Toward Better Conservation: A Spatial Analysis of Species Occurrence Data from the Global Biodiversity Information Facility

Susmita Dasgupta
Lead Environmental Economist, Development Economics Research Group,
MC- 3-347, World Bank,1818 H Street, Washington DC 20433, USA
sdasgupta@worldbank.org
Brian Blankespoor
Senior Geographer, Development Economics Data Group, MC-9-204,
World Bank,1818 H Street, Washington DC 20433, USA bblankespoor@worldbank.org
and
David Wheeler
Consultant, Development Economics Research Group,
World Bank,1818 H Street, Washington DC 20433, USA
wheelrdr@gmail.com

Corresponding author: Brian Blankespoor, Senior Geographer, Development Economics Data Group, MC-9-204, The World Bank, 1818 H Street, Washington, DC 20008, USA. Tel: 1-202-473-1546. Email: bblankespoor@worldbank.org

Abstract

The world is facing an unprecedented loss of biodiversity, with nearly one million species on the brink of extinction, and the extinction rate accelerating. Conservation efforts are often hindered by insufficient information on crucial ecosystems. To address this issue, our paper leverages advances in machine-based pattern recognition to estimate species occurrence maps using georeferenced data from the Global Biodiversity Information Facility (GBIF). Our algorithms have generated maps for more than 600,000 species, including vertebrates, arthropods, mollusks, other animals, vascular plants, fungi, and other organisms. Validation involved comparing these maps with expert maps for mammals, ants, and vascular plants. We found a close similarity in global distribution patterns, with regional differences attributed to technical variations or necessary revisions in existing expert maps based on GBIF data. As a practical application, we identified the global distributions of approximately 68,000 species with small ranges (25 km x 25 km or less) confined to a single country. Our maps reveal a skewed international distribution of these species, identifying 30 countries where 78.2 percent are concentrated. These results highlight the need to integrate the newly mapped GBIF data into global conservation planning. Our algorithms support rapid updates and the creation of new maps as GBIF occurrence reports increase. The data are available on the World Bank Development Data Hub at https://doi.org/10.57966/h21e-vq42 (Dasgupta et al. 2024).

Keywords: Conservation planning, global biodiversity, species' occurrence region, endemic and small-occurrence region, Kunming-Montreal Global Biodiversity Framework.

500-character non-technical summary including space

Short Summary

This study leverages recent advances in machine-based pattern recognition to estimate occurrence maps for over 600,000 species, using georeferenced data from the Global Biodiversity Information Facility (GBIF). A pilot application for priority-setting identifies 30 nations that host nearly 80 percent of threatened species with small ranges limited to a single country. The algorithms are designed for rapid map updates and estimating new maps as growth in GBIF species occurrence reports continues.

1. Introduction

The world is losing biodiversity at an unprecedented rate. One million plant and animal species may be near extinction, and the pace of extinction is accelerating. According to Pimm et al. (2014), the species extinction rate is at least one thousand times the background rate. The Living Planet Index, an indicator of global biodiversity based on population trends for vertebrate species in terrestrial, freshwater, and marine habitats, provides corroborating evidence, showing a 69 percent decline since 1970. In response to such alarming indicators, 188 governments in the Convention on Biological Diversity (CBD) ratified the Kunming-Montreal Global Biodiversity Framework (GBF) at the fifteenth meeting of the CBD's Conference of the Parties (COP 15) in December 2022. Among other measures, the GBF committed participants to protecting 30 percent of global biodiversity by 2030 (UNEP, 2022). Effectively implementing the GBF requires an understanding of (1) the spatial distribution of global biodiversity to be protected and (2) how protecting 30 percent of the planet can best conserve this biodiversity, taking the opportunity value of protected areas into account.

Unfortunately, conservation efforts worldwide are often hindered by limited information on critical ecosystems and biodiversity (Hughes et al. 2024). Comprehensive species coverage is significantly lacking; instead, the predominant focus is on vertebrates and vascular plants, neglecting crucial taxa like invertebrates and other major phyla. This gap creates a policy dilemma for meeting the 2030 GBF commitment of 30 percent global protection. If global biodiversity assessments are limited to previously mapped species, policy makers and the conservation community will effectively ignore the enormous population of other species whose occurrences are reported by the Global Biodiversity Information Facility (GBIF). In short, policymakers cannot automatically assume that previously mapped species adequately represent the larger population, a point further discussed by Kass et al. (2022) in the case of invertebrates.

To help bridge the data gap, this study uses GBIF species occurrence records to revisit global biodiversity's spatial distribution. GBIF's reporting network has expanded over 15 years to include over 2 million species' occurrences, with a daily increase of about 1.3 million records in the last two years. Most of these records include locational coordinates, making it possible to estimate the spatial distribution of previously unmapped species, as well as improve estimates for those with existing maps. The algorithm that is presented and implemented in this study generates species maps directly from the GBIF occurrence data, updating the maps automatically as new occurrence data become available. Maps are generated for all species whose data satisfy our computational criteria (currently around 600,000 species). By overlaying the maps with a high-resolution grid, the view of global biodiversity is broadened from the traditional focus on vertebrate animals to encompass greater representation for invertebrates, other animals, plants, fungi, and other non-animal and non-plant species. The maps are then used to develop new indicators of species endemism and identify species with small, vulnerable habitats. Traditionally, species are considered endemic if they reside 100% in one country; this study examines the effects of lowering this threshold to 95% and 90%. Additionally, since small-range status lacks a definitive minimum habitat size, the study explores different area sizes for species with limited occurrence regions.

The study's approach should be viewed as complementary to previous biodiversity assessments (e.g. Map of Life) and can benefit the policy process in several major ways. For existing mapped species, rapid updates using the study's algorithm can help to identify cases where newly reported occurrences suggest alteration of map boundaries. For unmapped species, the approach can provide new information useful for global biodiversity assessments. In addition, the mapping exercise can yield valuable insights on the global distribution of endemic species and small-range species that are especially vulnerable to human encroachment.

The open-access data, algorithm, and programming code developed in this study align with principles of open and equitable sharing of location-specific biodiversity information (Blankespoor et al., 2025; Tanalgo, 2025). It is hoped that this approach would facilitate coordination among countries and enable policymakers, researchers, and conservation practitioners to take informed, coordinated actions to safeguard biodiversity.

2. Data and methods

2.1 Data source and tools

Our data source is an international network funded by the world's governments that provides open access to data about all types of life on Earth. Other international organizations with which the GBIF collaborates include the Catalogue of Life partnership, Biodiversity Information Standards, Consortium for the Barcode of Life (CBOL), Encyclopedia of Life (EOL), and GEOSS. The GBIF provides a continuously updated, open-source repository of geolocated, date-stamped reports of species occurrences from many institutions and nongovernmental organizations (NGOs) worldwide. These reports can be accessed directly from the GBIF's Occurrence and Maps application programming interfaces (APIs) at https://www.gbif.org/developer/occurrence and https://www.gbif.org/developer/maps, respectively. Its georeferenced data for plants, animals, fungi, and microbes hold the potential for vastly expanding the species domain maps that provide a critical foundation for global conservation planning. Access to the GBIF's full database is offered through Google's BigQuery, Amazon Web Services (AWS), and other cloud-based services.

In this study, we use BigQuery to download the full GBIF database due to its convenient dataset size reduction tools (GBIF 2024). We limit the GBIF occurrence data to geolocated reports since 1970 for species with at least five unique reporting locations, which our mapping algorithm requires. Since our mapping algorithm operates reliably on several thousand points at most, we cap the data at a maximum of 20,000 randomly selected reports per species to ensure reliability. This limit drastically reduces the size of the download dataset since some species have millions of occurrence records (e.g., the American robin [*Turdus migratorius*] currently has 21,258,907 reported occurrences). We accept the GBIF's protocols for occurrence report admissibility. Detailed descriptions of the GBIF's protocols and database elements can be found at https://www.gbif.org/data-quality-requirements-occurrences.

2.2 Analysis

The premise of our analysis is that species maps underpin spatial analyses of global biodiversity. In theory, if all species are treated equally, a map of global biodiversity could be created by (1) choosing the best available map for each species; (2) overlaying all chosen maps on a high-resolution global grid; and (3) counting the total species incidence in each grid cell. In practice, however, global biodiversity analyses can modify species counting in several major ways. For example, species may be assigned weights, based on the branches they occupy in the "tree of life," which describes overall genetic variation in the global biome. Also, species weights for conservation priority-setting may vary, based on the species' widely differing vulnerabilities to human encroachment. In addition, the distribution of a species may not be uniform in the spaces enclosed by its maps. If its spatial distribution density is known, its counting weight for each cell can be made proportional to its likelihood of occurrence in that cell. Abundant scientific literature offers examples of weighted species counting (e.g., Guo et al., 2022; Jenkins et al., 2015; Pimm et al., 2014; Veach et al., 2017); however, the requisite research may require detailed genetic and environmental data, as well as expert analysis of their roles in assigning counting weights. Inevitably, the intensive processes involved are time-consuming, requiring technical resources that are in short supply. As a result, a large gap has emerged between the population of species with GBIF occurrence records and that for which research-driven counting weights are available.

The mapping algorithm utilized in this study is a by-product of recent advances in machine-based pattern recognition, cluster analysis, and image processing. In terms of computational geometry, it addresses the problem of efficient bounding of a spatial set, given a subset of actually observed points. Traditional algorithms that draw simple convex hulls poorly represent sets with irregular shapes as a polygon is considered convex if none of its corners bend inward and a convex hull is the smallest convex polygon that encloses all points in a set. In contrast, this study's alphahull algorithm, developed by Pateiro-López and Rodríguez-Casal (2010), which is a function in the R programming language can construct continuous non-convex boundaries for efficient representation. This powerful feature has motivated alphahull's rapid adoption for species range analysis (Guo et al., 2022; Kass et al., 2022).

In our study, alphahull successfully estimates occurrence maps for 92.9 percent (567,464) of the 610,694 species in our database. For each of the remaining 7.1 percent of species in the database, we employ a standard k-means algorithm to separate occurrence reports into spatial clusters and draw a convex hull around each. We assess the utility of these mapping algorithms for delineating species occurrence regions, with illustrative examples provided in the Annex. Our algorithms estimate occurrence maps for terrestrial, coastal, and marine species.

2.3 Spatial selection bias

We acknowledge that GBIF occurrence reports are often produced by voluntary exercises that do not utilize scientific sampling methods. As a result, spatial point densities in species occurrence reports are positively related to physical accessibility, population density, and income (Borgelt et al., 2022; Garcia-Rosello et al., 2023; Isaac and Pocock, 2015; Reddy and Dávalos, 2003). This means that (all else being equal) species sightings are more likely to occur in areas (1) near transport arteries; (2) with a greater number of inhabitants to identify species, and (3) where more inhabitants have enough disposable income to support species search and reporting costs.

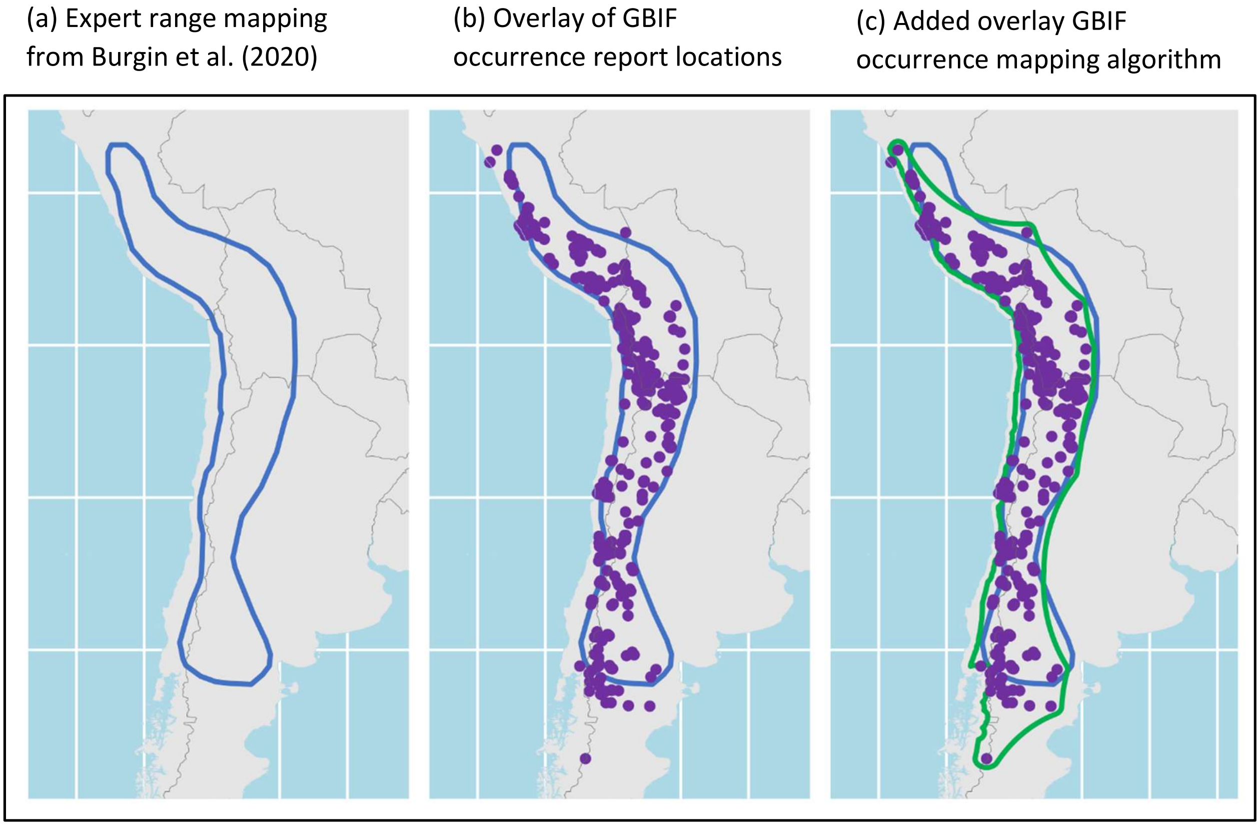
These factors complicate attempts to map species population densities from occurrence reports (e.g., Kass et al., 2022). Our alphahull and clustered convex hull estimators for boundaries differ because they focus on exterior points in spatial sets. Even so, accurate representation requires a critical minimum number of sightings in areas not advantaged by transport access, high population density, or sufficient disposable income. As the occurrence of sightings in non-advantaged areas increases, so does the accuracy of boundary estimation (Feeley and Smith, 2011). Given that the GBIF occurrence inventory is growing by about 1.3 million new reports per day, one can expect that, over time, increased sightings will improve the boundary estimates for sparsely reported species.

3. Results

3.1 Pilot example

To illustrate the results of our mapping algorithm exercise, we take the example of the species *Lagidium viscacia* (common name: Mountain Viscacha), whose range extends from areas in Argentina and Chile to Bolivia and Peru. In Fig. 1, a comparison of panels (a) and (b) shows that the reported sightings of this species include a few points beyond the northern boundary of the expert range map (Burgin et al., 2020a), along with many points beyond the southern boundary. Panel (c), which displays the output of our mapping algorithm, shows that the alphahull boundary follows the curvilinear north-south orientation of the point set, widening and narrowing as the point set expands and contracts. As shown, it overlaps heavily with the map of Burgin et al. (2020a), but extends its northern and southern boundary areas to incorporate these sightings.

Figure 1. Mapping exercise results for *Lagidium viscacia* (Mountain Viscacha), showing overlapping boundaries



(a) Expert range mapping from Burgin et al. (2020)

(b) Overlay of GBIF occurrence report locations

(c) Added overlay GBIF occurrence mapping algorithm

● GBIF Occurrence Reports

― Expert Range Map (Burgin et al. 2020)

― GBIF Occurrence Mapping Algorithm

Lagidium viscacia observed in Bolivia by Miglė Montrimaitė

3.2 Overall mapping results

Occurrence record maps were computed for 610,694 GBIF species, comprising 52,433 vertebrates (amphibians, birds, fish, mammals, and reptiles), 213,268 arthropods, 32,355 mollusks, 24,109 other animals, 232,693 plants, 38,122 fungi, and 17,714 other species (the kingdoms Archaea, Bacteria, Chromista, Protozoa, and Viruses) (Table 1).

Table 1. GBIF species occurrence maps

Species classification	Group	Count based on presence
Vertebrates		
Class	Amphibians	5,055
	Birds	11,064
	Mammals	4,881
	Reptiles	7,644
	Fish	23,789
Subtotal		**52,433**
Arthropods		
Order	Araneae	10,438
	Coleoptera	44,152
	Diptera	23,567
	Hemiptera	13,272
	Hymenoptera	26,159
	Lepidoptera	50,675
	Other	45,005
Subtotal		**213,268**
Mollusks		
Phylum		32,355
Other Animals		24,109
Vascular Plants		
Order	Asterales	22,978
	Asparagales	17,439
	Caryophyllales	9,554
	Ericales	8,574
	Fabales	16,277
	Gentianales	13,811
	Lamiales	16,545
	Malpighiales	12,790
	Myrtales	10,336
	Poales	16,845
	Other	87,544
Subtotal		**232,693**
Fungi		
Kingdom		38,122
Other Species		17,714
Total		**610,694**

The exercise was limited to GBIF species in three kingdoms (Animalia, Plantae, and Fungi) with at least three unique geolocated occurrences since 1970. To remove spurious observations from locales (e.g., zoos and botanical gardens), we relied on the following two methods: (1) exclusion of isolated outlier occurrences before map estimation, which happens automatically in our mapping algorithms and (2) exclusion of bounded point sets with fewer than three observations after map estimation. For the many species maps that have multiple bounded areas, we imposed a conservative interpretation of the evidence, dropping species maps with single-bounded areas when they contain fewer than three observations. Although this final condition may seem redundant; it is important to include as a species can pass the initial condition and fail the final one since our estimation algorithms may exclude an outlier point or two from their computations of the bounded areas. Examples of the convex and alpha hull comparisons are included in the Annex.

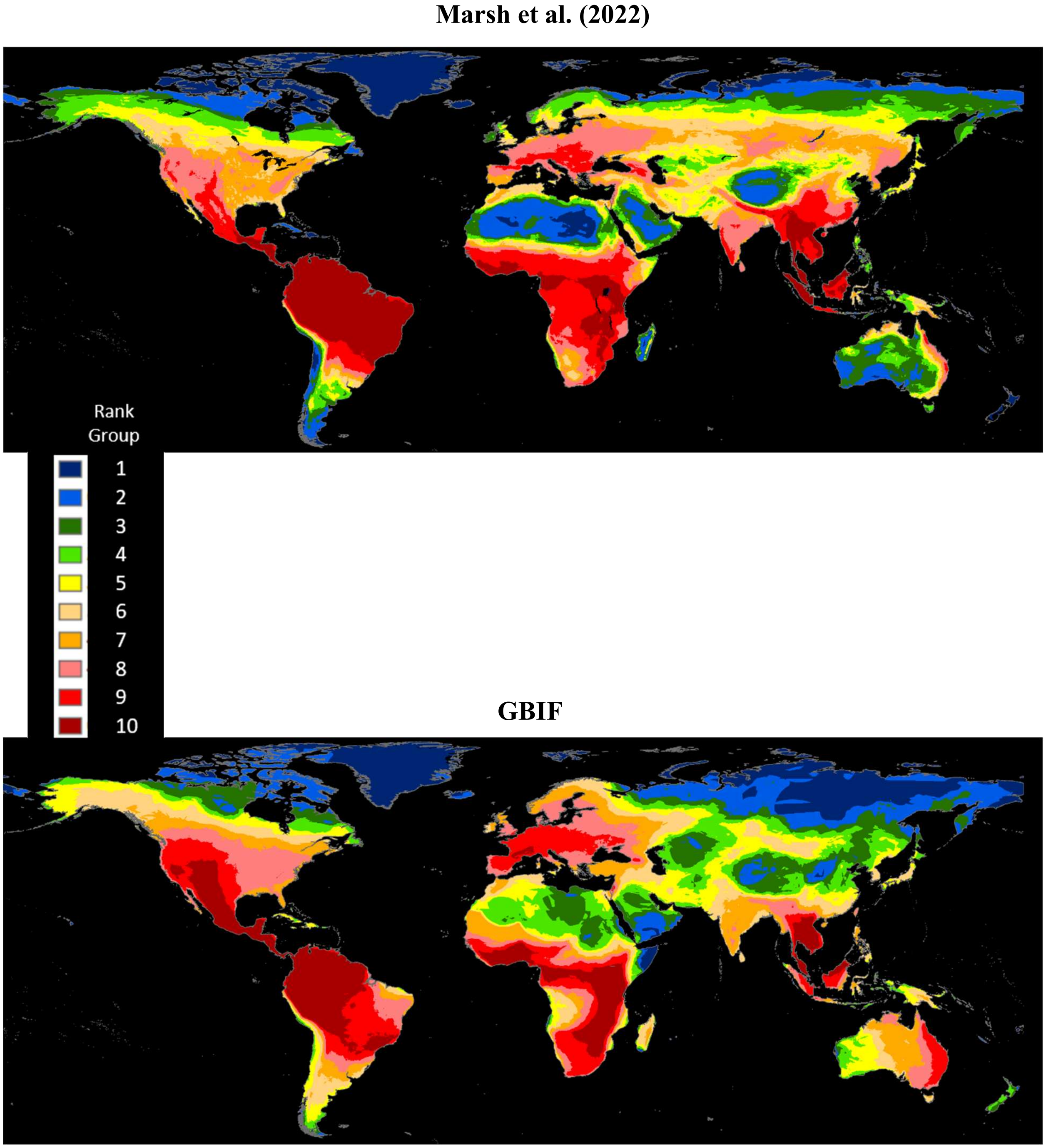
3.3 Case comparisons

To test our species boundary mapping from the current inventory, we ask whether the view of global biodiversity distribution it provides is consistent with that of existing expert maps. On comparing our estimated GBIF occurrence maps with expert maps from recently published research, we find that thousands of species with GBIF maps have been mapped by the research teams. Using these matched species, each comparison assesses the similarity in global biodiversity patterns produced by our GBIF maps and the expert research products by rank group based on the pixel level distribution of the species maps from 1 (lowest count) to 10 (highest count). Where the patterns diverge, we explore the technical factors that can explain the differences. The first case comparison retains the traditional focus on vertebrates, comparing mammal range maps estimated by Marsh et al. (2022). The second one focuses on a comparison with maps for ants developed by Kass et al. (2022), while the third centers on a limited set of vascular plants mapped by Borgelt et al. (2022). Invertebrates are significantly underrepresented in existing expert maps (Kass et al., 2022). At the outset, it should be noted that this study's major contribution is the expanded coverage of invertebrates. As shown in Table 1, our work offers more comprehensive representation by estimating maps for 213,268 arthropods.

3.3.1 Mammals

Marsh et al. (2022) map the native ranges of mammals globally using the authoritative taxonomy provided by the Mammal Diversity Database (Burgin et al., 2018). Their exercise harmonizes species maps from the *Checklist of the Mammals of the World* (Burgin et al., 2020b, 2020c), and the *Handbook of the Mammals of the World* published in nine volumes by Mittermeier, Rylands, and Wilson (2013), Wilson, Lacher, and Mittermeier (2016, 2017), and Wilson and Mittermeier (2009, 2011, 2014, 2015, 2018, 2019). In our GBIF occurrence maps database, we identify 3,530 mammals that are also mapped by Marsh. We rasterize both sets of maps using a global grid with 0.05 degree (5 km) resolution. For each species map, rasterization assigns a value of 1 to grid cells that overlap with the map and 0 to other cells. Next, we compute species densities by cellwise addition across 3,530 rasters for each set. Figure 2 compares cell counts, which are ranked in 10 groups. The maps' broad patterns are visibly similar. Both assign ranks in the highest two groups to Central America, northwest South America, West Africa, East Africa, the northern region of Central Africa, the eastern region of southern Africa, Western Europe and Southeast Asia. Notably, they also differ in other regions. The GBIF map assigns higher ranks to large areas of Mexico, western United States, and eastern Australia and lower ranks to the southeastern Amazon region and South Asia.

Figure 2. Matched mammal species densities: Marsh et al. (2022) versus GBIF occurrence reports

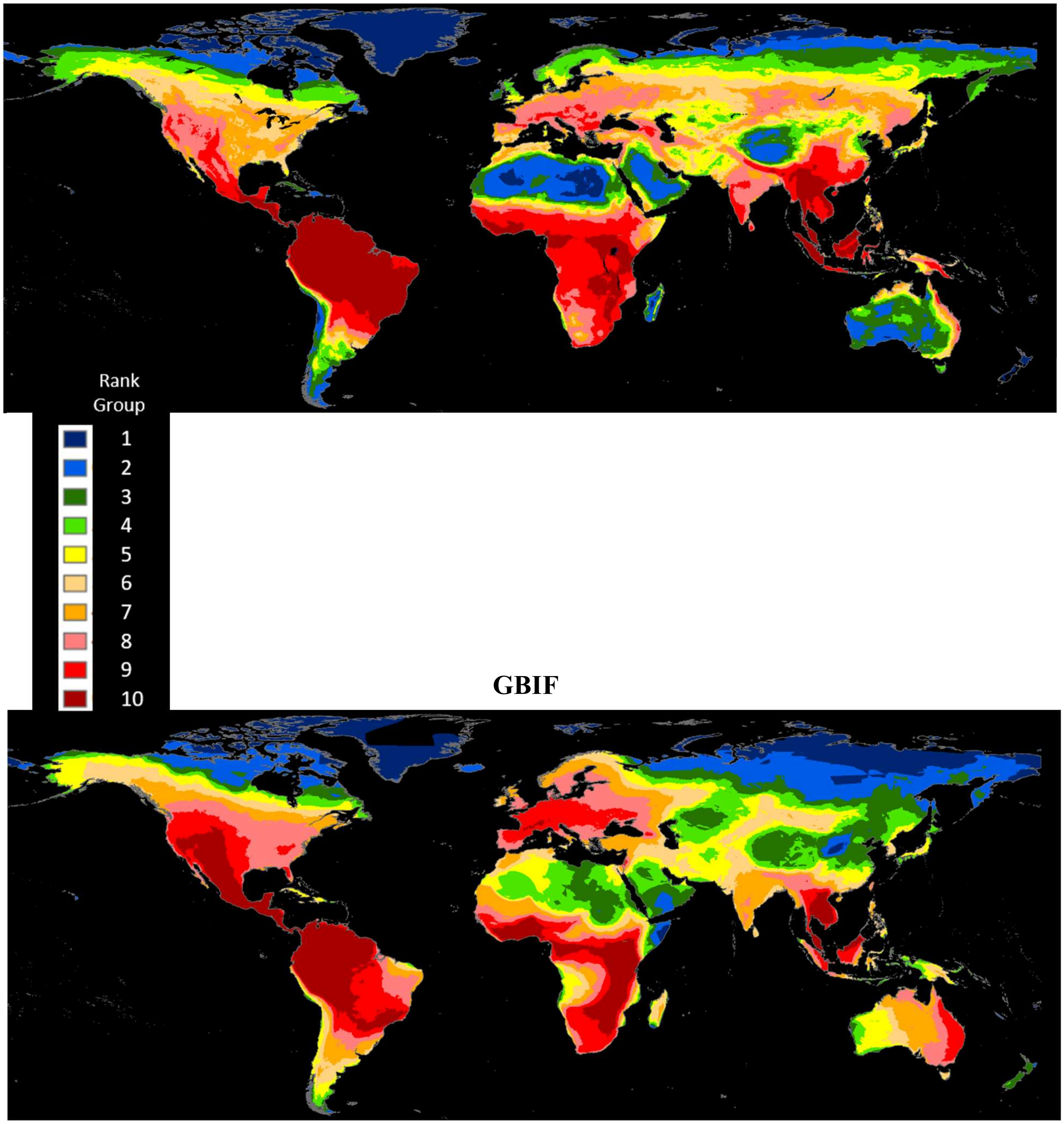

Marsh et al. (2022)

GBIF

Technical differences can explain divergences in the two patterns. For example, the Marsh et al. (2022) maps estimate native ranges of global mammals, taking into account recorded historical occurrences and biogeographical factors that correlate with the range of each mammal species. By contrast, the GBIF mammal maps bound the areas where species occurrences have been reported since 1970. Regions where GBIF ranks are higher than Marsh et al. (2022) ranks have many species with reported occurrences beyond their estimated native ranges; regions with lower GBIF ranks have occurrence reports clustered in subareas within native ranges. This difference could reflect underreporting for GBIF species in lower-ranked areas, although many higher-ranked areas appear similarly disadvantaged for species observation. In our view, the more plausible explanation is that lower-ranked regions are populated by many species whose ranges have contracted over time. The ongoing accumulation of GBIF species occurrence reports should help to resolve this issue.

Figure 2 compares 3,530 mammals with maps in both databases, while Fig. 3 does the same for all mapped terrestrial mammals (4,138 for GBIF and 6,360 for Marsh et al. [2022]). Comparison with Fig. 2 reveals almost no difference for GBIF; however, Marsh et al. (2022) have generally higher rankings for Indonesia and Papua New Guinea. Mammal species may be underrepresented in GBIF occurrence reports from the two countries, although this seems more likely for sparsely populated Papua New Guinea than densely populated Indonesia. The more likely explanation, in our view, is that the areas populated by many mammals have contracted.

Figure 3. Full mammal species densities: Marsh et al. (2022) versus GBIF occurrence reports



Marsh et al. (2022)

GBIF

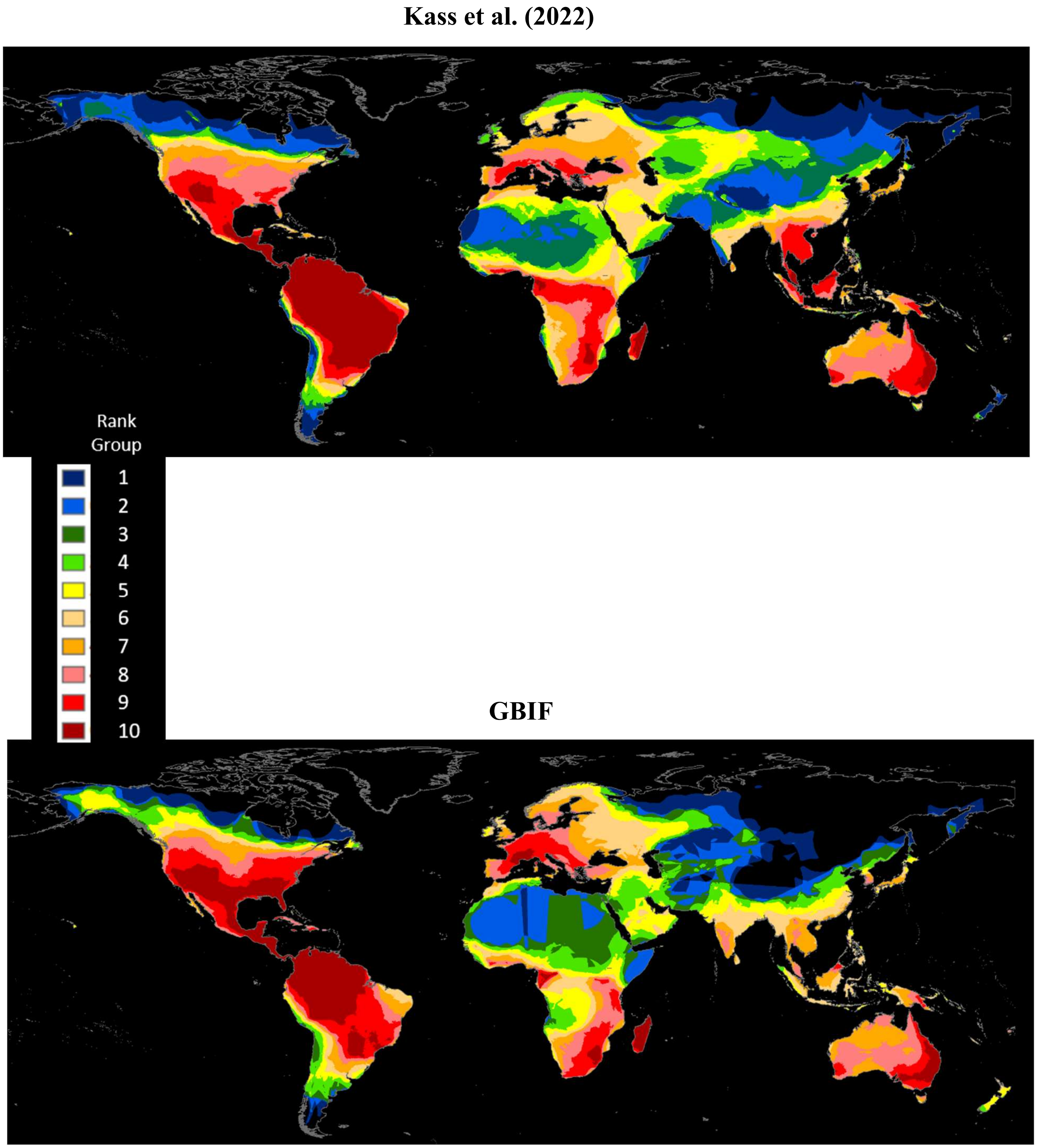
3.3.2 Ants

Clark and May (2002) identified a severe taxonomic bias in conservation research, finding that vertebrates accounted for only 3 percent of described species but 69 percent of published papers. Conversely, invertebrates accounted for 79 percent of described species and just 11 percent of published papers (Leather, Basset, and Hawkins, 2008). Kass et al. (2022) address this problem for ants using a variety of datasets and techniques, including the alphahull algorithm, to estimate the range maps. In our GBIF occurrence maps database, we identify 5,445 ant species also mapped by Kass et al. (2022). We rasterize both sets of maps using a global grid with 0.05 degree (5 km) resolution and compute species densities by cellwise addition across 5,445 rasters for each set.

Figure 4 compares the cell counts, which are ranked in 10 groups. Many areas exhibit similar patterns, including northern North America, Mexico, Central America, northwest South America, Eastern and Western Europe, West Africa, southern Africa, Madagascar, and eastern Australia. However, there are three notable differences. First, both maps identify a large high-ranking region in the Western Hemisphere, which is further north for GBIF than for Kass et al. (2022). Second, both maps identify a band of relatively high ranks across West and northern Central Africa, linking to a north-south band in East and southern Africa; however, the rankings for GBIF are generally lower than those for Kass et al. (2022). Third, Southeast Asia ranks uniformly higher for Kass et al. (2022) than for GBIF.

Since Kass et al. (2022) also rely heavily on the alphahull methodology, we attribute these differences to two technical factors. First, their database comes from intensive processing and error checking of records drawn from the Global Ant Biodiversity Informatics (GABI) database in July 2020. In our study, by contrast, the records are drawn from GBIF occurrence data, as of July 2023. Second, our approach is significantly more conservative. For example, we exclude unique species occurrences that number fewer than three, while Kass et al. (2022) include them; since alphahulls cannot be estimated for these 5,168 ant species, Kass et al. (2022) estimate their ranges by drawing 30 km buffer zones around the occurrence locations. Given this difference, comparing full database results for GBIF and Kass et al. (2022) would be, in effect, comparing apples and oranges.

Figure 4. Matched ant species densities: Kass et al. (2022) versus GBIF occurrence reports


Kass et al. (2022)

GBIF

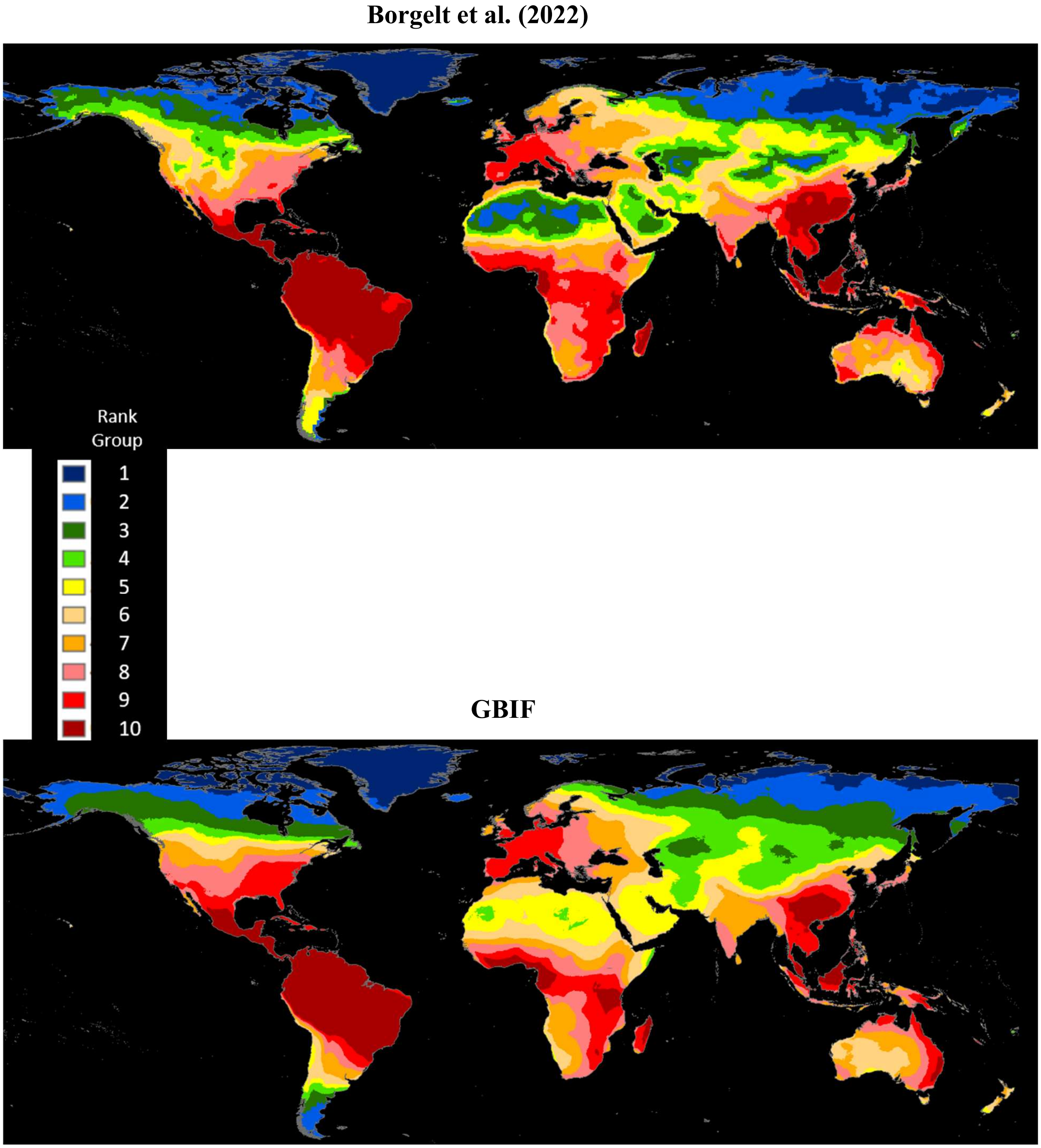
3.3.3 Vascular plants

Borgelt et al. (2022) have recently developed spatial density maps for vascular plants in the International Union for Conservation of Nature (IUCN) Red List (IUCN, 2021). They utilize maximum entropy (Maxent) models that predict the likelihood of species occurrences from the values of several environmental variables. For each species, they identify native regions from a web-scraping exercise using the Plants of the World Online (POWO) database, with regional identification standardized from the World Geographical Scheme for Recording Plant Distributions (WGSRPD). Typically, the resulting native regions are the boundaries of small countries or provinces (GADM level-1 administrative units) in large countries. Borgelt et al. (2022) estimate the models using GBIF occurrence data with restrictive prior conditions. To preserve compatibility with the environmental variables used for Maxent estimation, the data are confined to the 2000–20 period. For each species, georeferenced occurrence reports exclude all observations outside pre-identified native regions, and Maxent-estimated species distributions are also confined to native regions. The advantage of this approach is that it guarantees the exclusion of spurious observations from such entities as botanical gardens and private collections in other regions. One drawback, however, is that it incurs the cost of excluding potentially large numbers of occurrence observations that lie outside pre-identified native regions that are arbitrarily defined by national or provincial boundaries.

Unlike Borgelt et al. (2022), our exercise is not constrained by the need for compatibility with environmental modeling variables; therefore, we draw on a longer time period (1970–2023). Also, we impose no prior geographic restrictions on the data. As previously explained, our methodologies estimate occurrence map boundaries after eliminating spurious single outliers and small, isolated occurrence clusters. We identify 32,339 vascular plant species found in both databases, and, as before, rasterize our occurrence maps and compute cell counts at 5 km resolution. As Borgelt et al. (2022) provide species maps in a raster stack with much coarser resolution (50 km), we next extract raster layers for these 32,339 common species and add across layers to obtain relative incidence scores for the 50-km grid cells. Finally, we use mean smoothing to approximate the effect of higher resolution.

Figure 5 displays the comparative results as ranks in 10 groups; the two maps share essentially the same density pattern, except for the somewhat more extensive high-ranking areas in South and Southeast Asia for the Borgelt et al. (2022) maps and the United States for our study's maps.

Figure 5. Matched vascular-plant species densities: Borgelt et al. (2022) versus GBIF occurrence reports



Borgelt et al. (2022)

GBIF

3.4 Summing up

In all three case comparisons, we find quite similar global patterns of species density with strong agreement between the sources for each case. From a confusion matrix of 5 to 10 classes based on distribution of the species rank group, we find that across these taxa overall accuracy and agreement are significantly better than random (accuracy p-value = 0) with plants consistently outperforming ant and mammals (see Annex). Additionally, we calculated the local Spearman correlation coefficient for each case and found high agreement in areas with low population and levels of lights at night. Where the patterns diverge, the discrepancies can be traced to technical differences. In the case of mammals, differences between the GBIF and expert native-range maps can be attributed to either undershooting, where expert map boundaries exclude many GBIF occurrences, or overshooting, where GBIF occurrences are persistently absent in parts of the native-range maps. In the case of ants, where the research also utilizes alphahull estimation, differences are attributable to differences in source databases and our relatively conservative approach to map estimation. Finally, in the case of vascular plants, where the research employs GBIF occurrences and the global pattern similarity is most striking, the few discrepancies are attributable to temporal and spatial restrictions imposed by the expert research team.

4. Priority-setting applications

The effectiveness of biodiversity conservation plans will require identification of occurrence regions for species with elevated extinction risks. While multiple approaches exist for setting conservation priorities, decisions on resource allocation are typically made at national and subnational levels, particularly in the Global South. Endemic species and those with very restricted ranges provide a practical entry point, as they are highly vulnerable, confined to smaller geographic areas, and relatively straightforward to target with conservation measures. Protecting these species not only contributes to global biodiversity goals but also safeguards the unique natural heritage of individual nations, creating an added incentive for action. Therefore, using the maps developed with GBIF data, we explored (1) species endemic to a single country and (2) species under continuous threat owing to their small occurrence regions. Our database reflects GBIF-sourced occurrence maps for previously unmapped species, as well as revised estimates for those with existing maps.

4.1 Endemic species distribution by group and country

With the new dataset, we explored the endemic status assigned to species that are 100 percent resident in a single country. By this criterion, 44.6 percent (272,189) of the 610,694 species maps tabulated by country are classed as endemic. The incidence of endemism differs widely by species group (e.g., 54.5 percent for mollusks, 48.7 percent for vascular plants, 47 percent for other animals, 44 percent for arthropods, 37.1 percent for vertebrates, and 29.5 percent for fungi).

We also count endemic species by country and species group. Country scale plays a major role in raw counts, so we standardize by total country species to highlight the relative importance of endemicity in each country and species group. Table 2 provides a summary for the top 30 countries in each species group, sorted in descending order by average ranking for the seven groups. Overall, the top 30 have 86.6 percent (235,706 out of 272,189 species); and our results assign overall top

10 status to Australia, United States, Brazil, Mexico, South Africa, China, New Zealand, Madagascar, Japan, and Costa Rica. Even for the top 30 countries, endemicity varies enormously by species group. In terms of vertebrates, for example, 62 percent are endemic in Australia versus only 3 percent in the United Kingdom. For plants, the endemicity in Madagascar, Australia, and New Zealand is extremely high, at 89.6 percent, 88.1 percent, and 84.4 percent, respectively. For arthropods, the maximum endemicity is even higher in New Zealand and Australia, at 92.8 percent and 91.2 percent, respectively. Mollusks, fungi, and other species more closely resemble arthropods and vascular plants in the relative compactness of their ranges.

Table 2. Top 30 countries for species endemism, by group

Country	Endemic species (%)						
	Vertebrates	Arthropods	Mollusks	Other animals	Vascular plants	Fungi	Other species
Australia	62.3	91.2	76.5	71.5	88.1	59.8	57.3
United States	41.3	41.7	57.5	49.5	48.4	27.9	26.3
Brazil	38.5	46.1	70	67.9	51.6	30.1	45.5
Mexico	45.1	47.7	41.4	48	58.1	44.1	40.7
South Africa	37	41.5	81.9	71	67.2	46.1	36.5
China	29.6	56.5	58.1	55.9	34	38.9	16.3
New Zealand	57.5	92.8	89.6	66.2	84.4	81.8	58.9
Madagascar	77.4	85.3	81.6	27.5	89.6	25.7	11.8
Japan	41	55.5	67	65.8	59.2	66.8	28.1
Costa Rica	33.7	48.6	50.2	63.8	42	43.2	35.7
Colombia	37.7	32.7	49.7	40	32.9	35.7	73.7
France	19.9	22	50.4	33.1	25.2	6	7.1
Spain	24.8	39.6	47.8	40.5	43.9	29.4	27.7
New Caledonia	50.8	64.1	54.2	49.1	94.7	76.9	55.6
Ecuador	51.9	60.5	65.6	72.4	46.8	56	80
Papua New Guinea	40.7	53.3	37.4	24.4	61.4	48.8	6.2
Indonesia	29	38.9	13.6	13.9	24.4	10.9	17.7
Peru	32.5	35.3	53.4	44.4	36	39.5	48.3
Canada	7.1	29.7	17.4	34.5	11.3	25.4	19.7
India	39.8	38.7	43.8	66.1	49.4	34.1	28.5
Chile	44.6	62.3	45.5	47.1	50.3	44.8	44.3
Russian Federation	14.1	15.5	27.5	48.5	22.2	12.3	51.8
Philippines	66.7	72.7	57	37.4	57.9	0	15.4
Sweden	13	10.1	10	51	42.1	5.1	23.4
United Kingdom	3	23.1	5.2	11.1	39.7	29.3	21.9
Argentina	26.7	40.5	28.9	56.5	32.4	28	10
Cuba	31.2	32	10	0	70.4	4.5	6.9
Sri Lanka	95.9	88.9	63.2	100	89.3	72.7	88.1
Malaysia	34.2	41.1	61.3	44.4	25.3	51.9	80
Bolivia	27.4	53.6	50	28.6	50.3	46.4	50

It should be noted that Table 2 excludes small island territories that rank high in at least one group, including South Georgia, French Polynesia, Heard and McDonald Islands, Norfolk Island, and the Malvinas/Falklands disputed territory.

4.2 Distribution of small-region species

Small range size has been studied extensively in the empirical literature (Jenkins et al., 2015; Kraus et al., 2023; Manne, Brooks, and Pimm, 1999; Manne and Pimm, 2001; Purvis et al., 2000; Veach et al., 2017). Jenkins et al. (2015), for example, note that "small range size is the best predictor of extinction risk and, thus, the first metric for conservation priority." It has particular significance since it is a widely recognized indicator of extinction risk that is computable for any species that can be mapped.

However, it should be noted that small-range status is not determinate; there is no single, critical minimum habitat size, given the myriad interactions between species and habitat characteristics that affect extinction risks. Therefore, we examined the size and global distribution of species with small occurrence regions in our GBIF maps database, considering the effects of changing the criteria for small-occurrence-region status. Table 3 displays the cumulative global count for species groups as the occurrence region increases from 5 km x 5 km to 200 km x 200 km. Even for occurrence regions of 10 km x 10 km or less, 57,765 species are identified; this number increases to 85,310 at 25 km x 25 km or less. Differences across species groups reflect their varying representation in the database and group-specific factors.

Table 3. Species counts by group and grid scale

Occurrence region category (km)	Vertebrates	Arthropods	Mollusks	Other animals	Vascular plants	Fungi	Other species
5 x 5	3,029	17,587	3,336	2,843	12,908	3,410	2,046
10 x 10	3,897	22,245	4,502	3,611	17,234	3,921	2,355
20 x 20	5,385	29,016	6,166	4,575	24,611	4,674	2,948
25 x 25	6,020	31,734	6,748	4,931	27,785	4,936	3,156
50 x 50	8,580	42,894	8,976	6,214	41,285	6,125	3,872
100 x 100	12,215	60,914	12,169	8,149	63,173	8,248	5,213
200 x 200	17,522	88,204	16,425	10,927	94,036	11,755	7,303

We believe that an upper bound of 25 km x 25 km on critical scale for small-range species is appropriately conservative. The small-range species count increases to 117,946 at 50 km x 50 km or less, 170,081 at 100 km x 100 km or less, and 246,172 at 200 km x 200 km. From a policy perspective, the feasibility and sustainability of species protection tend to decline as the number of species protected increases. Since even the 25 km x 25 km limit qualifies nearly 85,310 species as having a small occurrence region, we retain it here, recognizing that other analyses may well opt for higher limits.

Using GIS overlays of GBIF maps and country boundaries, we count species with small occurrence regions by country, finding that their international distribution is skewed. The top 30 countries account for 75.5 percent of them (64,443 out of 85,310 species). Our overall results assign top 10 status to Australia, United States, Brazil, Mexico, France, South Africa, Costa Rica, China, Colombia, and Japan. Australia leads with 8,673 species, followed by the United States (7,791), Brazil (4,434), Mexico (4,217), and France (3,732). Comparing Table 3 with Table 4 suggests that small-occurrence-region species are endemic in most cases, so the dominant country is chosen by default. In other cases (e.g., Panama, Venezuela, RB, Thailand, and Italy), it is the country with greatest area share in the species' GBIF occurrence map. Among species groups, the top 30 countries' global share varies from 66 percent (vertebrates) to 78 percent (arthropods) (Table 4).

Table 4. Top 30 countries for species with small occurrence regions, by group

Country	Species (%)						
	Vertebrates	**Arthropods**	**Mollusks**	**Other animals**	**Vascular plants**	**Fungi**	**Other species**
Australia	6.5	11.7	13.8	15.9	7.6	8.8	10.4
United States	7	11	13.1	14.1	5.2	12.2	7.2
Brazil	7.8	3.2	1.4	2.4	8.8	4.1	2.2
Mexico	5.3	4.2	3.6	5.7	6.2	4	3.8
France	0.6	6.7	6	4.3	1.9	6.9	2.9
South Africa	2	2.2	1.8	2.9	7.5	0.9	0.8
Costa Rica	1.1	6.9	1.1	1.1	1.6	2.2	0.1
China	2.5	3.1	1.6	1	3.2	2.1	3.3
Colombia	3.9	1.8	0.5	1.1	3.2	5	6.9
Madagascar	2.5	1.1	5.9	0.2	4.4	0.2	0.1
Japan	1.4	3.1	3.9	2	1.3	2.2	1.5
Spain	0.4	2.9	2.1	1.4	2.2	2	0.5
Canada	0.2	3.9	0.5	2	0.3	3.6	1.1
New Zealand	0.6	2.1	1.9	4.2	1	4.9	1.9
Indonesia	5.5	1.2	1.9	1.3	1.6	0.5	1.6
Russian Federation	0.5	1.7	0.3	1.2	1.6	2.9	5.4
Ecuador	2.1	1.2	0.3	0.4	2.9	0.5	0.3
New Caledonia	0.7	0.9	2.8	2.4	2.1	0.1	0.1
Papua New Guinea	1.7	0.8	2.2	0.6	2	0.3	0.1
Sweden	0	1.1	0	2.1	0.7	3.2	9.9
Peru	2.1	0.6	0.1	0.1	2.2	0.2	0.5
India	2.6	1.1	0.3	0.4	0.9	0.8	2.4
Malaysia	1.2	1.2	0.8	0.2	1.2	0.2	0
Panama	1.3	0.5	2	1	1.3	0.2	0.1
Chile	1.3	0.7	0.7	0.9	1.1	0.8	0.6
United Kingdom	0.1	0.7	0.5	1.9	0.7	3	1.8
Philippines	2.6	0.4	4	1	0.3	0.1	0.1
Venezuela, RB	0.8	0.3	0.1	0.4	1.7	0.6	0.5
Italy	0.3	0.8	0.5	1.5	0.8	0.7	1.7
Thailand	1.3	0.9	0.5	0.1	0.6	0.6	1
Total	**65.9**	**78**	**74.2**	**73.8**	**76.1**	**73.8**	**68.8**

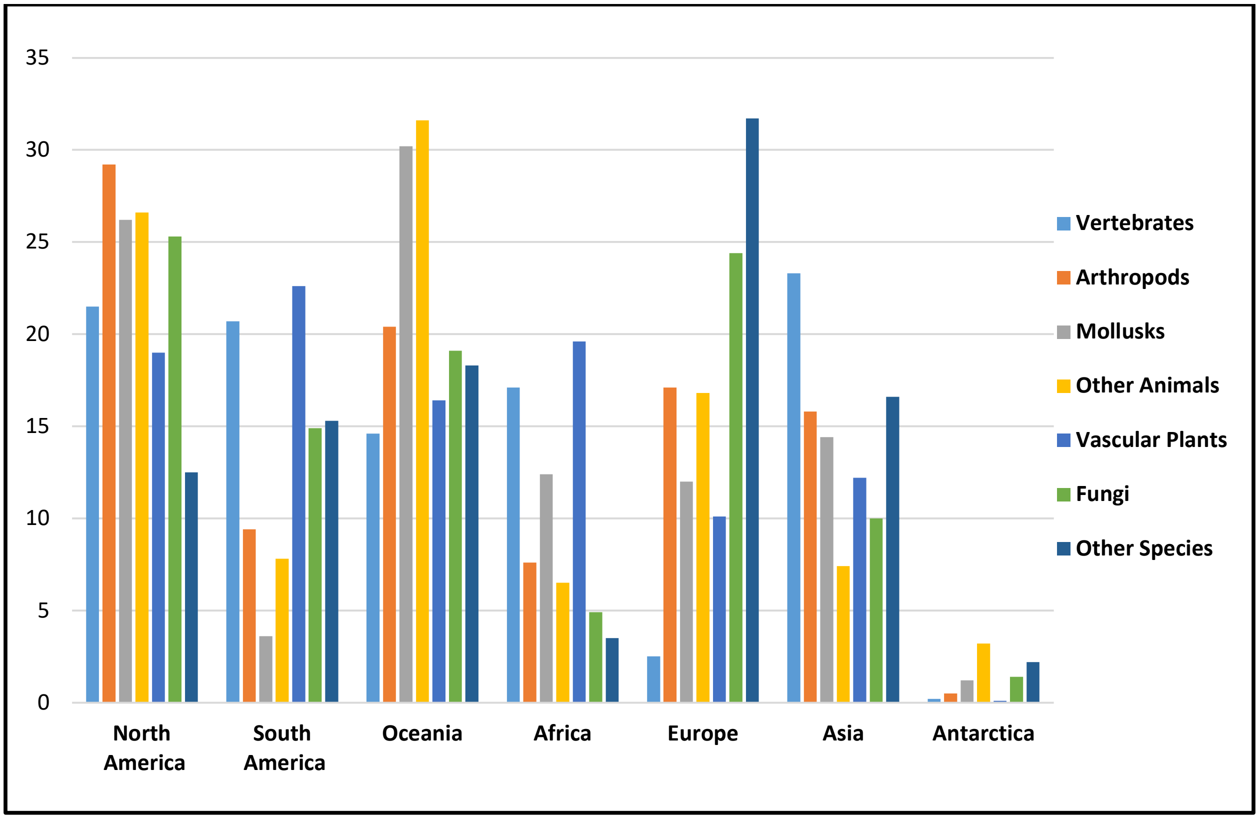
We also explored the geographical distribution of endemic species with small occurrence regions (25 km x 25 km size limit). Our results identified 67,941 species in a single country (Table 5).

Table 5. Top 30 countries for endemic species with small occurrence regions, by group

	Species (%)						
Country	**Vertebrates**	**Arthropods**	**Mollusks**	**Other animals**	**Vascular plants**	**Fungi**	**Other species**
Australia	6.7	14.1	15.2	17.8	8.7	10.8	12.7
United States	7.1	10.6	13.5	14.3	5.1	11	6.2
Brazil	7.9	3.3	1.5	2.7	8.9	3.4	2.6
Mexico	5.7	4.6	3.5	6	6.5	4.8	4.9
South Africa	2	2.3	2	3.3	8.3	1	1.2
Costa Rica	1.1	7.6	1.2	1.2	1.6	2.8	0.1
France	0.5	6	5.6	4.2	1.4	4.1	2
Madagascar	3	1.3	6.8	0.1	5.1	0.1	0
Colombia	3.9	1.7	0.5	1.1	3.1	6.7	10.2
Japan	1.4	3.4	4.1	2.1	1.3	3.2	1.4
China	2	3.3	1.6	0.9	2.3	2.1	1.5
Spain	0.3	3.1	2.1	1.4	2.2	2	0.4
New Zealand	0.7	2.5	2.3	4.6	1.1	7.2	2.3
Ecuador	2.4	1.4	0.3	0.5	2.9	0.6	0.4
Canada	0.1	3.7	0.4	1.7	0.2	4.3	0.9
New Caledonia	0.8	1.1	3	2.5	2.5	0.1	0.1
Indonesia	4.9	1.3	0.9	0.8	1.5	0.2	0.6
Papua New Guinea	1.7	0.9	2	0.4	2.1	0.3	0
Russian Federation	0.4	1.3	0.2	1.2	1.2	2.4	7.5
Peru	2.2	0.7	0.2	0.1	2.3	0.2	0.6
Malaysia	1.4	1.3	0.9	0.2	1.1	0.3	0
India	2.5	1.1	0.3	0.5	0.8	0.8	2.2
Chile	1.3	0.8	0.7	1	1.2	1	0.7
Sweden	0	0.6	0	2.4	0.7	1.2	10.5
Panama	1.1	0.5	2	1	1.3	0.2	0.1
Philippines	2.9	0.5	4	0.9	0.4	0	0.1
Venezuela, RB	0.7	0.3	0.1	0.4	1.7	0.2	0.5
Cuba	0.6	0.2	0.4	0	1.9	0	0
French Polynesia	0.9	0.6	1.9	1	0.8	0	0.4
Portugal	0.1	1.1	1.1	1	0.4	0.8	1.5
Total	**66.3**	**81.2**	**78.3**	**75.3**	**78.6**	**71.8**	**71.6**

As before, we find that the international distribution is skewed, with 78.2 percent (53,114) of the 67,941 species found in 30 countries. The overall results assign top 10 status to Australia, United States, Brazil, Mexico, South Africa, Costa Rica, France, Madagascar, Colombia, and Japan. Australia leads with by 8,072 species, followed by the United States (6,003), Brazil (3,629), Mexico (3,621) and South Africa (2,911). Among species groups, the top 30 countries have the following global shares: arthropods (81.2 percent), vascular plants (78.6 percent), mollusks (78.3 percent), other animals (75.3 percent), fungi (71.8 percent), other non-animal and non-plant species (71.6 percent), and vertebrates (66.3 percent).

Figure 6. Regional distribution (percentage) of endemic species with small occurrence regions



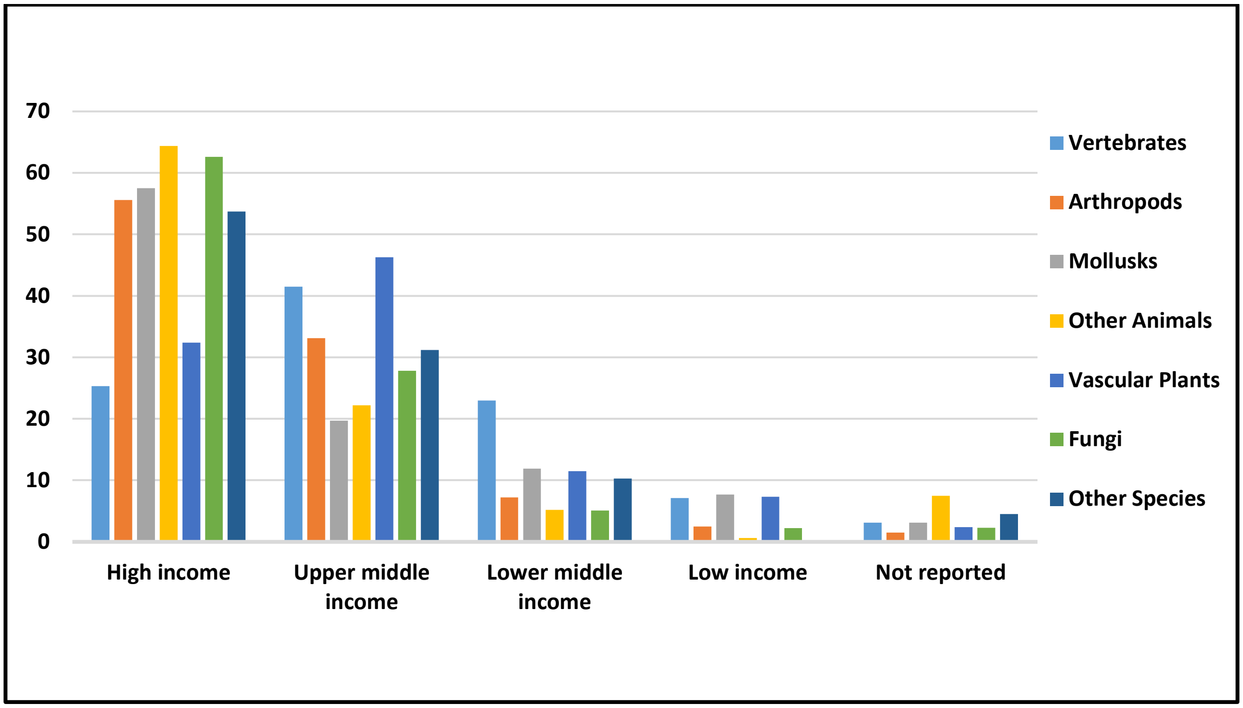
Among endemic species with small occurrence regions, the largest share is found in Oceania (28 percent), followed by North America (23 percent). Four regions are in the mid-range—South America (15 percent), Asia (13 percent), Africa (11 percent), and Europe (9 percent)—and Antarctica has small representation, at 1 percent (Fig. 6).

4.4 Candidate hotspot areas for protection

Limited resources for biodiversity conservation make it critical to prioritize protection efforts in regions inhabited by many unique, at-risk species. Endemism and small occurrence regions, as identified by our maps, can inform conservation policy priority-setting. This study’s findings indicate that 40 countries have significant opportunities for protecting areas with concentrations of endemic species, species with small occurrence regions, and species with both features.

Aligning countries with World Bank income groups reveals an encouraging trend for conservation. While over 4,000 endemic species with small ranges are in low and lower-middle-income countries, the majority, 82.7 percent, are in high and upper-middle-income countries (Fig. 7), which generally have substantial conservation resources. Many such areas may already be protected. Although a global assessment was beyond this study's scope, it would be a valuable future application of our GBIF species maps database.

Figure 7. Percent distribution of endemic species with small occurrence regions, by income class



It should be noted that the maps constructed with processed data also provide opportunities for understanding the geographic distribution of the species within countries.

5. Discussion

Transparent and accessible biodiversity data remain central to informed decision-making for conservation and to safeguarding a livable planet for current and future generations. However, comprehensive and all-taxa-inclusive biodiversity data have long been scarce, limiting the ability to assess species distribution and conservation priorities at a global scale. This study presents the first near-comprehensive global mapping of terrestrial, freshwater, and marine biodiversity derived from GBIF occurrence data. By applying the *alphahull* algorithm and complementary clustering techniques to over 600,000 species, the analysis demonstrates that high-resolution, occurrence-based approaches provide a robust complement to expert-drawn range maps. Building on these results and given the numerous competing demands on land and water, the spatial precision of conservation activities and policies can be enhanced through *alphahull* techniques, thereby increasing their potential for success by effectively focusing the area of consideration. Also, our effort aligns with that of Janicki et al. (2016), and a salient contribution of this work is the inclusion of previously unmapped taxa—particularly invertebrates, fungi, and plants—which substantially broadens the taxonomic and spatial scope of global biodiversity representation.

5.1 Expanding the biodiversity evidence base

The observed alignment between occurrence-based and expert-derived maps underscores the potential of crowdsourced georeferenced data for large-scale biodiversity mapping. Discrepancies between the two primarily reflect temporal and ecological scope. Expert maps typically represent potential distributions based on habitat suitability and species ecology, while occurrence-based maps capture realized distributions shaped by recent environmental and anthropogenic dynamics (Meyer et al., 2016). Consequently, GBIF-derived boundaries can more effectively detect recent range contractions or expansions associated with deforestation, climate change, or land-use conversion. The rapid updating of GBIF records—currently exceeding one million new georeferenced occurrences daily—creates opportunities for near-real-time monitoring of biodiversity change, a capability unavailable in static expert compilations.

5.2 Addressing taxonomic and geographic bias

A major contribution of this work lies in filling taxonomic gaps that have historically constrained global biodiversity assessments. Existing global datasets emphasize vertebrates, overlooking invertebrates and fungi that together comprise the majority of known species (Cardoso et al., 2011). The integration of over 213,000 arthropods, 38,000 fungi, and 24,000 other animal taxa demonstrates the feasibility of incorporating these groups systematically into global-scale analyses. While GBIF data remain spatially biased toward accessible and high-income regions (Yesson et al., 2007), ongoing growth in citizen science platforms and institutional data sharing is steadily improving geographic representativeness. As observation density increases, particularly across tropical and remote regions, boundary estimation will gain precision, enabling more accurate identification of small-range and data-poor species.

5.3 Implications for conservation priority setting

The identification of endemic and small-range species from GBIF-derived maps provides new insights for conservation targeting. A broader view of biodiversity provides an opportunity for more countries to contribute to a shared vision of conservation stewardship, as most nations host significant distributions of species within at least one major taxonomic group. Consistent with prior studies (Kier et al., 2009), this analysis confirms that small-range species are disproportionately at risk of extinction due to limited dispersal capacity and habitat specialization. The correlation between endemism and small-range occurrence suggests that national-level conservation actions will be critical for achieving global biodiversity outcomes. Australia, the United States, Brazil, Mexico, and South Africa emerge as global centers of endemism and small-occurrence-region species, underscoring their unique biodiversity assets. The identification of numerous localized hotspots—such as Madagascar, Costa Rica, and Southeast Asia—further highlights regions where protection gap analyses are most urgent.

5.4 Policy relevance and the Global Biodiversity Framework

Implementing the goal of protecting 30 percent of the planet's land and sea areas by 2030, as set out in Target 3 of the 2022 Kunming–Montreal Global Biodiversity Framework (GBF), requires precise identification of areas critical for conservation. Using occurrence data for over 600,000 species from GBIF, this study provides the largest set of species distribution maps derived from open-access data and highlights 30 countries that could play central roles in area-based conservation, capturing 86.6 percent of endemic species, 75.5 percent of small-occurrence-region species, and 78.2 percent of species that are both endemic and range-limited. These data can also be used to examine the overlap of existing protected areas with endemic-species distributions, highlighting significant variation in initial conservation conditions, including current protection levels and the spatial distribution of unprotected species (Dasgupta et al., 2025).

The GBIF database, expanding by approximately 1.3 million new records daily, allows continuous updates for previously unmapped species and refinement of existing maps. The estimation algorithm supports this growth, providing area estimates for all species and complementing traditional risk indicators. In addition, the alpha hull approach offers a parsimonious, occurrence-driven estimate of species range, requiring minimal assumptions and no environmental predictors. Our practical experience in client countries further indicates that communicating with policymakers is often more straightforward when using species occurrence regions derived from reported occurrences, compared to statistical modeling methods that infer occurrence from environmental predictors, which can be challenging to explain. Given the rate of biodiversity loss, it is urgent to initiate the biodiversity dialogue immediately, beginning with approaches to mapping that are readily understandable. The open-access nature of the data, algorithms, and maps aligns with the GBF's emphasis on equitable sharing, transparency, and evidence-based policy design. High-resolution, taxonomically inclusive maps can guide national and transboundary prioritization, while continuous updating allows adaptive policy responses—an essential feature for tracking progress toward GBF targets and ensuring accountability in biodiversity monitoring frameworks. This need for spatially explicit and scalable information is particularly relevant for international institutions and development organizations. An information paradox often confronts decision-makers in transnational institutions: while the scope of their operations is global, the scale of specific interventions is inevitably local or regional (Blankespoor et al. 2023).

Looking ahead, open and equitable sharing of biodiversity, threat, and protection data across administrative boundaries remains essential, as species and ecosystems frequently span political jurisdictions (Blankespoor et al., 2025; Tanalgo, 2025). Reliable, high-resolution, and up-to-date data facilitate coordinated conservation strategies across terrestrial, freshwater, and marine systems. Healthy biodiversity underpins ecosystem stability and sustains essential services such as agriculture, fisheries, water security, and climate resilience. It is therefore fundamental to sustainable development, poverty alleviation, equitable prosperity, and safeguarding a livable planet for current and future generations.

5.5 Limitations and future directions

Despite these advances, several limitations warrant acknowledgment. First, GBIF occurrence data remain uneven in spatial and temporal coverage, with collection bias favoring regions with established research infrastructure (Beck et al., 2022; Boakes et al., 2010). Second, presence-only data do not capture species abundance, which limits inference about population viability within mapped boundaries. Third, while the *alphahull* method provides an efficient means of delineating occurrence extents, it does not explicitly account for environmental suitability or ecological connectivity. Future research could enhance these methods by integrating species distribution modeling, remote-sensing covariates, and temporal updating to capture dynamic species–environment interactions (Elith & Leathwick, 2009).

6. Concluding Remarks

Overall, this study demonstrates the feasibility and scientific value of automated, open-access biodiversity mapping at global scale. By complementing expert-based datasets with data-driven, continuously updatable occurrence maps, it bridges long-standing taxonomic and geographic data gaps, improves spatial precision for conservation planning, and enhances the policy relevance of biodiversity information systems. As global data sharing and digital observation platforms continue to expand, the approach outlined here can provide a scalable and transparent foundation for evidence-based implementation of the Global Biodiversity Framework and related multilateral environmental agreements.

7. Code and Data availability

These data are available at the World Bank's Development Data Hub under Global Biodiversity Species Occurrence Gridded Data and Global Biodiversity Species Occurrence Endemism and Small Range Data. The datasets can be accessed at https://doi.org/10.57966/h21e-vq42 (Dasgupta et al. 2024). The authors utilized Google BigQuery to retrieve species occurrence records and R software for data processing. The associated scripts are available from the authors upon request.

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Competing interests: The authors confirm that there is no financial or personal relationship that could cause a conflict of interest or even influence and/or bias their judgment regarding this article.

Funding: This research was funded by the Global Environment Facility.

Author contributions

Susmita Dasgupta: Conceptualization, Funding Acquisition, Project Supervision, and Writing original draft.

Brian Blankespoor: Funding Acquisition; GIS Analysis; Visualization; and Writing, Review, and Editing.

David Wheeler: Conceptualization, Data Curation, Methodology, and Formal Analysis.

Acknowledgements

This research was funded by a grant from the Global Environment Facility to a World Bank program managed by the authors with Dr. Nagaraja Harshadeep Rao. We are thankful to the World Bank Sustainable Development Global Practice and Environment, Natural Resources and Blue Economy Global Practice for their review and comments. We are also grateful to Polly Means for the graphics and Norma Adams for editorial support.

The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

References

Blankespoor B, Dasgupta S, Wheeler D. Bridging Conflicts and Biodiversity Protection: The Critical Role of Reliable and Comparable Data. World Bank Policy Research Working Paper 11076; 2025 February. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099415202272518792

Blankespoor, B., Dasgupta, S., Wheeler, D., Jeuken, A., van Ginkel, K., Hill, K., and Hirschfeld, D.: Linking sea-level research with local planning and adaptation needs, Nature Climate Change, 13(8), 760–763, 2023. https://doi.org/10.1038/s41558-023-01789-8

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W.: Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics, 19*, 10–15. 2014. https://doi.org/10.1016/j.ecoinf.2013.11.002

Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M.: Distorted views of biodiversity: spatial and temporal bias in species occurrence data. *PLoS Biology, 8*(6), e1000385. 2010. https://doi.org/10.1371/journal.pbio.1000385

Borgelt, J., Sicacha-Parada, J., Skarpaas, O., et al.: Native range estimates for red-listed vascular plants, Nature Scientific Data, 9, 117, 2022. https://doi.org/10.1038/s41597-022-01233-5

Burgin, C., Colella, J., Kahn, P., and Upham, N.: How many species of mammals are there?, Journal of Mammalogy, 99, 1–14, 2018. https://doi.org/10.1093/jmammal/gyx147

Burgin, C., Wilson, D., Mittermeier, R., Rylands, A., Lacher, T., and Sechrest, W.: Illustrated Checklist of the Mammals of the World, Lynx Nature Books, 2020a. ISBN 10: 8416728364 ISBN 13: 9788416728367

Burgin, C., Wilson, D., Mittermeier, R., Rylands, A., Lacher, T., and Sechrest, W.: Illustrated Checklist of the Mammals of the World (Vol. 2, Eulipotyphla to Carnivora), Lynx Edicions, 2020b. ISBN: 978-84-16728-36-7

Burgin, C., Wilson, D., Mittermeier, R., Rylands, A., Lacher, T., and Sechrest, W.: Illustrated Checklist of the Mammals of the World (Vol. 1, Monotremata to Rodentia), Lynx Edicions, 2020c. ISBN: 978-84-16728-36-7

Cardoso, P., Erwin, T. L., Borges, P. A. V., & New, T. R.: The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation, 144*(11), 2647–2655. 2011. https://doi.org/10.1016/j.biocon.2011.07.024

Clark, J. and May, R.: Taxonomic bias in conservation research, Science, 297, 191–192, 2002. https://www.science.org/doi/10.1126/science.297.5579.191b

Dasgupta, S., Blankespoor, B., and Wheeler D.: Global Biodiversity Data, 2024. https://doi.org/10.57966/h21e-vq42

Dasgupta, S., Wheeler, D., & Blankespoor, B.: Pathways to 30×30: Evidence-Based Lessons from Global Case Studies in Biodiversity Conservation. Diversity, 17(6), 401. 2025. https://doi.org/10.3390/d17060401

Elith, J., & Leathwick, J. R.: Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics, 40*, 677–697. 2009. 2009. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Feeley, K. and Silman, M.: Keep collecting: accurate species distribution modelling requires more collections than previously thought, Diversity and Distributions, 17, 1132–1140, 2011. https://doi.org/10.1111/j.1472-4642.2011.00813.x

Garcia-Rosello, E., Gonzalez-Dacosta, J., Guisande, C., and Lobo, J.: GBIF falls short of providing a representative picture of the global distribution of insects, Systematic Entomology, 48, 489-497, 2023. https://doi.org/10.1111/syen.12589

GBIF.org: GBIF Occurrence Data accessed via Google BigQuery (https://console.cloud.google.com/bigquery) on 17 February 2024.

Guo, W., Serra-Diaz, J., Schrodt F., et al.: High exposure of global tree diversity to human pressure, Proceedings of the National Academy of Sciences, 119, 2022. https://doi.org/10.1073/pnas.202673311

Hughes, A. C., Dorey, J. B., Bossert, S., Qiao, H., & Orr, M. C.: Big data, big problems? How to circumvent problems in biodiversity mapping and ensure meaningful results. *Ecography*, *2024*(8), e07115, 2024. https://doi.org/10.1111/ecog.07115

Isaac, N. and Pocock, M.: Bias and information in biological records, Biological Journal of the Linnean Society, 115, 522–531, 2015. https://doi.org/10.1111/bij.12532

IUCN (International Union for Conservation of Nature): The IUCN Red List of Threatened Species, https://www.iucnredlist.org , 2021.

Janicki, J., Narula, N., Ziegler, M., Guénard, B., & Economo, E. P.: Visualizing and interacting with large-volume biodiversity data using client–server web-mapping applications: The design and implementation of antmaps.org. *Ecological Informatics, 32*, 185–193. 2016. https://doi.org/10.1016/j.ecoinf.2016.02.006

Jenkins, C., Van Houtan, K., Pimm, S., and Sexton, J.: US protected lands mismatch biodiversity priorities, Proceedings of the National Academy of Sciences, 112, 5081-5086, 2015. https://doi.org/10.1073/pnas.1418034112

Kass, J., Guénard, B., Dudley, K., et al.: The global distribution of known and undiscovered ant biodiversity, Science Advances, 8, 2022. https://www.science.org/doi/10.1126/sciadv.abp9908

Kier, G., Kreft, H., Lee, T. M., Jetz, W., Ibisch, P. L., Nowicki, C., Mutke, J., & Barthlott, W.: A global assessment of endemism and species richness across island and mainland regions. *Proceedings of the National Academy of Sciences, 106*(23), 9322–9327. 2009. https://doi.org/10.1073/pnas.08103061

Kraus, D., Enns, A., Hebb, A., et al.: Prioritizing nationally endemic species for conservation, Conservation Science and Practice, 5, 2023. https://doi.org/10.1111/csp2.12845

Leather, S., Basset, Y., and Hawkins, B.: Insect conservation: finding the way forward, Insect Conservation and Diversity, 1, 67–69, 2008. https://doi.org/10.1111/j.1752-4598.2007.00005.x

Manne, L. and Pimm, S.: Beyond eight forms of rarity: which species are threatened and which will be next, Animal Conservation, 4, 221–229, 2001. https://doi.org/10.1017/S1367943001001263

Manne, L., Brooks, T., and Pimm, S.: Relative risk of extinction of passerine birds on continents and islands, Nature, 399, 258–261, 1999. https://doi.org/10.1038/20436

Marsh, C., et al.: Expert range maps of global mammal distributions harmonized to three taxonomic authorities, Journal of Biogeography, 49, 979-992, 2022. https://doi.org/10.1111/jbi.14330

Meyer, C., Weigelt, P., & Kreft, H.: Multidimensional biases, gaps and uncertainties in global plant occurrence information. *Ecology Letters, 19*(8), 992–1006. 2016. https://doi.org/10.1111/ele.12624

Mittermeier, R., Rylands, A., and Wilson, D.: Handbook of the mammals of the world (Vol. 3: Primates), Lynx Edicions, 2013.

Pateiro-López, B. and Rodríguez-Casal, A.: Generalizing the convex hull of a sample: the R package alphahull, Journal of Statistical Software, 34, 2010. DOI: 10.18637/jss.v034.i05

Pimm, S., Jenkins, C., Abell, R., Brooks, T., Gittleman, J., Joppa, L., Raven, P., Roberts, C., and Sexton, J.: The biodiversity of species and their rates of extinction, distribution, and protection, Science, 344, 2014. DOI: 10.1126/science.1246752

Purvis, A., Gittleman, J., Cowlishaw, G., and Mace, G.: Predicting extinction risk in declining species, Proceedings of the Royal Society, Biological Sciences, 267, 1947–1952, 2000. https://doi.org/10.1098/rspb.2000.1234

Reddy, S. and Dávalos, L.: Geographical sampling bias and its implications for conservation priorities in Africa, Journal of Biogeography, 30, 1719–1727, 2003. https://doi.org/10.1046/j.1365-2699.2003.00946.x

Tanalgo, K. C. (2025). Open and FAIR data sharing are building blocks to bolster biodiversity conservation in Southeast Asia. *Biological Conservation, 307,* 111192. https://doi.org/10.1016/j.biocon.2025.111192

Veach, V., Di Minin, E., Pouzols, F., and Moilanen, A.: Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity, Diversity and Distributions, 23, 715–726, 2017. https://doi.org/10.1111/ddi.12571

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 1: Carnivores), Lynx Edicions, 2009. ISBN: 978-84-96553-49-1

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 2: Hoofed Mammals), Lynx Edicions, 2011.ISBN: 978-84-96553-77-4

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 4: Sea Mammals), Lynx Edicions, 2014. https://doi.org/10.1093/jmammal/gyv071

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 5: Monotremes and Marsupials), Lynx Edicions, 2015. https://doi.org/10.1093/jmammal/gyw012

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 8: Insectivores, Sloths and Colugos), Lynx Edicions, 2018. ISBN: 978-84-16728-08-4

Wilson, D. and Mittermeier, R.: Handbook of the mammals of the world (Vol. 9: Bats), Lynx Edicions, 2019. ISBN: 978-8416728194

Wilson, D., Lacher, T., and Mittermeier, R.: Handbook of the mammals of the world (Vol. 6: Lagomorphs and Rodents), Lynx Edicions, 2016.

Wilson, D., Lacher, T., and Mittermeier, R.: Handbook of the mammals of the world (Vol. 7: Rodents II). Lynx Edicions, 2017. ISBN: 978-8416728046

Yesson, C., Brewer, P. W., Sutton, T., Caithness, N., Pahwa, J. S., Burgess, M., Gray, W. A., White, R. J., Jones, A. C., & Bisby, F. A.: How global is the Global Biodiversity Information Facility? *PLoS ONE, 2*(11), e1124. 2007. https://doi.org/10.1371/journal.pone.0001124