SUPPLEMENTAL INFORMATION

A. Descriptions of Datasets – The following datasets were used in geostatistical analyses and their results are given in figure 3 of the main manuscript. All were obtained on December 24, 2019 and all information below is made with respect to that date.

PlutoF:

Formal Name	PlutoF platform observations
Description	PlutoF is an online biodiversity data workbench. Observation records from Citizen Science and research projects hosted by PlutoF platform. All observations have been moderated and verified on PlutoF workbench by appointed persons. See: https://plutof.ut.ee/
Provider	PlutoF
Occurrences	1,288,808; 90% with geographical coordinates
DOI	https://www.gbif.org/dataset/169fa761-2fb9-4022-93bd-e22b7a062efd
Citation	PlutoF (2018). PlutoF platform observations. Version 1.17. Occurrence dataset https://doi.org/10.15156/bio/587440 accessed via GBIF.org

ZOBODAT:

Formal Name	Biologiezentrum Linz
Description	ZOBODAT is a digitally organized, biogeographical database, including an analysis, documentation and communication facility. The database contains data on the distribution of plant and animal species and various supporting digital libraries. The majority of the data traditionally relates to insects and geographically to Austria. ZOBODAT is usable for all organism groups and worldwide. See: http://www.landesmuseum.at/en/collection/nature/zobodat-zoological-botany-database.html
Provider	Biologiezentrum Linz Oberoesterreich, the Museum of zoology in Linz, Austria.
Occurrences	2,243,034; 97% with geographical coordinates
DOI	https://www.gbif.org/dataset/857bce66-f762-11e1-a439-00145eb45e9a
Citation	Biologiezentrum Linz Oberoesterreich. Biologiezentrum Linz. Occurrence dataset https://doi.org/10.15468/ynjblx accessed via GBIF.org on 2020-01-12.

Hattika:

Formal Name	Hatikka Observation Database
Description	Hatikka Observation Database is a digital field journal for Finnish amateur naturalists. It contains occurrence data for all taxon groups from Finland and several other countries. Data are mostly from Finland, smaller amount of data from other countries.
Provider	Finnish Biodiversity Information Facility: https://laji.fi/en
Occurrences	1,616,778; 99.6% with geographical coordinates
DOI	https://www.gbif.org/dataset/b84a3711-b4ca-4e4f-adac-80dfaea98d1c
Citation	Lahti T (2017). Hatikka Observation Database. Version 1.1. Finnish Biodiversity Information Facility. Occurrence dataset https://doi.org/10.15468/te1t6l accessed via GBIF.org on 2020-01-12.

***Edaphobase*:**

Formal Name	Edaphobase
Description	Edaphobase is a GBIF-D project collecting information from literature and museum collections about distribution and ecology of soil animals (earthworms, small earthworms, nematodes, springtails, moss/ beetle mites, gamasina mites, centipedes, millipedes, and woodlice).
Provider	Senckenberg Museum für Naturkunde Görlitz: https://museumgoerlitz.senckenberg.de/en/
Occurrences	394,134; 91% with geographical coordinates
DOI	https://www.gbif.org/dataset/82a421d4-f762-11e1-a439-00145eb45e9a
Citation	Senckenberg Museum für Naturkunde Görlitz (2018). Edaphobase. Occurrence dataset https://doi.org/10.15468/rk9xc7 accessed via GBIF.org on 2020-01-12.

***WTU*:**

Formal Name	Vascular Plant Collection - University of Washington Herbarium (WTU)
Description	Specimen collections, primarily from Pacific Northwest region of North America. Dataset contains approximately 154,000 specimens. Only Darwin Core 2 fields are accessible through this interface; additional fields can be accessed through the web site referenced herein. Localities and geographic coordinates for rare taxa have been withheld.
Provider	University of Washington Burke Museum
Occurrences	183,585; 58% with geographical coordinates
DOI	https://www.gbif.org/dataset/8310f570-f762-11e1-a439-00145eb45e9a
Citation	University of Washington Burke Museum. Vascular Plant Collection - University of Washington Herbarium (WTU). Occurrence dataset https://doi.org/10.15468/plngb6 accessed via GBIF.org on 2020-01-12.

***PRECIS*:**

Formal Name	PRECIS
Description	The largest computerized botanical database in Africa. The National Herbarium collection of South Africa contains more than 1.2 million specimens, of which over 900,000 are found in the PRECIS database.
Provider	South African National Biodiversity Institute:
Occurrences	1,124,142; 80% with geographical coordinates
DOI	https://www.gbif.org/dataset/8310f570-f762-11e1-a439-00145eb45e9a
Citation	South African National Biodiversity Institute (2018). PRECIS. Occurrence dataset https://doi.org/10.15468/rckmn2 accessed via GBIF.org

SAFRING:

Formal Name	SAFRING: Historical Bird Ringing Records (2005-2009)
Description	Bird ringing started in 1948 in southern Africa and is thus the longest running bird monitoring project in the sub-region. Although the data are biased to areas where ringers operated, it is an invaluable record of bird occurrence and distribution changes, arrival and departure times of migrants, survival and mass data in historical times.The South African Bird Ringing Unit (SAFRING), which is an essential arm of the Animal Demography Unit, administers bird ringing in southern Africa.
Provider	South African National Biodiversity Institute.
Occurrences	2,162,753; 98% with geographical coordinates
DOI	https://www.gbif.org/dataset/b4ae1720-1431-49ee-bfeb-8146fc42b1a3
Citation	Oschadleus D, Ranwashe F (2017). SAFRING: Historical Bird Ringing Records (2005-2009). Version 1.1. South African National Biodiversity Institute. Occurrence dataset https://doi.org/10.15468/jclwaz accessed via GBIF.org on 2020-01-12.

CONN:

Formal Name	CONN
Description	George Safford Torrey Herbarium specimens. The herbarium combines significant palaeobotanical, bryological, lichenological, mycological, phycological and vascular plants totaling over 160,000 specimens, all housed in a fully modern, state-of-the-art facility. See: https://biodiversity.uconn.edu/herbarium/
Provider	University of Connecticut
Occurrences	172,098; 91% with geographical coordinates
DOI	https://www.gbif.org/dataset/5288946d-5fcf-4b53-8fd3-74f4cc6b53fc
Citation	Capers R (2014). CONN. University of Connecticut. Occurrence dataset https://doi.org/10.15468/w35jmd accessed via GBIF.org on 2020-01-12.

UCFC:

Formal Name	Stuart M. Fullerton Collection of Arthropods (UCFC), University of Central Florida
Description	Vouchered occurrence records for arthropods, primarily insects, from the Stuart M. Fullerton Collection of Arthropods at the University of Central Florida.
Provider	Museum of Biological Diversity, The Ohio State University
Occurrences	564,020; 99.9% with geographical coordinates
DOI	https://www.gbif.org/dataset/262f8270-f9c2-4bc6-a562-8ed71c0790e6
Citation	Song H, Johnson N (2018). Stuart M. Fullerton Collection of Arthropods (UCFC), University of Central Florida. Version 85.36. Museum of Biological Diversity, The Ohio State University. Occurrence dataset https://doi.org/10.15468/kyulwg accessed via GBIF.org on 2020-01-12.

SAHFOS:

Formal Name	Continuous Plankton Recorder Dataset (SAHFOS)
Description	SAHFOS contains data on the geographical distribution, seasonal cycles and year-to-year changes in abundance of plankton over a large spatial area. There is no other survey in the world with 80 years of marine plankton data. CPR Data are available in paper form back to 1931 and from January 1946 onwards in a computerised relational database. The database contains results with 2,629,628 taxonomic abundance entries.
Provider	Sir Alister Hardy Foundation for Ocean Science (SAHFOS)
Occurrences	2,629,628; 99.9% with geographical coordinates
DOI	https://www.gbif.org/dataset/67c54f85-7910-4cbf-8de4-6f0b136a0e34
Citation	Sir Alister Hardy Foundation for Ocean Science (SAHFOS). Continuous Plankton Recorder Dataset (SAHFOS). Version 1.0. Occurrence dataset https://doi.org/10.15468/ygwilu accessed via GBIF.org on 2020-01-12.

SAIAB:

Formal Name	Occurrence records of southern African aquatic biodiversity
Description	Specimen-records (of physical specimens) of fishes, mostly from southern Africa and surrounding oceans, but also from elsewhere in the world.
Provider	South African Institute for Aquatic Biodiversity (SAIAB)
Occurrences	136,855; 83% with geographical coordinates
DOI	https://www.gbif.org/dataset/1aaec653-c71c-4695-9b6e-0e26214dd817
Citation	Coetzer W (2017). Occurrence records of southern African aquatic biodiversity. Version 1.10. The South African Institute for Aquatic Biodiversity. Occurrence dataset https://doi.org/10.15468/pv7vds accessed via GBIF.org on 2020-01-12.

TexaAM:

Formal Name	Texas A&M University Insect Collection
Description	Selected insect specimen records from the Texas A&M University Insect Collection.
Provider	Texas A&M University Insect Collection
Occurrences	404,842; 84% with geographical coordinates
DOI	https://www.gbif.org/dataset/96193ea2-f762-11e1-a439-00145eb45e9a
Citation	Texas A&M University Insect Collection. Texas A&M University Insect Collection. Occurrence dataset https://doi.org/10.15468/caprqh accessed via GBIF.org on 2020-01-12.

CMHerps:

Formal Name	CM Herps Collection
Description	Herpetology maintains a collection of more than 207,500 specimens and ranks as about the ninth largest amphibian and reptile collection in the United States. Ninety percent are fluid preserved; others are preserved as skeletons, skins, mounts, or cleared and stained preparations. Notable historic collections rich in type specimens are the Taylor Philippine collection, the Le Boutellier collection of South American snakes, and specimens from early museum expeditions to the Isle of Pines and Angola. Collection data are completely computerized.
Provider	VertNet: http://vertnet.org/
Occurrences	170,040; 67% with geographical coordinates
DOI	https://www.gbif.org/dataset/76dd8f0d-2daa-4a69-9fcd-55e04230334a#description
Citation	Rogers S (2017). CM Herps Collection. Version 8.5. Carnegie Museums. Occurrence dataset https://doi.org/10.15468/lxe6h4 accessed via GBIF.org on 2020-01-12.

Arctic:

Formal Name	Arctic Ocean Diversity
Description	Occurrence data of biodiversity found in the sea ice, water column and sea floor located in Arctic, Pacific and coastal Alaska.
Provider	Alaska Ocean Observing System: https://aoos.org/
Occurrences	180,013; 83% with geographical coordinates
DOI	https://www.gbif.org/dataset/84b72ee4-f762-11e1-a439-00145eb45e9a
Citation	Rogers S (2017). CM Herps Collection. Version 8.5. Carnegie Museums. Occurrence dataset https://doi.org/10.15468/lxe6h4 accessed via GBIF.org on 2020-01-12.

UofVienna:

Formal Name	University of Vienna, Institute for Botany - Herbarium WU
Description	The foundation of this collection of dried plant specimens dates back to 1879. Today the herbarium of the Institute of Botany is estimated to contain about 1,400,000 specimens covering all plant groups. See: https://herbarium.univie.ac.at/database/search.php
Provider	University of Vienna, Institute for Botany - Herbarium WU
Occurrences	117,480; 57% with geographical coordinates
DOI	https://www.gbif.org/dataset/0afba960-be3b-4202-a7de-736ae05aec9e
Citation	University of Vienna, Institute for Botany - Herbarium WU. University of Vienna, Institute for Botany - Herbarium WU. Occurrence dataset https://doi.org/10.15468/tnj8wm accessed via GBIF.org on 2020-01-12.

IllNat:

Formal Name	Illinois Natural History Survey Insect Collection
Description	The insect collection at the Survey houses over 6,000,000 curated specimens. The INHS insect collection is one of the largest in North America. The INHS Insect Collection includes more than 13,000 primary types (syntypes inclusive).
Provider	Illinois Natural History Survey
Occurrences	656,321; 78% with geographical coordinates
DOI	https://www.gbif.org/dataset/68513375-3aa5-4f6f-9975-d97d56c21d61#description
Citation	Dmitriev D (2015). Illinois Natural History Survey Insect Collection. Illinois Natural History Survey. Occurrence dataset https://doi.org/10.15468/eol0pe accessed via GBIF.org on 2020-01-12.

UPS:

Formal Name	Botany (UPS)
Description	Database of the botany collection of the Museum of Evolution in Uppsala.
Provider	GBIF-Sweden
Occurrences	706,679; 4% with geographical coordinates
DOI	https://www.gbif.org/dataset/c1a13bf0-0c71-11dd-84d4-b8a03c50a862
Citation	Telenius A, Shah M (2018). Botany (UPS). GBIF-Sweden. Occurrence dataset https://doi.org/10.15468/ufmslw accessed via GBIF.org on 2020-01-12.

LACM:

Formal Name	LACM Entomology Collection
Description	The Natural History Museum's entomology collection has more than 5.8 million specimens of insects and spiders. The collection's strength lies in its holdings of specimens of ants, phorid flies, scarab beetles, and moths from North and Central America. Museum scientists conduct world-class research on systematics, studying species and their relationships, the evolution of major groups, and fossil insects in amber. They conduct field work on insect biodiversity at home and in tropical countries. Entomology at the Natural History Museum of Los Angeles County goes back as far as 1913, when the Museum first opened. The entomology collection is the largest in Southern California. Although the phorid flies represented in this data are from all over the world, they are primarily from North and South America with an emphasis on Costa Rica, Colombia, USA, Equador, and Panama.
.Provider	VertNet: http://vertnet.org/
Occurrences	184,364; 98% with geographical coordinates
DOI	https://www.gbif.org/dataset/0ec927cf-325a-4d63-9499-d721c734463a
Citation	Mertz B (2018). LACM Entomology Collection. Version 5.2. Natural History Museum of Los Angeles County. Occurrence dataset https://doi.org/10.15468/kc9hyp accessed via GBIF.org on 2020-01-12.

FishBase:

Formal Name	Fishbase
Description	Fishbase occurrences hosted by GBIF-Sweden.
Provider	GBIF-Sweden
Occurrences	731,045; 67% with geographical coordinates
DOI	https://www.gbif.org/dataset/197908d0-5565-11d8-b290-b8a03c50a862
Citation	Norén M, Shah M (2017). Fishbase. FishBase. Occurrence dataset https://doi.org/10.15468/wk3zk7 accessed via GBIF.org on 2020-01-12.

JFish:

Formal Name	Fish collection of National Museum of Nature and Science
Description	Fish specimens deposited at the Department of Zoology, National Museum of Nature and Science, Japan. See: https://www.kahaku.go.jp/english/research/specimen/index.html
Provider	National Museum of Nature and Science, Japan
Occurrences	108,775; 21% with geographical coordinates
DOI	https://www.gbif.org/dataset/84a1a7e0-f762-11e1-a439-00145eb45e9a
Citation	Nakae M, Shinohara G (2018). Fish collection of National Museum of Nature and Science. National Museum of Nature and Science, Japan. Occurrence dataset https://doi.org/10.15468/w3dzv1 accessed via GBIF.org on 2020-01-12.

CMBird:

Formal Name	CM Birds Collection
Description	The section cares for over 187,000 specimens of birds, and a database over 206,000 records which include exchanged specimens and other specimens no longer in the collection. The most important of these are the 519 holotypes and 40 syntypes. We also care for approximately 196 extinct birds as well as specimens of many rare species collected decades if not more than a century ago. The collection on whole is ranked roughly ninth in the United States. The Carnegie Collection has over 154,000 study skins, almost 16,000 skeletons of which over 5650 have an accompanying spread wing prepared, many with tails, over 10,000 egg sets, 6760 fluid specimens, 440 flat skins and about 1250 taxidermy mounts.
Provider	VertNet: http://vertnet.org/
Occurrences	205,967; 1% with geographical coordinates
DOI	https://www.gbif.org/dataset/a82d2421-be5a-4779-babc-da0b634a7ba8
Citation	Rogers S (2016). CM Birds Collection. Version 9.1. Carnegie Museums. Occurrence dataset https://doi.org/10.15468/dv1ojv accessed via GBIF.org on 2020-01-12.

OSUC:

Formal Name	C.A. Triplehorn Insect Collection (OSUC), Ohio State University
Description	The Triplehorn Insect Collection is comprised of over 3.5 million cataloged specimens, including one of the world's largest leafhopper collections. Initiated by Professor Josef N. Knull in 1934, the collection has strong holdings in the Coleoptera, Hemiptera, Hymenoptera, Odonata, and Orthoptera. The collection was a recent recipient of a National Science Foundation facilities grant. See: https://mbd.osu.edu/collections#insects
Provider	Museum of Biological Diversity, The Ohio State University
Occurrences	498,736; 99.9% with geographical coordinates
DOI	https://www.gbif.org/dataset/84ab7b76-f762-11e1-a439-00145eb45e9a
Citation	Johnson N, Cora J. C.A. Triplehorn Insect Collection (OSUC), Ohio State University. Museum of Biological Diversity, The Ohio State University. Occurrence dataset https://doi.org/10.15468/efb17f accessed via GBIF.org on 2020-01-12.

ROM:

Formal Name	Mammalogy Collection - Royal Ontario Museum
Description	The Mammal collection consists of 120,000 skins, skeletons, wet specimens and is strong in Chiroptera and New World Rodents. The geographic strengths are Ontario (pre-1985), Guyana, Suriname, Colombia, southeastern Asia, Kenya, Uganda and Zimbabwe. Approximately 24,000 of the more recent voucher specimens have associated frozen tissues.
Provider	Royal Ontario Museum
Occurrences	119,382; 75% with geographical coordinates
DOI	https://www.gbif.org/dataset/c5c4a23e-2035-4416-ab64-032d6df52ddb
Citation	Millen B, Lim B (2018). Mammalogy Collection - Royal Ontario Museum. Version 11.5. Royal Ontario Museum. Occurrence dataset https://doi.org/10.15468/2rlrvh accessed via GBIF.org on 2020-01-12.

PaleoB:

Formal Name	Paleobiology Database
Description	The Paleobiology Database (PBDB) is a non-governmental, non-profit public resource for paleontological data. It has been organized and operated by a multi-disciplinary, multi-institutional, international group of paleobiological researchers. Its purpose is to provide global, collection-based occurrence and taxonomic data for organisms of all geological ages, as well data services to allow easy access to data for independent development of analytical tools, visualization software, and applications of all types. The Database's broader goal is to encourage and enable data-driven collaborative efforts that address large-scale paleobiological questions.
Provider	VertNet: http://vertnet.org/
Occurrences	1,310,925; 98% with geographical coordinates
DOI	https://www.gbif.org/dataset/bb5b30b4-827e-4d5e-a86a-825d65cb6583
Citation	McClennen M, Jenkins J, Uhen M (2017). Paleobiology Database. Paleobiology Database. Occurrence dataset https://doi.org/10.15468/jfqhiu accessed via GBIF.org on 2020-01-12.

RoyBot:

Formal Name	Royal Botanic Garden Edinburgh Herbarium (E)
Description	RBGE's extensive Herbarium numbers nearly three million specimens representing half to two thirds of the world's flora. It is considered a leading botanical collection. See: https://www.rbge.org.uk/
Provider	Royal Botanic Garden Edinburgh
Occurrences	903,681; 9% with geographical coordinates
DOI	https://www.gbif.org/dataset/bb5b30b4-827e-4d5e-a86a-825d65cb6583
Citation	McClennen M, Jenkins J, Uhen M (2017). Paleobiology Database. Paleobiology Database. Occurrence dataset https://doi.org/10.15468/jfqhiu accessed via GBIF.org on 2020-01-12.

Eumyceto:

Formal Name	Planetary Biodiversity Inventory Eumycetozoan Databank
Description	The eumycetozoans (or slime molds) are a monophyletic group of microscopic organisms characterized by an amoeba-like trophic stage and aerial spore-bearing reproductive structures. Eumycetozoans are widespread and often common to even abundant in nature, where they are major predators of other microorganisms (bacteria, yeasts, cyanobacteria and green algae). The project, currently being carried out at the University of Arkansas and funded by two major grants from the National Science Foundation (NSF), seeks to expand, standardize, systematize and (ultimately) summarize the body of information available on the taxonomy, ecology and biogeographical distribution of all three groups of eumycetozoans (myxomycetes, dictyostelids and protostelids). See: http://slimemold.uark.edu/aboutframe.htm
Provider	University of Arkansas
Occurrences	223,245; 87% with geographical coordinates
DOI	https://www.gbif.org/dataset/85af82a6-f762-11e1-a439-00145eb45e9a
Citation	University of Arkansas. Planetary Biodiversity Inventory Eumycetozoan Databank. Occurrence dataset https://doi.org/10.15468/zxmhsy accessed via GBIF.org on 2020-01-12.

Macaulay:

Formal Name	Macaulay Library Audio and Video Collection
Description	The Macaulay Library is the world's largest and oldest scientific archive of biodiversity audio and video recordings. Our mission is to collect and preserve recordings of each species' behavior and natural history, to facilitate the ability of others to collect and preserve such recordings, and to actively promote the use of these recordings for diverse purposes spanning scientific research, education, conservation, and the arts.
Provider	VertNet: http://vertnet.org/
Occurrences	220,286; 64% with geographical coordinates
DOI	https://www.gbif.org/dataset/7f6dd0f7-9ed4-49c0-bb71-b2a9c7fed9f1
Citation	Scholes III, Ph.D. E (2015). Macaulay Library Audio and Video Collection. Cornell Lab of Ornithology. Occurrence dataset https://doi.org/10.15468/ckcdpy accessed via GBIF.org on 2020-01-12.

Geneva:

Formal Name	Geneva Herbarium – General Collection (G)
Description	The General Collection of the Geneva Herbarium is estimated to contain 6 million specimens of plants (vascular plants, bryophytes and algae), fungi (Basidiomycetes, Ascomycetes, lichens), Oomycetes and Myxomycetes from around the world.
Provider	Conservatoire et Jardin botaniques de la Ville de Genève - G
Occurrences	188,801; 99.3% with geographical coordinates
DOI	https://www.gbif.org/dataset/f577c9f3-ae71-4278-b6bf-512ba1dfaa21
Citation	Conservatoire et Jardin botaniques de la Ville de Genève - G. Geneva Herbarium – General Collection (G). Occurrence dataset https://doi.org/10.15468/rvjdu1 accessed via GBIF.org on 2020-01-12.

MICROBIS:

Formal Name	MICROBIS database
Description	MICROBIS is a database of marine microbial biota, most gained from genomic analysis used in support of ICoMM. The International Census of Marine Microbes (ICoMM) facilitates the inventory of marine microbial diversity. It will develop a strategy to (1) catalogue all known diversity of single-cell organisms inclusive of the Bacteria, Archaea, Protista and associated viruses, (2) explore and discover unknown microbial diversity, and (3) place that knowledge into ecological and evolutionary contexts.
Provider	Marine Biology Laboratory
Occurrences	889,985; 100% with geographical coordinates
DOI	https://www.gbif.org/dataset/840ebec6-f762-11e1-a439-00145eb45e9a
Citation	Marine Biology Laboratory. MICROBIS database. Occurrence dataset https://doi.org/10.15468/wi6v9k accessed via GBIF.org on 2020-01-12.

AMNH-M:

Formal Name	AMNH Mammal Collections
Description	The scope of the collections is worldwide, with areas of particular strength including Australia and New Guinea (30,000 specimens), Central Asia (12,000 specimens), Central Africa (30,000 specimens), and South America (50,000 specimens). Taxonomic coverage is particularly broad (100% of the 26 mammalian orders; 96% of the 136 families; about 50% of the 1,135 genera and about 60% of the 4,629 species), and the collection contains over 1,200 name-bearing type specimens. The collections of marsupials, insectivores, bats, primates, rodents, carnivores, whales and ungulates are recognized as being among the best in the world.
Provider	VertNet: http://vertnet.org/
Occurrences	290,333; 5% with geographical coordinates
DOI	https://www.gbif.org/dataset/96ca66b4-f762-11e1-a439-00145eb45e9a#description
Citation	Trombone T (2016). AMNH Mammal Collections. American Museum of Natural History. Occurrence dataset https://doi.org/10.15468/wu3poe accessed via GBIF.org on 2020-01-12.

Mosquito:

Formal Name	Mosquito Occurrence Dataset
Description	Detailed data on the past and present distribution of mosquito vectors. See: http://www.wrbu.org/
Provider	Walter Reed Biosystematics Unit, Smithsonian Institution
Occurrences	126,031; 99.9% with geographical coordinates
DOI	https://www.gbif.org/dataset/88e38292-f762-11e1-a439-00145eb45e9a
Citation	Walter Reed Biosystematics Unit, Smithsonian Institution. Mosquito Occurrence Dataset. Occurrence dataset https://doi.org/10.15468/tw2y9b accessed via GBIF.org on 2020-01-12.

Neptune:

Formal Name	Neptune Deep-Sea Microfossil Occurrence Database
Description	Micropaleontologic study of deep-sea sediments cored over several decades by the DSDP, ODP and IODP drilling programs. The Neptune database provides integrated access to this global dataset. It contains over 500,000 records for the occurrences of species in individual, numerically age-dated samples, age models for hundreds of deep-sea sections, and extensive, annotated, quality-controlled taxonomic lists for thousands of fossil marine plankton species. See: https://www.gbif.org/dataset/e1e16cf0-ada2-11e2-8fbc-00145eb45e9a#description
Provider	Museum für Naturkunde Berlin
Occurrences	500,808; 100% with geographical coordinates
DOI	https://www.gbif.org/dataset/e1e16cf0-ada2-11e2-8fbc-00145eb45e9a
Citation	Museum für Naturkunde Berlin. Neptune Deep-Sea Microfossil Occurrence Database. Occurrence dataset https://doi.org/10.15468/bvu6a8 accessed via GBIF.org on 2020-01-12.

MCDB:

Formal Name	Mammal Community Database
Description	Taken from the data paper's Abstract: A data set that includes species lists for 1000 mammal communities, excluding bats, with species-level abundances available for 940 of these communities. All communities found in the literature that included complete, site-specific sampling data, composed of species lists with or without associated abundances, were included in the data set. Most, but not all, sites are limited to species groups that are sampled using a single technique (e.g., small mammals sampled with Sherman traps). The dataset consists of 7977 records from 1000 sites encompassing habitats throughout the world, and including data on 660 mammal species with sizes ranging from 2 g to >500 kg.
Provider	Ecological Archives E092-201-D1
Occurrences	1,000; 100% with geographical coordinates
DOI	http://esapubs.org/archive/ecol/E092/201/
Citation	Katherine M. Thibault, Sarah R. Supp, Mikaelle Giffin, Ethan P. White, and S. K. Morgan Ernest. 2011. Species composition and abundance of mammalian communities. Ecology 92:2316.

EMP:

Formal Name	Earth Microbiome Project
Description	The Earth Microbiome Project (EMP) is a massively collaborative effort to characterize microbial life on this planet. We use DNA sequencing and mass spectrometry of crowd-sourced samples to understand patterns in microbial ecology across the biomes and habitats of our planet. The EMP is a comprehensive example of open science, leveraging a collaborative network of 500+ investigators, supporting pre-publication data sharing, and crowdsourcing data analysis to enable universal principles to be explored. The standardized collection, curation, and analysis are enabling a robust interpretation of ecological trends.
Provider	Earth Microbiome Project: http://www.earthmicrobiome.org/
Occurrences	27,751; 80.5% with geographical coordinates
DOI	http://www.earthmicrobiome.org/
Citation	Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., …, Jansson, J. K., Gilbert, J. A., Knight, R., & The Earth Microbiome Project Consortium. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. Nature, 551:457-463. doi:10.1038/nature24621

GPC:

Formal Name	Global Prokaryotic Census
Description	A compilation of microbial community data obtained from previous studies. See: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000106
Provider	Loucalab:http: //www.loucalab.com/archive/GPC
Occurrences	34,368; 73.5% with geographical coordinates
DOI	http://www.earthmicrobiome.org/
Citation	Louca, S., Mazel, F., Doebeli, M., & Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. PLoS biology, 17(2), e3000106.

GBI:

Formal Name	Global Biotic Interactions
Description	Global Biotic Interactions (GloBI) provides open access to finding species interaction data (e.g., predator-prey, pollinator-plant, pathogen-host, parasite-host) by combining existing open datasets using open source software. GloBI is made possible by a community of software engineers, bioinformaticists and biologists. GloBI is sustained by a network of open source, open data and open science communities in addition to receiving donations, grants, awards or being written into grants, including, but not limited to, EOL's EOL Rubenstein Fellows Program (CRDF EOL-33066-13/F33066, 2013) and the David M. Rubenstein Grant (FOCX-14-60988-1, 2014), and the Smithsonian Institution (SI) (T15CC10297-002, 2016). If you would like to contribute to GloBI please visit our contribute page.
Provider	Global Biotic Interactions: https://www.globalbioticinteractions.org/
Occurrences	4,385,686; 50% with geographical coordinates
DOI	https://www.globalbioticinteractions.org/data.html
Citation	Jorrit H. Poelen, James D. Simons and Chris J. Mungall. (2014). Global Biotic Interactions: An open infrastructure to share and analyze species-interaction datasets. Ecological Informatics.