krizsán K, Földi C, Dima B, Sánchez-García M, Sánchez-Ramírez S, Szöllösi GJ, Szarkándi JG, Papp V, Albert L, Andreopoulos W, Angelini C, Antonín V, Barry KW, Bougher NL, Buchanan P, Buyck B, Bense V, Catcheside P, Chovatia M, Cooper J, Dámon W, Desjardin D, Finy P, Geml J, Haridas S, Hughes K, Justo A, Karasiński D, Kautmanova I, Kiss B, Kocsubé S, Kotiranta H, LaButti KM, Lechner BE, Liimatainen K, Lipzen A, Lukács Z, Mihaltcheva S, Morgado LN, Niskanen T, Noordeloos ME, Ohm RA, Ortiz-Santana B, Ovrebo C, Rácz N, Riley R, Savchenko A, Shiryaev A, Soop K, Spirin V, Szebenyi C, Tomšovský M, Tulloss RE, Uehling J, Grigoriev IV, Vágvölgyi C, Papp T, Martin FM, Miettinen O, Hibbett DS and Nagy LG (2019) Megaphylogeny resolves global patterns of mushroom evolution. *Nature Ecology & Evolution*. Springer Science and Business Media LLC 3(4): 668–678. Available at:

<https://doi.org/10.1038/s41559-019-0834-1>

Author countries/areas: Argentina, Australia, Austria, Canada, Czechia, Dominican Republic, Estonia, Finland, France, Hungary, Italy, Netherlands, New Zealand, Norway, Poland, Russia, Slovakia, Sweden, USA

Playing a variety of roles in ecosystems, mushroom-forming fungi (Agaricomycetes) are the most diverse and complex group within the fungal kingdom, with more than 20,000 described species.

In this paper, researchers from institutions in 19 countries present a new, comprehensive megaphylogeny of mushroom-forming fungi, comprising 5,284 species of which more than 1,200 species were specifically sampled as part of the study.

The resulting phylogeny revealed a major class-wide species radiation in the Jurassic, coinciding with angiosperm expansions. Several clade-specific adaptive radiations also took place, offering an explanation for mushroom genera with thousands of species.

Using GBIF-mediated mushroom occurrences, the authors found the highest speciation rates in the temperate zone, contrasting with the latitudinal diversity gradient observed in plants and animals, for which diversity tends to be higher in the tropics.

The classic toadstool morphology is associated with increased rates of lineage diversification and thus appears to have been a key innovation in the evolution of mushroom-forming fungi, shaping the extant diversity of the class.

## MAPPING ENVIRONMENTAL PROPERTIES IN DATA-POOR AREAS USING SPECIES OCCURRENCES

DATA USED: 30,000 SPECIES OCCURRENCES

Zuquim G, Stropp J, Moulatlet GM, Van doninck Jasper, Quesada CA, Figueiredo FOG, Costa FRC, Ruokolainen K and Tuomisto H (2019) Making the most of scarce data: Mapping soil gradients in data-poor areas using species occurrence records. *Methods in Ecology and Evolution*. Wiley 10(6): 788–801. Available at: <https://doi.org/10.1111/2041-210X.13178>

Author countries/areas: Brazil, Ecuador, Finland

Environmental data is commonly used to define ecological niches and to model species distributions. When environmental data is scarce, however, the presence of indicator species can be used to infer environmental conditions.

In this study, authors outline a method for generating environmental maps derived from plot data in three layers: 1) indicator species with known environmental tolerances, 2) species occurrences from e.g. GBIF, and 3) environmental data only.

From the first layer, the authors are able to derive an environmental optimum. In the next step, this can be used to infer estimates of environmental properties at species-only plots. These estimates, when interpolated with known environmental plots, generate a map of an area of interest that can be validated using an external dataset.

The authors employ the proposed framework to successfully produce a map of soil quality (as measured through the concentration of positively charged ions) for Amazonia using ferns and lycophytes as indicator species, leading to a 12-fold increase in environmental data.

## EXPLORING THE GLOBAL DISTRIBUTION OF MYCOHETEROTROPHIC PLANTS

DATA USED: 32,752 SPECIES OCCURRENCES

**a** Gomes SIF, van Bodegom PM, Merckx VSFT and Soudzilovskaia NA (2019) Global distribution patterns of mycoheterotrophy. *Global Ecology and Biogeography*. Wiley. Available at: <https://doi.org/10.1111/geb.12920>

**Author countries/areas:** Netherlands

**Research funding:** Nederlandse Organisatie voor Wetenschappelijk Onderzoek

In mycorrhizal symbioses, plants exchange photosynthesized carbohydrates for mineral nutrients from fungi in the soil. Mycoheterotrophic plants, however, cheat—and obtain carbon from their fungal partners as well, sometimes leading to the complete loss of photosynthesis.

In this study, Dutch researchers explore the environmental preferences of mycoheterotrophic plants and drivers of their global distribution. Using GBIF-mediated occurrences of all known mycoheterotrophic species, the authors compile a large dataset with information on soil, climate, type of mycorrhizal fungi and autotrophic hosts.

Their analysis shows that mycoheterotrophic plants tend to avoid cold and highly seasonal climates. Plants associated with so-called arbuscular fungi are found in broadleaved tropical forests, whereas those preferring the company of ectomycorrhizal fungi occur in temperate regions—primarily needle-leaved forests.

This pattern appears to be driven by climatic factors—primarily temperature and precipitation—rather than abundance of specific autotrophic hosts from which the fungi-mediated nutrients are derived.

## ESTIMATING RANGE EXTENTS OF MUSHROOM-FORMING FUNGI USING OTUs

DATA USED: 155,363 SPECIES OCCURRENCES

Bazzicalupo AL, Whitton J and Berbee ML (2019) Over the hills, but how far away? Estimates of mushroom geographic range extents. *Journal of Biogeography*. Wiley. Available at: <https://doi.org/10.1111/jbi.13617>

**Author countries/areas:** Canada

**Research funding:** Natural Sciences and Engineering Research Council of Canada

Genetically more closely related to animals than to plants, fungi constitute the third kingdom of eukaryotic organisms. Essential to terrestrial ecosystems, they facilitate wood decay and form mycorrhizal symbioses with autotrophic plants.

Accurate knowledge on fungus distributional ranges, however, may be hampered by overestimations due to broad species delimitations.

Using genetically defined operational taxonomic units (OTUs) from the UNITE database, this study by Canadian researchers attempts to improve this knowledge by analysing global data from 12 mushroom-forming fungus genera. They augmented this dataset by sampling 600+ mushroom specimens found in Washington and British Columbia.

As many OTU records lack precise locality data, the authors used country centroids to estimate ranges—a methodology they verified by comparing estimated range extents with GBIF-mediated occurrences of species of four well-studied tree genera.

The overall results put the median range extent of mushroom OTUs at 1,200 to 4,039 km and suggest a biogeographic structure as opposed to global distributions—of mushroom-forming fungi.

## COLD-ADAPTED PLANT RICHNESS IN THE NORTHERN HEMISPHERE

DATA USED: 140,229,903 SPECIES OCCURRENCES

Hagen O, Vaterlaus L, Albouy C, Brown A, Leugger F, Onstein RE, Santana CN, Scotese CR and Pellissier L (2019) Mountain building, climate cooling and the richness of cold-adapted plants in the Northern Hemisphere. *Journal of Biogeography*. Wiley 46(8): 1792–1807. Available at: <https://doi.org/10.1111/jbi.13653>

**Author countries/areas:** France, Switzerland, Germany, USA  
**Research funding:** Deutsche Forschungsgemeinschaft

Plants in cold climates have evolved to have lower stature, tougher leaves, and biochemical properties preventing the formation of ice crystals. Even though lineages of such cold-adapted vascular plants from mid-latitude mountain regions—such as the Himalayas and Rocky Mountains—are also found in Arctic regions, species richness is lower in the Arctic.

This study used GBIF-mediated occurrences of more than 5,000 species occupying cold climates to generate stacked distribution maps confirming this observation. The authors used a model to derive environmental predictors of cold-adapted richness patterns, showing that richness is strongly associated with ecoregions.

Mapping the cold climate of the entire Northern Hemisphere from 60 million years ago to now, the authors showed that while cold regions—with temperatures below 0°C—started appearing in mid-latitude regions with the Himalayan uplift about 42



million years ago, cooling in the Arctic did not start until about 10 million years later, and the region as a whole was not comparably cold until much later.

The overall results indicate that early cooling of alpine regions allowed for evolution and diversification of cold-adapted vascular plants, followed much later by colonization of the Arctic.

### HIGHER RISK OF LOCAL EXTINCTION FOR SPECIES WITH SMALL RANGES

DATA USED: 247,231,174 SPECIES OCCURRENCES

**a** Staude IR, Navarro LM and Pereira HM (2019) Range size predicts the risk of local extinction from habitat loss. *Global Ecology and Biogeography*. Wiley 29(1): 16–25. Available at: <https://doi.org/10.1111/geb.13003>

**Author countries/areas:** Germany

**Research funding:** Deutsche Forschungsgemeinschaft

At the global scale, species with small ranges are more likely to face extinction when threatened by changes in climate or land use. It is not well-documented, however, whether this risk also applies at the local level.

This study by researchers from Germany used historical land-use data combined with GBIF-mediated plant occurrences to derive species range based on the number of grid cells occupied by a given species. They then identified cells that had experienced natural habitat loss and the species present before and after the change.



*GENTIANA DEPRESSA* OBSERVED IN LANGTANG, NEPAL BY YALING LIN. PHOTO VIA INATURALIST (CC BY-NC 4.0)

Grouping species into small or large ranges, the author calculated the probability of a group persisting at a given amount of habitat loss, showing that small-range species have a steeper decline than large-range plants.

At 80 per cent habitat loss, for example, the probability of persisting for species with larger ranges is higher than 85 per cent, while small-range species have a significantly lower persistence probability of 75 per cent.

### DISCRIMINATION ACCURACY UNINFORMATIVE WHEN EVALUATING SPECIES DISTRIBUTION MODELS

DATA USED: 5,969,252 SPECIES OCCURRENCES

**a** Warren DL, Matzke NJ and Iglesias TL (2019) Evaluating presence-only species distribution models with discrimination accuracy is uninformative for many applications. *Journal of Biogeography*. Wiley 47(1): 167–180. Available at: <https://doi.org/10.1111/jbi.13705>

**Author countries/areas:** Australia, Germany, Japan, New Zealand

**Research funding:** Macquarie University, Australian Research Council

Species distribution models (SDMs) are popular tools—easy to build using data freely available—and, in some cases the only tractable means of estimating habitat suitability. The ability to predict withheld data (discrimination accuracy) is often used to make decisions on models, methods and data.

Using a simulation approach, this study evaluates the relationship between discrimination accuracy and functional accuracy to assess commonly-used methods for selection of models. The authors simulated occurrence data across 11 levels of sampling bias—modelled using GBIF-mediated data of all plants in Australia—performing 20 simulations for each and using four experimental scenarios to build models based on seven algorithms, resulting in a total of 6,160 models.

The results showed a fairly good functional accuracy of models overall—that is, modelled suitability matched true suitability of habitat. However, across nearly all algorithms and levels of complexity the discrimination capacity was a very poor predictor of functional accuracy, indicating that assessing models based on this approach might be less than ideal.

## DESCRIBING BROAD-SCALE DIVERSITY PATTERNS USING HERBARIUM DATA

DATA USED: 4,251,424 SPECIES OCCURRENCES

Bottin M, Peyre G, Vargas C, Raz L, Richardson JE and Sanchez A (2019) Phytosociological data and herbarium collections show congruent large-scale patterns but differ in their local descriptions of community composition. *Journal of Vegetation Science*. Wiley 31(1): 208–219. Available at: <https://doi.org/10.1111/jvs.12825>

Author countries/areas: Colombia, UK

The study of vegetation turnover, composition gradients and beta-diversity requires systematically collected sampling data from vegetation plots and relevés. Occurrence data from museum and herbarium specimens, however, may be used in cases where such data is scarce.

This study compares phytosociological vegetation data with mainly GBIF-mediated herbarium data from páramos (alpine ecosystems) in Colombia in their ability to describe diversity patterns and species composition gradients.

The authors show that overall species composition differed between the two datasets—perhaps explained by a bias towards charismatic and/or rare species in the herbarium data. So while a large

number of species were present in a few locations in the herbarium data—and absent from the plot data—many species present in the plot data were common to a larger number of plots than described by the herbarium data.

Pairwise beta-diversity analyses showed no significant differences between the datasets, and among common species, ecological distributions patterns were similar. The authors conclude that while plot data is best for describing local patterns, combining data from the two sources may be efficient at broader scales.

## ECO-GEOGRAPHICAL ANALYSIS OF RARE ENDEMIC PLANTS OF SIBERIA

Baikov KS, Baikova EV and Banaev EV (2019) Ecological and Geographical Analysis of Endemic Vascular Plants in Siberia and Problems of Their Conservation Ex Situ. *Contemporary Problems of Ecology*. Pleiades Publishing Ltd 12(5): 418–433. Available at: <https://doi.org/10.1134/S1995425519050032>

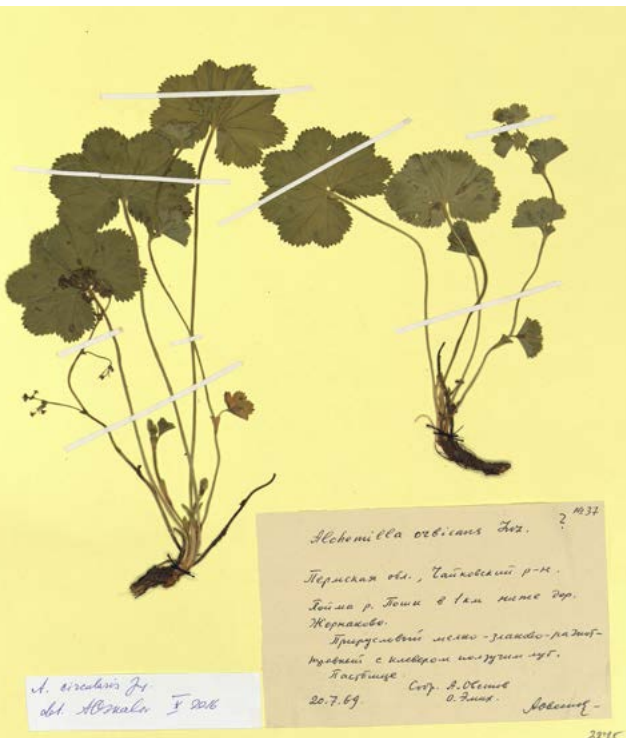
Author countries/areas: Russia

Endemic species are only found in a single geographical location, zone or habitat type. Some very rare species are only known from the first location in which they were originally described (stenoendemics).

In this study, researchers from Russia developed a custom nine-point ranking system based on characteristics such as stenoendemism, typification and inclusion in red lists to identify the 50 rarest vascular plant species in Siberia. Using herbarium data published through GBIF, the authors analysed the taxonomic composition and eco-geographical features of these species in order to develop scientific recommendations for their conservation.

The 50 identified species make up about 1.1 per cent of the total number of known vascular plants in Siberia. The most widely represented families are Asteraceae (nine species), Rosaceae (seven species), and Ranunculaceae (four species). More than 80 per cent of the studied species are found in the southern hemi-boreal zone (areas of mixed boreal and temperate forests), and nearly half belong to the Altai-Siberian ecoregion.

The study finds a low level of state protection of the studied species and recommends inclusion of endemic Siberian species into lists of rare and endangered plants of Russia.



ALCHEMILLA CIRCULARIS RECORDED IN CHAYKOVSKIY RAYON, PERM KRAI, RUSSIA BY OVESNOV A. PHOTO VIA MOSCOW STATE UNIVERSITY (CC BY 4.0)

## HELP PROMOTE OPEN BIODIVERSITY DATA



SAULCY'S LEAFCUTTER (*MEGACHILE SAULCYI*) OBSERVED IN SANTIAGO, CHILE BY ORLANDOMONTES. PHOTO VIA INATURALIST (CC BY-NC 4.0)

- ✓ You can provide at least one example in which you have shared biodiversity data through GBIF, used GBIF-mediated data, and/or advocated open data in your professional capacity
- ✓ You agree with the ICSU-World Data System Data Sharing Principles—in short, that data should be shared openly in a timely manner, with the fewest restrictions possible and used with proper citation
- ✓ You agree to share your contact information on GBIF.org, and possibly on websites run by GBIF nodes and partners
- ✓ You allow the GBIF Secretariat and GBIF nodes to contact you about opportunities to promote open biodiversity data at specific events
- ✓ You commit to providing details annually about at least one event, publication or process in which you have advocated for open biodiversity data

If you can check all of the above, then maybe you're the next **Biodiversity Open Data Ambassador!**

For more information about the programme and how to apply, please visit

<http://bit.ly/bodamba>

## Environmental Sciences



POLAR BEARS (*URSUS MARITIMUS*) BY SANDRA EGLITE VIA INATURALIST. PHOTO LICENSED UNDER CC BY-NC 4.0.

### AMAZONIA FACING THREATS OF DEFORESTATION AND CLIMATE CHANGE

DATA USED: OCCURRENCES OF 6,394 SPECIES

Gomes VHF, Vieira ICG, Salomão RP and ter Steege H (2019) Amazonian tree species threatened by deforestation and climate change. *Nature Climate Change*. Springer Science and Business Media LLC 9(7): 547–553. Available at: <https://doi.org/10.1038/s41558-019-0500-2>

Author countries/areas: Brazil, Netherlands

Holding close to 13 per cent of all trees of the world, the Amazonian lowlands is the largest rainforest block on Earth with a richness of up to 16,000 species. This immense tree diversity is threatened by both deforestation and climate change, and assessing the relative impacts of these threats is crucial to conservation.

In this study, researchers used environmental suitability models based on GBIF-mediated occurrences of nearly 7,000 Amazonian trees to estimate impacts of climate change. They also used historical deforestation data to predict future losses.

The authors found that by 2050, deforestation alone may cause a decline in richness of 19-36 per cent, while climate change may account for 31-37 per cent. Combined, however, the effect may reach 58 per cent

loss of richness—with the worst outlook in the portion of the Amazonian forest outside current protected areas, accounting for roughly half of the total.

With the major importance of the rainforest in absorbing atmospheric CO<sub>2</sub> and regulating climate, this study urges expanded protected area networks and an urgent halt to deforestation, to help mitigate climate change and avoid species collapse in Amazonia.

### FRESHWATER MANAGEMENT USING MULTI-OBJECTIVE OPTIMIZATION OF SPECIES DISTRIBUTION MODELS

DATA USED: 65,000 SPECIES OCCURRENCES

GGobeyn S and Goethals PLM (2019) Multi-objective optimisation of species distribution models for river management. *Water Research*. Elsevier BV 163: 114863. Available at:

<https://doi.org/10.1016/j.watres.2019.114863>

Author countries/areas: Belgium

Modelling freshwater species distributions can be challenging, as biases in the case of species that are obscure or difficult to observe may lead to underestimating the true habitat suitability. In this paper, authors present a solution to this problem involving optimizing models based on multiple



objectives rather than just one.

Using data from the GBIF-mediated Limnadata Neerlandica dataset, the authors evaluated the performance of their multi-objective optimization (MOO) approach—called the non-dominated sorting genetic algorithm II—against regular single objective models of 11 pollution-sensitive macroinvertebrate species (including *Cloeon dipterum* shown above).

Their results showed that the MOO approach is two-to-four times more efficient at identifying large range distributions, while only requiring four per cent longer runtimes per training. To support decision-making, the authors propose a closer collaboration between model developers and freshwater managers to set environmental standard limits on a more objective basis.

### THE IMPLICATIONS OF THE PARIS AGREEMENT ON CLIMATE CHANGE FOR GLOBAL BIODIVERSITY HOTSPOTS

DATA USED: OCCURRENCES OF 80,000 SPECIES

**a** Warren R, Price J, VanDerWal J, Cornelius S and Sohl H (2018) The implications of the United Nations Paris Agreement on climate change for globally significant biodiversity areas. *Climatic Change*. Springer Science and Business Media LLC 147(3-4): 395–409. Available at: <https://doi.org/10.1007/s10584-018-2158-6>

Author countries/areas: Australia, UK

Research funding: Natural Environment Research Council



CARRION CROW BUSH (*SENNA RETICULATA*) OBSERVED IN MANAUS, BRAZIL BY CRAX. PHOTO VIA INATURALIST (CC BY-NC 4.0)

Under the Paris Agreement of the UN Framework Convention on Climate Change, countries have pledged to reduce greenhouse gas emissions in order to limit global warming. Current data, however, shows that contributions are falling short of the required levels towards limiting warming to even 2°C above pre-industrial levels.

In this study, researchers modelled the distributions of ~80,000 plants and animals using GBIF-mediated data to assess the risks to species in globally significant biodiversity conservation areas, quantifying the benefits of national emission reductions and the effects of limiting warming to 2°C against 4.5°C in the unmitigated scenario.

The results show that without mitigation, on average only 33 per cent of each conservation area will be available as a climate refugium. If warming is limited to 2°C, however, the average refugium area is doubled. The authors conclude that limiting warming to well below 2°C would further reduce risk of local extinctions.

### CROP SECURITY AND ADAPTIVE STRATEGIES IN SUB-SAHARAN AFRICA UNDER CLIMATE CHANGE

DATA USED: OCCURRENCES OF 807 SPECIES

Pironon S, Etherington TR, Borrell JS, Kühn N, Macias-Fauria M, Ondo I, Tovar C, Wilkin P and Willis KJ (2019) Potential adaptive strategies for 29 sub-Saharan crops under future climate change. *Nature Climate Change*. Springer Science and Business Media LLC 9(10): 758–763. Available at: <https://doi.org/10.1038/s41558-019-0585-7>

Author countries/areas: New Zealand, UK

With climate change expected to negatively impact agriculture, devising adaptive strategies for crops is an important step in overcoming this global challenge.

In this study, researchers developed an analytical framework for assessing crop security and mitigating strategies for agriculture in sub-Saharan Africa, considered especially at risk due to its tropical location, socioeconomic policy and farming characteristics.

Based on occurrence data from GBIF and other sources, the authors modelled the climatic niches of 29 major crops and 778 of their wild relatives—now and in the future—to determine crop security.

For each crop they then assessed the potential success of three strategies: switching to cultivars from other continents, using wild relatives, or—in the worst cases—replacing with another sub-Saharan African crop.



## Evolutionary Biology



*MACRACANTHA ARCUATA* OBSERVED IN GUNUNG MULU NATIONAL PARK, MALAYSIA. PHOTO 2019 KAL SQUIRES VIA INATURALIST UNDER (CC BY-NC 4.0)

### EFFECTS OF SAMPLING BIAS ON STABILITY OF AREAS OF ENDEMISM

**DATA USED:** 34,000 SPECIES OCCURRENCES

Casagrande MD and Goloboff PA (2019) On stability measures and effects of data structure in the recognition of areas of endemism. *Biological Journal of the Linnean Society*. Oxford University Press [OUP] 127(1): 143–155. Available at: <https://doi.org/10.1093/biolinnean/blz019>

**Author countries/areas:** Argentina

Biases in data—whether geographic or taxonomic—all introduce uncertainty in downstream analyses which can affect results and misrepresent the real world.

This study attempts to quantify the effects of different types of biases by introducing two stability measures—one for geography and one for taxonomy—to indicate the degree to which a biased dataset agrees with its unbiased version.

Using three datasets, including a GBIF download of all Amazonian amphibians, researchers artificially introduced "biases" by randomly removing increasing fractions of records for 1) all species, 2) specific subsets of species, and 3) within defined geographical sectors—to emulate problems of poor sampling, uneven sampling and geographical bias, respectively.

In analyses focused on specific areas of endemism, the authors compared results from the original data with intentionally biased data to uncover measures of stability—that is, the degree to which the biased data leads to the same predictions as the unbiased data.

In nearly all cases, stability diminished with increased removal of data. However, it seems that data incompleteness had a bigger impact on taxonomic stability, indicating perhaps on a positive note that even scattered data can lead to fair spatial identification of areas of endemism.

### UNTANGLING THE PHYLOGENETIC WEBS OF ORB-WEAVING SPIDERS

**DATA USED:** 66,479 SPECIES OCCURRENCES

Scharff N, Coddington JA, Blackledge TA, Agnarsson I, Framenau VW, Szűts T, Hayashi CY and Dimitrov D (2019) Phylogeny of the orb-weaving spider family Araneidae [Araneae: Araneoidea]. *Cladistics*. Wiley 36(1): 1–21. Available at: <https://doi.org/10.1111/cla.12382>

**Author countries/areas:** Australia, Denmark, Hungary, Norway, USA

The third-most speciose family of spiders, Araneidae (the orb-weaving spiders) has been widely studied from within various branches of research including behavioral science, ecology, material science, genomics and pharmacology. Modern phylogenetic revisions of the family, however, have been geographically limited.

Continuing previous work on Araneid evolution and diversification, the authors of this study present a new phylogeny, based on genetic samples from 158 taxa. Using GBIF-mediated occurrences they generate a map of the global distribution of Araneid species and add many understudied species from the southern hemisphere to get a true global representation of

the family. Surprisingly, many clades named in the new phylogeny cannot be corroborated by current knowledge from morphological or behavioral studies.

Based on the phylogeny, the authors also analysed the evolution of sexual size dimorphism and web architecture. While genera exhibiting the former did not form a distinct clade, several patterns were discovered for web architecture including a monophyletic group of spanning thread web spiders.

### EXPLORING THE MECHANISMS UNDERLYING CONIFER HOTSPOT FORMATION

DATA USED: 324,485 SPECIES OCCURRENCES

Sundaram M, Donoghue MJ, Farjon A, Filer D, Mathews S, Jetz W and Leslie AB (2019) Accumulation over evolutionary time as a major cause of biodiversity hotspots in conifers. *Proceedings of the Royal Society B: Biological Sciences. The Royal Society* 286(1912): 20191887.

Available at: <https://doi.org/10.1098/rspb.2019.1887>

Author countries/areas: USA

Biodiversity is not distributed evenly across the planet, but follows patterns such as the latitudinal species gradient or assemblages in regional hotspots. Understanding these differences in spatial distribution and what causes them is a crucial aspect of evolutionary ecology.

This paper uses the global distribution of conifers to explore mechanisms underlying biodiversity hotspot formation. Its authors derived species ranges and—subsequently—assemblages from GBIF-mediated occurrence data to identify eight richness hotspots, while analysing a wide range of characteristics to compare these with non-hotspots.

In their analysis, they find greater topographic heterogeneity in hotspots than surrounding areas, although not consistently associated with differences in climate or soil types. Comparing traits, the authors show minor, but non-consistent differences between hotspot and non-hotspot species.

Overall, the study suggests a bigger role played by geography than biology, and that hotspots form not as a result of unique diversification processes, but rather a simple accumulation of regional diversity in stable, mountainous areas.

### SEX BIASES IN NATURAL HISTORY COLLECTIONS

DATA USED: 6.6 M SPECIES OCCURRENCES

Cooper N, Bond AL, Davis JL, Portela Miguez R, Tomsett L and Helgen KM (2019) Sex biases in bird and mammal natural history collections. *Proceedings of the Royal Society B: Biological Sciences. The Royal Society* 286(1913): 20192025. Available at:

<https://doi.org/10.1098/rspb.2019.2025>

Author countries/areas: Australia, UK

Often treated as a nuisance variable—or overlooked altogether—the biological sex of an organism is an important factor affecting many aspects of the individual's ecology and behaviour. Among museum collection specimens, biased sex ratios could have serious implications for studies that rely on them.

This study examines GBIF-mediated records of more than 2.4 million bird and mammal specimens from five major international museums. For birds, 20 per cent were female, 31 per cent male and 49 per cent unsexed, while for mammals—likely easier to sex—only 15 per cent of specimens were unsexed, 44 per cent were male and 41 per cent were female.

This bias was not due to unsexed specimens being female. Even species that exhibit female-skewed ratios in the wild, showed male bias among specimens.

The authors found no significant improvement in sex balance over the last 130 years.

The most notable finding of the study was among the name-bearing type specimens (holotypes, syntypes, lectotypes, and neotypes), where only 25 per cent of birds and 39 per cent of mammals were females.



FEMALE SPECIMEN OF A NORTHERN YELLOW-SHAFTED FLICKER (*COLAPTES AURATUS* SUBSP. *LUTEUS*) COLLECTED IN EVANSTON, ILLINOIS, USA. PHOTO VIA THE FIELD MUSEUM (CC-BY-NC 4.0)



## DISTURBANCE SHAPING THE ELEVATION RANGES OF TREES IN COSTA RICA

DATA USED: 126,066 SPECIES OCCURRENCES

**a** Muñoz Mazón M, Klanderud K, Finegan B, Veintimilla D, Bermeo D, Murrieta E, Delgado D and Sheil D (2019) Disturbance and the elevation ranges of woody plant species in the mountains of Costa Rica. *Ecology and Evolution*. Wiley 9(24): 14330–14340. Available at: <https://doi.org/10.1002/ece3.5870>

**Author countries/areas:** Costa Rica, Germany, Norway  
**Research funding:** Norges Miljø- og Biovitenskapelige Universitet

In ecology, disturbance caused by transient reduction in competition resulting from vegetation death or removal—through e.g. grazing, mowing or controlled burning—can help promote species diversity at local scales, but how it might shape distributions at larger scales is unknown.

In this study, researchers examined and compared the elevation range limits of tree species in undisturbed old-growth and secondary forests of Costa Rica. Using plot data and GBIF-mediated occurrences of 7,750 individual tree stems representing more than 490 species, the authors calculated elevation ranges of all species and analysed relationships with other parameters.

In their results, species ranges increased with elevation. Consistent with the authors' predictions, species in secondary growth forests had wider elevation ranges than species in old growth—a pattern also seen in larger stems compared to smaller stems. Despite some noise and uncertainty, these results together add evidence to the role of disturbance in expanding tree elevation ranges.



*QUERCUS COSTARICENSIS* OBSERVED IN COSTA RICA BY BOTANICALKATZ. PHOTO VIA INATURALIST (CC BY-NC 4.0)

## Multidisciplinary Sciences



HARLEQUIN RACERUNNER (*PLICA UMBRA*) OBSERVED IN BORBA, AMAZONAS, BRAZIL BY JONATHAN NEWMAN. PHOTO VIA INATURALIST (CC BY-NC 4.0)

### REDUCING EFFECTS OF SAMPLING GAPS IN CONSERVATION MAPPING

DATA USED: 659,062 SPECIES OCCURRENCES

Oliveria U, Soares-Filho BS, Santos AJ, Paglia AP, Brescovit AD, de Carvalho CJB, Silva DP, Rezende DT, Leite FSF, Batista JAN, Barbosa JPPP, Stehmann JR, Ascher JS, Vasconcelos MF, Marco PD, Löwenberg-Neto P and Ferro VG (2019) Modelling Highly Biodiverse Areas in Brazil. *Scientific Reports*. Springer Science and Business Media LLC 9(1). Available at:

<https://doi.org/10.1038/s41598-019-42881-9>

Author countries/areas: Argentina, Brazil, Singapore

In megadiverse regions and countries, gaps in biological sampling can make traditional mapping approaches for conservation difficult. Strategies developed for well-sampled regions may produce unsatisfactory results when applied in countries with less even sampling—such as Brazil.

In this study, authors developed a new comprehensive spatial framework for mapping highly biodiverse areas in Brazil using species occurrences as mediated by GBIF and applying special techniques to reduce the effect of irregular sampling. From thoroughly cleaned species records and a phylogenetic supertree, the authors derive a range of quantitative biodiversity variables, including composition, richness and endemism—both at the species and phylogenetic level.

Summarizing the quantitative variables at a regional scale, the model identifies the smallest possible

areas with the most unique biodiversity, producing a map of relevant priority areas for biodiversity conservation. In the optimal solution, the model is able to encompass ~90 per cent of known biodiversity in only 10 per cent of the country's area.

### PALEOCLIMATIC ANALYSIS OF EARLY EOCENE TRANSIENT GLOBAL WARMING EVENTS

DATA USED: 6,282,569 SPECIES OCCURRENCES

Willard DA, Donders TH, Reichgelt T, Greenwood DR, Sangiorgi F, Peterse F, Nierop KGJ, Frieling J, Schouten S and Sluijs A (2019) Arctic vegetation, temperature, and hydrology during Early Eocene transient global warming events. *Global and Planetary Change*. Elsevier BV 178: 139–152. Available at:

<https://doi.org/10.1016/j.gloplacha.2019.04.012>

Author countries/areas: Canada, Netherlands, USA

Global temperatures have been both much higher and much lower than the current average, and understanding paleoclimatic variations is important as they provide valuable analogues for models of future climate change.

This study examines two transient warming events—the Paleocene-Eocene Thermal Maximum (PETM) and the Eocene Thermal Maximum (ETM2)—in which the temperature increased by 4–8°C for periods of less than 200,000 years each.

Applying pollen and spore microfossil analyses of marine sediment cores from nearly 400 metres below

(continued on next page...)



the Arctic Ocean seabed, researchers reconstructed the terrestrial vegetation of PETM and ETM2.

Using a bioclimatic analysis of GBIF-mediated occurrences of modern living relatives assumed to have similar climatic niches, the authors then derived climate parameters during the two events.

Based on pollen taxa present and bioclimatic reconstructions, the study provides evidence of Arctic broad-leaved swamp forests during both PETM and ETM2. PETM mean annual temperatures and mean winter temperatures were both ~2 to 3.5°C warmer than the late Paleocene.

### PRIORITIZING IN SITU CONSERVATION OF CROP WILD RELATIVES

DATA USED: 136,576 SPECIES OCCURRENCES

**a** Vincent H, Amri A, Castañeda-Álvarez NP, Dempewolf H, Dulloo E, Guarino L, Hole D, Mba C, Toledo A and Maxted N (2019) Modeling of crop wild relative species identifies areas globally for in situ conservation. *Communications Biology*. Springer Science and Business Media LLC 2(1). Available at: <https://doi.org/10.1038/s42003-019-0372-z>

Author countries/areas: Colombia, Germany, Italy, Macao, Morocco, UK, USA

Research funding: Bioforsk, University of Birmingham

The wild cousins of agricultural plants, crop wild relatives (CWRs) are important sources of genetic diversity with potential for breeding more nutritious varieties with higher yields and climate change tolerance.

Exploring relevant areas for in situ conservation of CWRs, this study applies species distribution modelling to 1,261 wild species from 167 major crop gene pools using a global CWR database mediated by GBIF. The models reveal the highest concentrations of CWRs in the Mediterranean basin with a hotspot on the northeast Lebanese/Syrian border.

In future climates, the crop types with the most CWRs predicted to lose more than 50 per cent of their distribution potential are root, bulb or tuberous vegetables, but cereals and legumes are also likely to face substantial losses. Citrus is the least affected crop.

While prioritizing areas for conservation, the authors find that protecting the top 150 sites would only involve 0.01 per cent of the total terrestrial area of the world. The conservation of just the top 10 sites inside and outside existing protected areas, respectively, would protect 475 CWR species and more than 1,200 CWR/crop combinations.

### MODELLING THE GLOBAL DIVERSITY OF PHYTOPLANKTON

DATA USED: 790,000 SPECIES OCCURRENCES

**a** Righetti D, Vogt M, Gruber N, Psomas A and Zimmermann NE (2019) Global pattern of phytoplankton diversity driven by temperature and environmental variability. *Science Advances*. American Association for the Advancement of Science (AAAS) 5(5): eaau6253.

Available at: <https://doi.org/10.1126/sciadv.aau6253>

Author countries/areas: Switzerland

Marine phytoplankton play a crucial role in ocean productivity, yet very little is known about the global patterns of diversity of these extremely diverse marine organisms, varying from photosynthesizing cyanobacteria to plant-like diatoms.

Combining more than a million occurrence records from GBIF, the Ocean Biodiversity Information System (OBIS), and others with 10 environmental variables representing key dimensions of phytoplankton ecological niches, authors of this paper modeled the distributions of more than 500 species and globally projected the models onto monthly data fields.

Based on these projections, the authors produced a global phytoplankton map, showing threefold richness in the tropics compared to higher latitudes, confirming the metabolic theory, with sea surface temperature as the most important predictor.

In cooler waters, however, richness is lower than expected, with regions between 35 and 60 degrees latitude showing the greatest divergence. The authors suggest that environmental variability may play a role in reducing richness—either directly or through enhancing competitive exclusion.



VIRGINIA STRAWBERRY (*FRAGARIA VIRGINIANA*) OBSERVED IN AMHERST, MA, USA BY AARON HULSEY. PHOTO VIA INATURALIST (CC BY-NC 4.0)

## UNDERSTANDING MACROEVOLUTIONARY PROCESSES IN TEMPERATE BIOTAS

DATA USED: 1,650,109 SPECIES OCCURRENCES

**a** Folk RA, Stubbs RL, Mort ME, Cellinese N, Allen JM, Soltis PS, Soltis DE and Guralnick RP (2019) Rates of niche and phenotype evolution lag behind diversification in a temperate radiation. *Proceedings of the National Academy of Sciences*. *Proceedings of the National Academy of Sciences* 116(22): 10874–10882.

Available at: <https://doi.org/10.1073/pnas.1817999116>

Author countries/areas: Switzerland, USA

Evolution takes place all the time. When shifts in environmental regimes provide ecological opportunities, diversification can happen fast. New ecological niches and phenotypic traits evolve, but what might explain shifts in rates—and is there an association between the timing of diversification of species, niche and traits?

Focusing on a major flowering plant clade—Saxifragales—a group of nearly 2,500 species encompassing trees such as sweet gum and shrubs such as gooseberry, authors of this study built a phylogenetic tree of 1,736 species and addressed macroevolutionary questions through analyses of traits and ecological niches, derived from 1.6 million GBIF-mediated occurrences and climate data.

Their results showed that in Saxifragales, species first diversified 15 million years ago, driven by the Earth's cooling climate. It wasn't, however, until about five million years later that these plants invaded new habitats and evolved new physical traits. This suggests an evolutionary pattern never observed before, involving a strong temporal lag with diversification preceding niche and phenotype evolution in temperate biotas.

## CLIMATE-INDUCED CO-EXTINCTIONS IN SEVEN EUROPEAN POLLINATOR NETWORKS

DATA USED: OCCURRENCES OF 59 SPECIES

**a** Bascompte J, García MB, Ortega R, Rezende EL and Pironon S (2019) Mutualistic interactions reshuffle the effects of climate change on plants across the tree of life. *Science Advances*. *American Association for the Advancement of Science (AAAS)* 5(5): eaav2539.

Available at: <https://doi.org/10.1126/sciadv.aav2539>

Author countries/areas: Spain, Switzerland, UK

Research funding: Ministerio de Ciencia y Tecnología

Species distribution models are commonly used for predicting the effects of climate change including potential extinctions.

Such models treat species independently and neglect interactions between species.

Based on a rich trait dataset of seven European pollinator networks, this study explores potential climate-induced extinctions while considering mutual dependencies and thus co-extinctions. Adding GBIF-mediated occurrences, authors progressively projected models of 244 plant species into future climates, assigning a co-extinction probability each time an interaction would be lost.

The results show substantial variability across networks with dramatic increases in co-extinction rates in the Mediterranean networks—compared to the other networks. In fact, the 2080 projection shows a near-doubling of climate-induced extinctions when potential co-extinctions are considered.

Not surprisingly, some trait profiles are less likely to be driven to climatic extinction than others, however, the same profile may indeed have a significantly larger risk of disappearing through subsequent co-extinctions.

## REPLACING COFFEE WITH COCOA TO OVERCOME EFFECTS OF CLIMATE CHANGE

DATA USED: 48,239 SPECIES OCCURRENCES

**a** De Sousa K, van Zonneveld M, Holmgren M, Kindt R and Ordoñez JC (2019) The future of coffee and cocoa agroforestry in a warmer Mesoamerica. *Scientific Reports*. *Springer Science and Business Media LLC* 9(1). Available at: <https://doi.org/10.1038/s41598-019-45491-7>

Author countries/areas: Costa Rica, Ecuador, Kenya, Netherlands, Norway, Peru, Taiwan

With productivity of *Arabica* coffee projected to decline drastically, replacing coffee with cocoa (*Theobroma cacao*) and integrating trees in combined agroforestry systems may become an important strategy for climate change adaptation in Mesoamerica.

This study investigated the feasibility of this strategy by assessing current and future climatic suitability of coffee, cocoa and 100 common agroforestry trees using primarily GBIF-mediated occurrences.

By modelling distributions of all species and projecting these into 2050 climate scenarios, the authors showed that coffee is indeed more vulnerable to climate change than cocoa, as up to 62 per cent of current areas for production may become unsuitable by 2050. Eighty-five per cent of these vulnerable areas may, however, remain suitable for cocoa.

(continued on next page...)

The results also showed that despite overall losses in suitability for the most popular tree species, agroforestry may remain a viable alternative, as 72 per cent of future coffee areas will be able to support more than 30 tree species. For cocoa, the agroforestry potential is even higher—with 95 per cent of areas suitable for more than 30 species..

### USING MAMMAL EXCREMENT TO EXPLAIN THE RICHNESS OF THE CAPE FLORISTIC REGION

DATA USED: 165,365,849 SPECIES OCCURRENCES

Chase BM, Boom A, Carr AS, Chevalier M, Quick LJ, Verboom GA and Reimer PJ (2019) Extreme hydroclimate response gradients within the western Cape Floristic region of South Africa since the Last Glacial Maximum. *Quaternary Science Reviews*. Elsevier BV 219: 297–307. Available at: <https://doi.org/10.1016/j.quascirev.2019.07.006>

Author countries/areas: France, South Africa, Switzerland, UK

One of the world's biodiversity hotspots, the Cape Floristic Region (CFR) in South Africa is characterized by remarkable vegetation with high levels of endemism and species richness. The floral diversity of the CFR is thought to be driven by relative climatic stability, but it is not homogenous across the whole region.

This study uses GBIF-mediated occurrences of nine classes of vascular and non-vascular plants to map the species richness of the CFR—highlighting the extreme diversity in the southwestern part of the region. Using high-resolution paleoclimatic data derived from stable isotope analysis of hyraceum samples (fossilized faeces and urine of the rock hyrax) up to ~20,000 years old, the authors explore the climate change dynamics in the western CFR.

The results of their analysis point to major transitions in mean climate state including series of abrupt changes in water availability of up 70 per cent—each lasting 500–2,000 years. The authors speculate that these changes may have driven a speciation pump, ultimately responsible for the high diversity in the western CFR today.

### AGE AND PATTERNS OF ENDEMISM ON ISLANDS WORLDWIDE

DATA USED: 160,301,578 SPECIES OCCURRENCES

Veron S, Haevermans T, Govaerts R, Mouchet M and Pellens R (2019) Distribution and relative age of endemism across islands worldwide. *Scientific Reports*. Springer Science and Business Media LLC 9[1]. Available at: <https://doi.org/10.1038/s41598-019-47951-6>

Author countries/areas: France, UK

Isolated from continental life, evolution on islands can take its own course, allowing many unique lineages and high levels of endemism. In some sites, the age of endemic species varies significantly from the average, a phenomenon thought to be associated with the origin of the island. Recent endemism is associated with oceanic islands, i.e. islands formed de novo, whereas ancient endemism is linked to continental islands, i.e. islands formed by fragmentation.

Focusing on monocots, this study sought to identify islands with significant endemic flora of more recent, ancient or mixed lineages, and to investigate factors contributing to these patterns. Using GBIF-mediated plant occurrences representing 15,964 species found on 4,306 islands, the authors calculated a measure of phylogenetic endemism and identified 142 islands as being either recent, ancient or mixed—with a subset of 42 islands being 'super-endemic', an extreme case of the last category.



ROCK HYRAX (*PROCAVIA CAPENSIS* SUBSP. *CAPENSIS*) OBSERVED IN SOUTH AFRICA BY MICHELE BERTONCINI. PHOTO VIA INATURALIST (CC BY-NC 4.0).

The islands identified—irreplaceable for uniqueness and evolutionary history—can help guide biodiversity conservation. While distributed across the entire world, most are present at low latitudes. In addition, the authors found that habitat availability and climate stability were the most important environmental factors contributing to endemism.

### THE ROLE OF WILDERNESS AREAS IN PROTECTING BIODIVERSITY

DATA USED: 65,733,880 SPECIES OCCURRENCES

Di Marco M, Ferrier S, Harwood TD, Hoskins AJ and Watson JEM (2019) Wilderness areas halve the extinction risk of terrestrial biodiversity. *Nature*. Springer Science and Business Media LLC 573(7775): 582–585. Available at: <https://doi.org/10.1038/s41586-019-1567-7>

Author countries/areas: Australia, Italy, USA

Wilderness areas—nature with no or minimal human disturbance—are rapidly decreasing in global extent. How important this last stronghold of intact ecosystems is in mitigating the biodiversity crisis, however, is unknown.

Seeking to quantify the value of wilderness areas, this study applied models of compositional turnover—or beta-diversity—based on more than 65 million GBIF-mediated occurrences of vascular plants and invertebrates to predict differences in composition between wilderness and non-wilderness.

By identifying the areas on every continent that make the highest relative contribution to persistence of biodiversity, the study finds that wilderness buffers extinction risks in all biogeographical realms. The effect is more pronounced, however, in areas with larger remaining extent of wilderness, such as the Palearctic.

Overall, the probability of species extinction in non-wilderness areas is more than twice as high as in wilderness areas, and for some communities the loss of a single square km grid cell of wilderness signifies a reduction in species persistence of up to 14 per cent.

### THE IMPACT OF CLIMATE CHANGE ON ISLANDS

DATA USED: 2,550,396 SPECIES OCCURRENCES

Veron S, Mouchet M, Govaerts R, Haevermans T and Pellens R (2019) Vulnerability to climate change of islands worldwide and its impact on the tree of life. *Scientific Reports*. Springer Science and Business Media LLC 9(1). Available at: <https://doi.org/10.1038/s41598-019-51107-x>

Author countries/areas: France, UK

Having evolved in isolation, islands represent unique communities with extremely high rates of

endemism—exceeding mainland species almost tenfold. This uniqueness, however, also makes insular communities especially vulnerable to rapid changes caused by human activities, such as climate change.

To understand the impact a changing climate may have on islands, this study investigated relative vulnerability of monocots, a major group of flowering plants with many taxa of high economic value (e.g. palms and bananas).

Analysing 1,497 monocot genera distributed across 5,565 islands worldwide, the authors calculated the overall vulnerability of all islands to shifts in sea level, temperature and rainfall, and linked this to the expected phylogenetic diversity, derived from assessing 2.5 million GBIF-mediated occurrences, and the phylogenetic losses from predicted climate-induced local extinctions.

The most vulnerable islands were mainly situated at low latitudes and/or in the Southern Hemisphere, while the highest estimated phylogenetic losses mainly occurred in the Eastern Hemisphere. Considering both factors, the study highlights Cuba and New Guinea as highly vulnerable, both representing large phylogenetic diversity at stake.

### ONE THIRD OF TROPICAL AFRICAN FLORA MAY BE FACING EXTINCTION

Stévant T, Dauby G, Lowry PP, Blach-Overgaard A, Droissart V, Harris DJ, Mackinder BA, Schatz GE, Sonké B, Sosef MSM, Svenning J-C, Wieringa JJ and Couvreur TLP (2019) A third of the tropical African flora is potentially threatened with extinction. *Science Advances*. American Association for the Advancement of Science (AAAS) 5(11): eaax9444. Available at:

<https://doi.org/10.1126/sciadv.aax9444>

Author countries/areas: Belgium, Denmark, France, USA

Research funding: Fondation pour la Recherche sur la Biodiversité

Despite their crucial importance for terrestrial ecosystems, most tropical plants lack extinction risk assessments. IUCN Red List assessments are generated species-by-species, requiring knowledge, reliable data and thus, time. In 2019, less than eight per cent of the estimated number of plant species worldwide had been assessed.

(continued on next page...)



Preliminary Automated Conservation Assessments (PACA) may help the process along, as suggested by this paper, in which researchers used the RAINBIO dataset of tropical African plants to carry out a simultaneous extinction risk analysis of more than 22,000 vascular plant species in tropical Africa. The authors used GBIF-mediated data to filter out non-endemics with wide distributions outside the continent.

The final results of the PACA analysis, based on two IUCN Red List criteria, placed more than 30 per cent of species in the "likely/potentially threatened" category—with highest proportion of these found in Ethiopia, West Africa, Central Tanzania and southern Democratic Republic of the Congo.

### CLIMATE STABILIZATION—A DOUBLE-EDGED SWORD?

DATA USED: 566,184,545 SPECIES OCCURRENCES

📍 Ohashi H, Hasegawa T, Hirata A, Fujimori S, Takahashi K, Tsuyama I, Nakao K, Kominami Y, Tanaka N, Hijioka Y and Matsui T (2019) Biodiversity can benefit from climate stabilization despite adverse side effects of land-based mitigation. *Nature Communications*. Springer Science and Business Media LLC 10(1). Available at: <https://doi.org/10.1038/s41467-019-13241-y>

Author countries/areas: Austria, Japan

Climate change is a threat to all life on Earth and reducing greenhouse gas (GHG) emissions is a crucial mission of The Paris Agreement within the UN Framework Convention on Climate Change. Substantial GHG mitigation, however, may require significant land-use change, such as large-scale bioenergy crop production and afforestation. With land-use change being the largest driver of biodiversity loss to date, can the positive effects of GHG mitigation outweigh the negative effects of land-use change?

By comparing a GHG mitigation scenario to baseline in terms of changes to distribution models powered by GBIF-mediated data of more than 8,000 species of plants and animals, this study by researchers from Japan suggests that biodiversity overall can benefit from climate stabilization despite potential side effects of land-based mitigation.

Their results show that while some regions with much growth in land-based mitigation efforts (i.e., Europe and Oceania) are projected to suffer biodiversity losses, stringent GHG mitigation may bring a net benefit to global biodiversity—particularly in the latter half of the century.

### WHY ARE SOME SPECIES COMMON WHILE OTHERS ARE RARE?

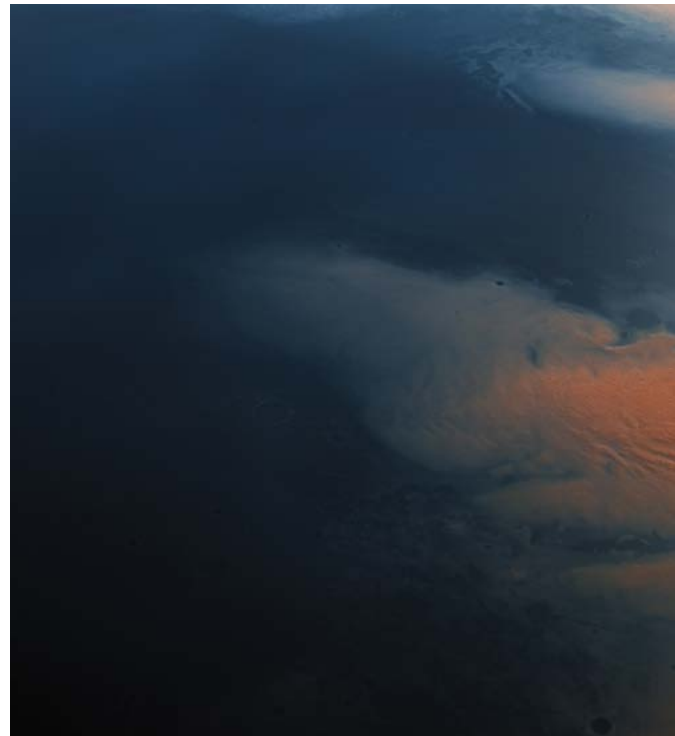
📍 Enquist BJ, Feng X, Boyle B, Maitner B, Newman EA, Jørgensen PM, Roehrdanz PR, Thiers BM, Burger JR, Corlett RT, Couvreur TLP, Dauby G, Donoghue JC, Foden W, Lovett JC, Marquet PA, Merow C, Midgley G, Morueta-Holme N, Neves DM, Oliveira-Filho AT, Kraft NJB, Park DS, Peet RK, Pillet M, Serra-Diaz JM, Sandel B, Schildhauer M, Šímová I, Violle C, Wieringa JJ, Wisser SK, Hannah L, Svenning J-C and McGill BJ (2019) The commonness of rarity: Global and future distribution of rarity across land plants. *Science Advances*. American Association for the Advancement of Science (AAAS) 5(11): eaaz0414. Available at: <https://doi.org/10.1126/sciadv.aaz0414>

Author countries/areas: Brazil, Chile, China, Czechia, Denmark, France, Netherlands, New Zealand, South Africa, UK, USA

Research funding: National Science Foundation

The phenomenon that some species are very common while others incredibly rare has always puzzled ecologists. Studies have investigated this on a local scale, but as most species are common in some parts of their ranges while rare in others, results are usually poor estimates of true global rarity.

This paper uses the Botanical Information and Ecology Network (BIEN) global plant database, which draws on GBIF as the main source of occurrences, to perform



an unprecedented assessment of global patterns of rarity. By quantifying the global species abundance distribution (gSAD)—a functional measure of true abundance—of 435,000 species, the authors showed that 36.5 per cent are rare, with just five records or fewer.

While distributed globally, very rare species are more common in hotspots found in Central and South America, southern Africa and Madagascar, and across Asia-Pacific from China to New Guinea.

When exploring drivers of rarity, models showed that the most important predictor of plant rarity is a stable climate, and that rapid changes in temperature negatively affects retention of rare species.

### CHOOSING LIFE ON MARS

📍 Vaz E and Penfound E (2020) Mars Terraforming: A Geographic Information Systems Framework. *Life Sciences in Space Research*. Elsevier BV 24: 50–63. Available at: <https://doi.org/10.1016/j.lssr.2019.12.001>

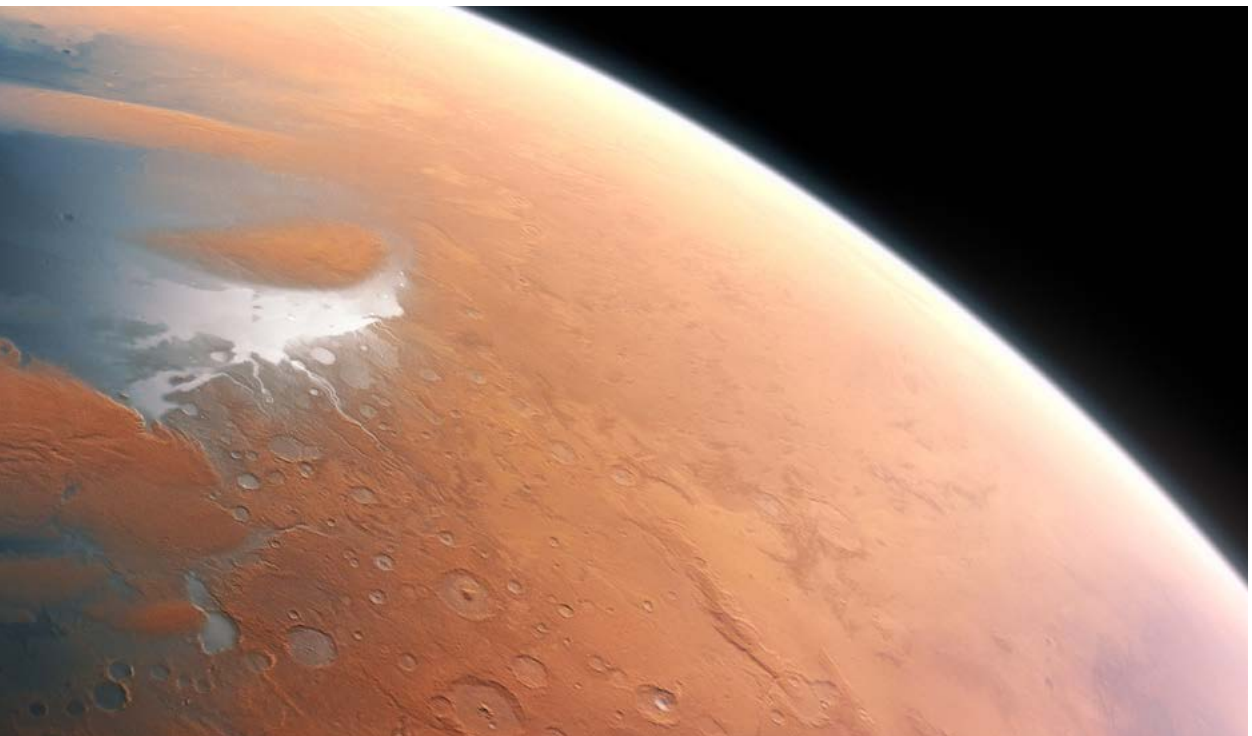
Author countries/areas: Canada

Home to millions of species, Earth is the only planet in our solar system able to sustain life as we know it. Travelling to and settling on other planets, like Mars, remains a fascination of humans, however, one that will require bringing hardy oxygen-producing plants to the Red Planet from Earth.

This study explores the potential of Mars terraforming through a suitability analysis identifying places on Earth with Mars-like environments, i.e. the coldest, driest and sunniest places on the planet, and using GBIF-mediated occurrences in these locations to identify relevant species likely to withstand hostile Martian environments.

Having the most Mars-like conditions, The Antarctic Peninsula and Ellesmere and Devon Islands in the Arctic were the focus of a vegetation analysis, in which the authors identified 21 tracheophyte and bryophyte genera with significantly higher proportions of occurrences within two of the study areas.

Further analysis indicated that the Arctic areas might be the better option in terms of finding suitable vegetation for Mars terraforming—and pointed to Bluegrass (*Poa*) as the most promising genus for further studies.



AN ARTIST'S IMPRESSION OF MARS FOUR BILLION YEARS AGO BY EUROPEAN SOUTHERN OBSERVATORY / M. KORNMESSENER VIA WIKIMEDIA COMMONS (CC BY-SA 4.0)

## Other research



HARVESTING CASSAVA BY IKHLASUL AMAL VIA FLICKR (CC BY-NC 2.0)

### GLOBAL ANALYSIS OF POTENTIAL MARGINAL LAND RESOURCES OF CASSAVA

DATA USED: 290,447 SPECIES OCCURRENCES

Jiang D, Wang Q, Ding F, Fu J and Hao M (2019) Potential marginal land resources of cassava worldwide: A data-driven analysis. *Renewable and Sustainable Energy Reviews*. Elsevier BV 104: 167–173. Available at: <https://doi.org/10.1016/j.rser.2019.01.024>

Author countries/areas: China

A drought-resistant tropical shrub capable of producing a rich, starchy root in marginal soil, cassava has been shown to be a suitable source of biofuel and a useful alternative to petroleum-based fuels.

Seeking to evaluate the potential of cassava to be grown on marginal lands at a global scale as a biofuel, researchers from the Chinese Academy of Sciences (CAS) used GBIF-mediated cassava occurrences combined with data on land use, climate and soil to create a boosted regression tree model mapping the environmental suitability of the plant.

With high accuracy and good performance, the obtained model was influenced primarily by land cover and annual precipitation with relative contributions of 46 and 35 per cent, respectively.

Focusing on relevant biomes, i.e. shrubland, savanna and grassland, the researchers estimated the worldwide marginal land area suitable for cassava to be 1.6 billion hectares—half of which are in Africa. With 394 million hectares, Brazil has the largest suitable land area followed by Angola with 105 million hectares.

### A NOVEL APPROACH TO CREATING AGGREGATED MAPS OF SPECIES OCCURRENCES

DATA USED: 64M SPECIES OCCURRENCES

Beilschmidt C, Mattig M, Fober T and Seeger B (2019) An efficient aggregation and overlap removal algorithm for circle maps. *Geoinformatica*. Springer Science and Business Media LLC 23(3): 473–498. Available at: <https://doi.org/10.1007/s10707-019-00342-5>

Author countries/areas: Germany

Research funding: Deutsche Forschungsgemeinschaft

Visualizing spatial data is an important tool when working with large volumes of species occurrences. Plotting millions of points on maps, however, is computationally expensive and overlaps can lead to occluded visualization.

In an attempt to overcome this, researchers from the University of Marburg in Germany designed a novel visualization algorithm for creating aggregated,

non-overlapping circle representation maps of point coordinates. Their approach—Circle Merging Quadtree—is based on an iterative transformation of all points into circles of a certain radius and then merging and (thus expanding) overlapping circles until no circles overlap.

Using GBIF-mediated occurrences of 50 species, varying from hundreds to millions of records per species, the authors compared the performance and quality of their algorithm to existing methods. While providing similar or better quality than the other tested methods, CMQ proved superior in terms of processing runtime—at up to two orders of magnitude.

### WIKIPEDIA PAGEVIEWS REFLECT OBSERVABLE PATTERNS IN PHENOLOGY

**a** Mittermeier JC, Roll U, Matthews TJ and Grenyer R (2019) A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLoS Biology*. Public Library of Science (PLoS) 17(3): e3000146. Available at: <https://doi.org/10.1371/journal.pbio.3000146>

Author countries/areas: Israel, Portugal, UK

Seasonal and interannual variations affect plant and animal life cycle events, such as flowers blooming and birds migrating. Phenology, the study of such events, can play a role in how and when humans interact with nature, a phenomenon with potential importance to conservation.

Using three years of log data from the online encyclopedia, Wikipedia, authors of this study analysed pageviews of articles of more than 31,000 species matched against the GBIF taxonomic backbone across 245 languages, to explore seasonal patterns in Wikipedia activity.

They found seasonal variation in more than 20 per cent of species pages, compared to just 6.5 per cent in a random sample of non-species pages. Most pages showed a single annual peak, while ~10 per cent had two annual peaks.

Reflecting actual phenological patterns, pageviews for groups like insects and flowering plants exhibited higher seasonality than mammals. Seasonal patterns were not consistent across Wikipedia languages, as fewer than two per cent of species showed the same pattern in all languages, with greater seasonality shown for languages spoken at higher latitudes, i.e. further from the Equator.

### RELYING ON BIODIVERSITY SCIENCE TO INFORM ART HISTORY

**a** Piñar G, Poyntner C, Tafer H and Sterflinger K (2019) A time travel story: metagenomic analyses decipher the unknown geographical shift and the storage history of possibly smuggled antique marble statues. *Annals of Microbiology*. Springer Science and Business Media LLC 69(10): 1001–1021. Available at: <https://doi.org/10.1007/s13213-019-1446-3>

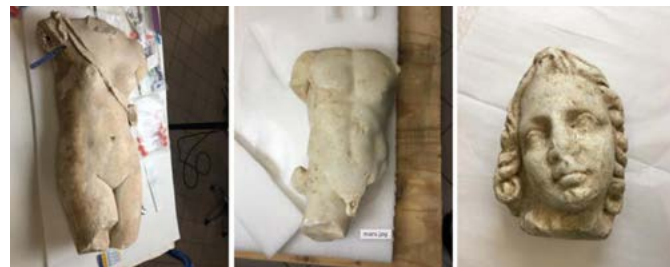
Author countries/areas: Austria

In forensic science, even the smallest traces of DNA can be used as biological markers to signify the presence of a person or object in a given context. But when no knowledge of context is available, advances in metagenomic barcoding coupled with knowledge of species distributions may be usefully exploited.

In this study, researchers from Austria sampled three marble statues of unknown origin for DNA extraction. They amplified the collected DNA and constructed libraries for quantification and taxonomic assignment. Matching the identified taxa with distributions of GBIF-mediated occurrences, the authors were able to draw some conclusions on the possible origins of the statues.

Two of the statues, representing torsos, appear to have been stored in agricultural soil, perhaps near an animal farm. As the genetic fingerprints were similar, the torsos are likely to have been stored together in recent times. Both carried sequences matching taxa endemic to Eastern Asia. The third statue, representing a head, however, differed considerably with a microbiome indicating longer storage time in arid conditions as well as marine environments in the past.

The most relevant find was the relative abundance of DNA of the genus *Taiwania* present on one torso. With a single extant species growing in the mountains of central Taiwan and locally in Southwest China, the torso is likely to have stayed or been transported through these areas.



THREE ANTIQUE MARBLE STATUES OF UNKNOWN ORIGIN ANALYSED IN THE STUDY. PHOTO FROM FIGURE 1 IN PIÑAR ET AL. 2019 (CC BY 4.0)



## SPATIOTEMPORAL BIASES IN ARCTIC HERBARIUM SPECIMEN COLLECTIONS

DATA USED: 29,749 SPECIES OCCURRENCES

**a** Panchen ZA, Doubt J, Kharouba HM and Johnston MO (2019) Patterns and biases in an Arctic herbarium specimen collection: Implications for phenological research. *Applications in Plant Sciences*. Wiley 7(3): e01229. Available at: <https://doi.org/10.1002/aps3.1229>

Author countries/areas: Canada

Research funding: Natural Sciences and Engineering Research Council of Canada, Killam Trusts

Biases in natural history collections may arise around certain climates and in remote areas, which in turn may impact the results of scientific studies based on data from such collections. This can be especially pronounced in time-sensitive phenological studies.

This study examines 29,000 herbarium specimens from Nunavut—the largest and most northerly territory of Canada—deposited at the National Herbarium of Canada [available through GBIF] to assess biases across time, space, and phenology, among others.

The limited access to the region as well as very short snow-free periods are apparent in the results of the study, showing 85 per cent of specimens collected in July and August. In terms of spatial distribution, only 0.63 per cent of the region contained one or more specimens—with the highest density of records found near the capital, Iqaluit.

The study also finds a preference towards collecting flowering plants in peak flower, as more than two-thirds of specimens fall into these categories when assessing phenological phase and flowering stage, respectively.

## THE EVOLUTION OF CLEANING BEHAVIOUR IN MARINE FISHES

DATA USED: OCCURRENCES OF FIVE FAMILIES

Baliga VB and Mehta RS (2019) Morphology, Ecology, and Biogeography of Independent Origins of Cleaning Behavior Around the World. *Integrative and Comparative Biology*. Oxford University Press (OUP) 59(3): 625–637. Available at: <https://doi.org/10.1093/icb/icz030>

Author countries/areas: USA

Cleaning is an often mutualistic behaviour among some marine species in which typically small and/or juvenile fishes (cleaners) locate and feed on ectoparasites off the bodies of larger taxa (clients). Some species obtain nearly all calories through cleaning while others use cleaning as a supplement.



DOTTED MAORI (*CHEILINUS CHLOROURUS*) BEING TENDED TO BY A BLUESTREAK CLEANER WRASSE (*LABROIDES DIMIDIATUS*). PHOTO BY MARK ROSENSTEIN VIA INATURALIST (CC BY-NC-SA 4.0)

In this study, researchers examined the evolutionary history of cleaning in five marine fish families in order to determine how and when this behaviour arose and whether the trait is restricted to species of a certain shape or size.

The authors found that cleaning evolved relatively recently with no one common ancestor in any family. By assigning species to distinct non-overlapping biogeographic regions using GBIF-mediated occurrences, they show that obligate cleaners in two families evolved in parallel and almost at same time in separate regions of the world: Labroides in the Indian Ocean and Elacatinus in the Western Atlantic.

By comparing photos of cleaner species with superimposed digitized body shape landmarks, they also showed that obligate cleaning is only present in elongate-bodied species with anterior pectoral fin attachment and eye positioning.

## OPENBIODIV: A KNOWLEDGE GRAPH FOR LINKED OPEN BIODIVERSITY DATA

DATA USED: GBIF TAXONOMIC BACKBONE

**a** Penev L, Dimitrova M, Senderov V, Zhelezov G, Georgiev T, Stoev P and Simov K (2019) OpenBiodiv: A Knowledge Graph for Literature-Extracted Linked Open Data in Biodiversity Science. *Publications. MDPI AG* 7(2): 38. Available at: <https://doi.org/10.3390/publications7020038>

Author countries/areas: Bulgaria, Sweden

Decades of biodiversity research have contributed immense knowledge about life on Earth. To fully exploit this knowledge, data must be available, openly and freely, while sources of data should be linkable through the use of stable and unique identifiers.

OpenBioDiv is a Open Biodiversity Knowledge Management System that uses semantic publishing workflows, graph database technologies, and text and data mining to establish a robust infrastructure for managing biodiversity knowledge.

Presented as a Linked Open Dataset, OpenBioDiv builds on data extracted from more than 5,000 scientific articles from Pensoft Publishers and even more taxonomic treatments mediated by Plazi. All this data is matched against an RDF version of the GBIF backbone taxonomy to ensure consistency.

The tool enables fast answers to complicated questions that would otherwise require querying numerous separate databases—e.g. "how many articles about taxon X has author Y published in the past 10 years", or, "which taxon treatments mention both scientific name X and Y?". Users can perform simple searches in the web portal while more complex queries are possible through a SPARQL endpoint.

### INSECTS OF MADAGASCAR

**DATA USED:** 286,764 SPECIES OCCURRENCES

**a** Plannella M, D'Alessandro P and Biondi M (2019) Entomological knowledge in Madagascar by GBIF datasets: estimates on the coverage and possible biases (Insecta). *Fragmenta Entomologica*. PAGEPress Publications 51(1): 1–10. Available at: <https://doi.org/10.4081/fe.2019.329>

**Author countries/areas:** Italy

One of the world's most important biodiversity hotspots, Madagascar is home to a disproportionately large and diverse flora and fauna with high levels of cryptic diversity and endemism.

This study examines available GBIF-mediated species occurrences of nine insect orders to evaluate the current coverage in knowledge of Malagasy entomofauna. With access to a dataset of 286,000 occurrences, researchers mapped and identified areas of high and low sampling, while investigating potential associations with specific ecosystems, protected areas and access infrastructure.



*CALLIMATION VENUSTUM* OBSERVED IN MORAMANGA, MADAGASCAR BY DAVID RABEHEVITRA. PHOTO VIA INATURALIST (CC BY-NC 4.0)

Their analysis revealed that the Eastern Madagascar subhumid forest and lowland rainforest host a significantly higher number of insect records than expected. More than 60 per cent of sampling happens within 1 km of a road, while no records are present beyond 10 km from a road.

The current protected areas of Madagascar house around 70 per cent of insect occurrences. A proposed expansion of the PA system would significantly improve this, especially for orders that are not currently well represented within PAs—Neuroptera, Odonata and Lepidoptera.

### XGBOOST: ACCURATE HIGH-RESOLUTION VEGETATION MAPPING

**DATA USED:** OCCURRENCES FROM 11,234 VEGETATION SURVEY PLOTS

**a** Zhang H, Eziz A, Xiao J, Tao S, Wang S, Tang Z, Zhu J and Fang J (2019) High-Resolution Vegetation Mapping Using eXtreme Gradient Boosting Based on Extensive Features. *Remote Sensing*. MDPI AG 11(12): 1505. Available at: <https://doi.org/10.3390/rs11121505>

**Author countries/areas:** China, France

**Research funding:** National Key Research and Development Program of China, National Natural Science Foundation of China

Mapping of ecosystems and vegetation types is important for environmental planning and resource management, but producing accurate, high-resolution maps is a challenge due to mosaic and heterogeneous environments.

This paper presents a workflow to produce high-resolution multiclass vegetation maps based on a machine learning algorithm for regression and classification referred to as XGBoost—or extreme gradient boosting.

To demonstrate the workflow, researchers used two strikingly different cases. First, they surveyed the Dzungarian Basin in China by photographic sampling covering more than 3,000 points 1-2 km apart. From the photos, they identified plant species and developed a hierarchical vegetation classification system.

In the second case, they gathered GBIF-mediated vegetation survey data for New Zealand. The authors supplemented both datasets with an extensive suite of bioclimatic data and remote sensing data.

In both cases, the authors used the XGBoost approach to produce high-resolution vegetation maps, effectively separating vegetation classes with higher accuracy than other mapping projects.

## FORMAL DESCRIPTION IS A PREDICTOR OF EXTINCTION RISK IN WEAKLY ELECTRIC FISHES

Tagliacollo VA, Camelier P, Zanata AM and Reis RE (2019) A shocking discovery of threat risks on newly described species of weakly electric fishes. *Journal of Fish Biology*. Wiley 96(5): 1077–1086. Available at: <https://doi.org/10.1111/jfb.14096>

**Author countries/areas:** Brazil

Among the most diverse of all vertebrate groups with nearly 16,000 described species, freshwater fishes are also the most highly threatened. Body and range size are good predictors of extinction risk, which has the implication that newly described species—usually small in size and with narrow ranges—could be more likely to be threatened.

This study examines the relationship between species traits, publication date and extinction risk in two groups of weakly electric fishes—Mormyroidea and Gymnotiformes. The authors compiled data on body size from FishBase and derived species distributions based on IUCN range maps or GBIF-mediated occurrences for 470 species.

Their analyses revealed that all except one of the larger species are considered Least Concern (LC) or Data Deficient (DD), while all widespread species are listed as LC. The models confirmed that newly-described species are more often small-bodied with narrow ranges and likely to be prone to extinction—especially in the African mormyroids, suggesting that their existence after being described may be short-lived.

## USING WW FOR SPECIES DISTRIBUTION MODELLING

**DATA USED:** 291,392 SPECIES OCCURRENCES

Deneu B, Servajean M, Botella C and Joly A (2019) Evaluation of Deep Species Distribution Models Using Environment and Co-occurrences. *Experimental IR Meets Multilinguality, Multimodality, and Interaction*. Springer International Publishing 213–225. Available at: [https://doi.org/10.1007/978-3-030-28577-7\\_18](https://doi.org/10.1007/978-3-030-28577-7_18)

**Author countries/areas:** France

Species distribution models (SDM) are key tools to obtaining knowledge about the spatial distribution of species, often achieved through ecological niche modelling, where environmental features are used to predict where species may occur.

Based on a conference challenge, this paper explores the use of deep neural networks (NN) in species distribution modelling, introducing a novel, convolutional NN model based on

environmental features and co-occurrences (species interdependencies) combined, and comparing its performance to three other SDM approaches.

The authors use a GBIF-mediated dataset of 3,300 plant species occurring in France, of which three quarters were used for training models and one quarter for testing. They showed that all models performed well, with environment-based models scoring highest. While the novel NN combined with co-occurrence data did get the highest score, its performance was not significantly higher than the model based on environmental features alone.

The study highlights that while NN can improve the performance of SDM, performance of models based solely on environmental data is outperformed by models capturing species interdependencies as well.

## CHECKLIST, TAXONOMY AND DISTRIBUTION OF RODENTS IN ETHIOPIA

**DATA USED:** 4,299 SPECIES OCCURRENCES

DBryja J, Meheretu Y, Šumbera R and Lavrenchenko LA (2019) Annotated checklist, taxonomy and distribution of rodents in Ethiopia. *Folia Zoologica*. Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic 68(3): 117. Available at: <https://doi.org/10.25225/fozo.030.2019>

**Author countries/areas:** Czechia, Ethiopia, Russia

Specialized to particular ecosystems and with low dispersal ability, rodents are generally good indicators of biodiversity levels. In eastern Africa, the geomorphologically diverse country of Ethiopia provides habitat for rich vertebrate fauna, yet the latest checklist of Ethiopian rodents is almost 25 years old.

Aimed at providing up-to-date knowledge on the biodiversity of rodents on Ethiopia, this paper compiles critically-reviewed data from literature, unpublished museum collections, field surveys and GBIF, providing maps of known distributions of 104 rodent species—including 40 species not included in the previous checklist (Yalden et al. 1996)

The study also produces complete taxonomic and ecological accounts for each species. The authors note the highest proportion of endemic species in the Ethiopian highlands and indicate that rodent diversity in Ethiopia is much higher than in other African countries of comparable size.

## USING GENETIC HEAT INDEX (GHI) FOR CONSERVATION PRIORITIZATION

Hawthorne WD and Marshall CAM (2019) Rapid Botanic Survey, Bioquality and improving botanical inventory in the tropics by integrating across spatial scales. *Gardens' Bulletin Singapore*. National Parks Board 71(suppl.2): 315–333. Available at: [https://doi.org/10.26492/gbs71\(suppl.2\).2019-21](https://doi.org/10.26492/gbs71(suppl.2).2019-21)

Author countries/areas: UK

Identifying areas of high biodiversity is important when prioritizing areas for conservation. Such hotspots can be defined by the number of species or levels of endemism, but richness alone is considered to have limited value for highlighting conservation priorities.

In this study, researchers propose 'bioquality' as a novel indicator of conservation priority—defined as biodiversity value measured by the concentration of globally rare taxa in a given plant community. Calculated by categorizing species based on an assessment of their global range and then summarizing in a Genetic Heat Index (GHI), bioquality can complement other parameters like richness and rarity.

Compiling occurrence data for more than 40,000 plant species from GBIF and other sources, the authors derive GHI scores for East Africa at varying resolutions, identifying hotspots and data gaps at both regional and local levels. To help fill gaps in sampling, the paper suggests Rapid Botanic Surveys, combining strengths of taxonomic and ecological recording, aiming to provide information about all plant species, community types and location of hotspots in a given area.

## ENVIRONMENTAL INFLUENCE ON ANGIOSPERM DIVERSITY OF DIFFERENT GROWTH FORMS

Zhou Y, Boru BH, Wang S and Wang Q (2020) Species richness and phylogenetic diversity of different growth forms of angiosperms across a biodiversity hotspot in the horn of Africa. *Journal of Systematics and Evolution*. Wiley. Available at: <https://doi.org/10.1111/jse.12559>

Author countries/areas: China, Ethiopia

The environment affects the evolutionary and biogeographical history of all species, affecting both richness and phylogenetic diversity. However, different forms of organisms may respond in different ways in their adaptation to changes.

In this study, researchers assessed patterns of species richness and phylogenetic diversity in angiosperm plants of East Africa, while considering differences between groups of woody and herbaceous species.

Using GBIF-mediated occurrences of more than 5,600 species, the authors analysed correlations between environmental variables and plant diversity metrics, finding similar tendencies among both angiosperm groups with positive responses to precipitation measures, topography and soil nutrients.

The results did, however, also show that environmental variables explained larger variance in phylogenetic diversity for woody plants compared to herbaceous plants, suggesting a greater role of the environment in shaping community structures among trees and shrubs than herbs.



*DRACAENA OMBET* OBSERVED NEAR ABALA, ETHIOPIA BY PFAUCHER. PHOTO VIA INATURALIST (CC BY-NC 4.0)





# GBIF

Global Biodiversity  
Information Facility

**GBIF Secretariat**

Universitetsparken 15  
DK-2100 Copenhagen Ø  
Denmark

tel +45 35 32 14 70

email [communication@gbif.org](mailto:communication@gbif.org)

web [www.gbif.org](http://www.gbif.org)



SILVER

PurePrint® by KLS