1 **Toward Better Conservation: A Spatial Analysis of Species Occurrence Data**

2 **from the Global Biodiversity Information Facility**

3

4 Susmita Dasgupta

5 Lead Environmental Economist, Development Economics Research Group,

6 MC- 3-347, World Bank,1818 H Street, Washington DC 20433, USA

7 sdasgupta@worldbank.org

8 Brian Blankespoor

9 Senior Geographer, Development Economics Data Group, MC-9-204,

10 World Bank,1818 H Street, Washington DC 20433, USA bblankespoor@worldbank.org

11 and

12 David Wheeler

13 Consultant, Development Economics Research Group,

14 World Bank,1818 H Street, Washington DC 20433, USA

15 wheelrdr@gmail.com

16

17 *Corresponding author:* Brian Blankespoor, Senior Geographer, Development Economics Data
18 Group, MC-9-204, The World Bank, 1818 H Street, Washington, DC 20008, USA. Tel: 1-202-
19 473-1546. Email: bblankespoor@worldbank.org

20

21

22

23

Abstract

24

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

42

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

45

46 *500-character non-technical summary including space*

47

Short Summary

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

55

56

57

58

59

60

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.

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

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

113 2. Data and methods

114 *2.1 Data source and tools*

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

127 In this study, we use BigQuery to download the full GBIF database due to its convenient dataset
128 size reduction tools (GBIF 2024). We limit the GBIF occurrence data to geolocated reports since
129 1970 for species with at least five unique reporting locations, which our mapping algorithm
130 requires. Since our mapping algorithm operates reliably on several thousand points at most, we
131 cap the data at a maximum of 20,000 randomly selected reports per species to ensure reliability.
132 This limit drastically reduces the size of the download dataset since some species have millions of
133 occurrence records (e.g., the American robin [*Turdus migratorius*] currently has 21,258,907
134 reported occurrences). We accept the GBIF's protocols for occurrence report admissibility.
135 Detailed descriptions of the GBIF's protocols and database elements can be found at
136 https://www.gbif.org/data-quality-requirements-occurrences.

137

138 *2.2 Analysis*

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

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

166 In our study, alphahull successfully estimates occurrence maps for 92.9 percent (567,464) of the
167 610,694 species in our database. For each of the remaining 7.1 percent of species in the database,
168 we employ a standard k-means algorithm to separate occurrence reports into spatial clusters and
169 draw a convex hull around each. We assess the utility of these mapping algorithms for delineating
170 species occurrence regions, with illustrative examples provided in the Annex. Our algorithms
171 estimate occurrence maps for terrestrial, coastal, and marine species.

172
173 *2.3 Spatial selection bias*

174 We acknowledge that GBIF occurrence reports are often produced by voluntary exercises that do
175 not utilize scientific sampling methods. As a result, spatial point densities in species occurrence
176 reports are positively related to physical accessibility, population density, and income (Borgelt et
177 al., 2022; Garcia-Rosello et al., 2023; Isaac and Pocock, 2015; Reddy and Dávalos, 2003). This
178 means that (all else being equal) species sightings are more likely to occur in areas (1) near

179 transport arteries; (2) with a greater number of inhabitants to identify species, and (3) where more
180 inhabitants have enough disposable income to support species search and reporting costs.

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

3. Results

3.1 Pilot example

192 To illustrate the results of our mapping algorithm exercise, we take the example of the species
193 *Lagidium viscacia* (common name: Mountain Viscacha), whose range extends from areas in
194 Argentina and Chile to Bolivia and Peru. In Fig. 1, a comparison of panels (a) and (b) shows that
195 the reported sightings of this species include a few points beyond the northern boundary of the
196 expert range map (Burgin et al., 2020a), along with many points beyond the southern boundary.
197 Panel (c), which displays the output of our mapping algorithm, shows that the alphahull boundary
198 follows the curvilinear north-south orientation of the point set, widening and narrowing as the
199 point set expands and contracts. As shown, it overlaps heavily with the map of Burgin et al.
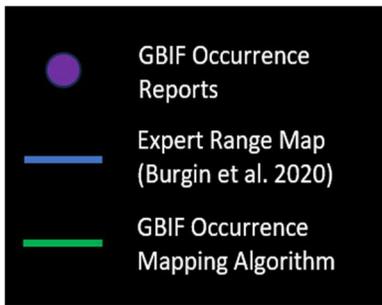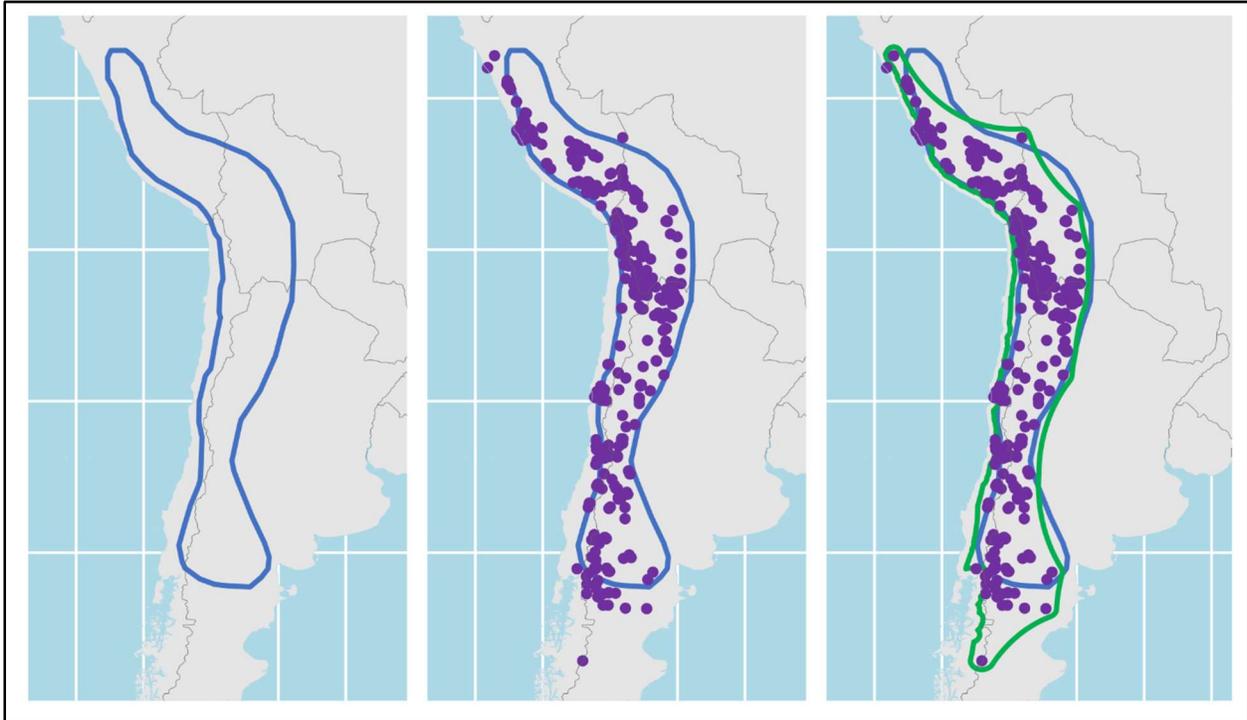200 (2020a), but extends its northern and southern boundary areas to incorporate these sightings.

201

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

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**

8

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,

257   western United States, and eastern Australia and lower ranks to the southeastern Amazon region
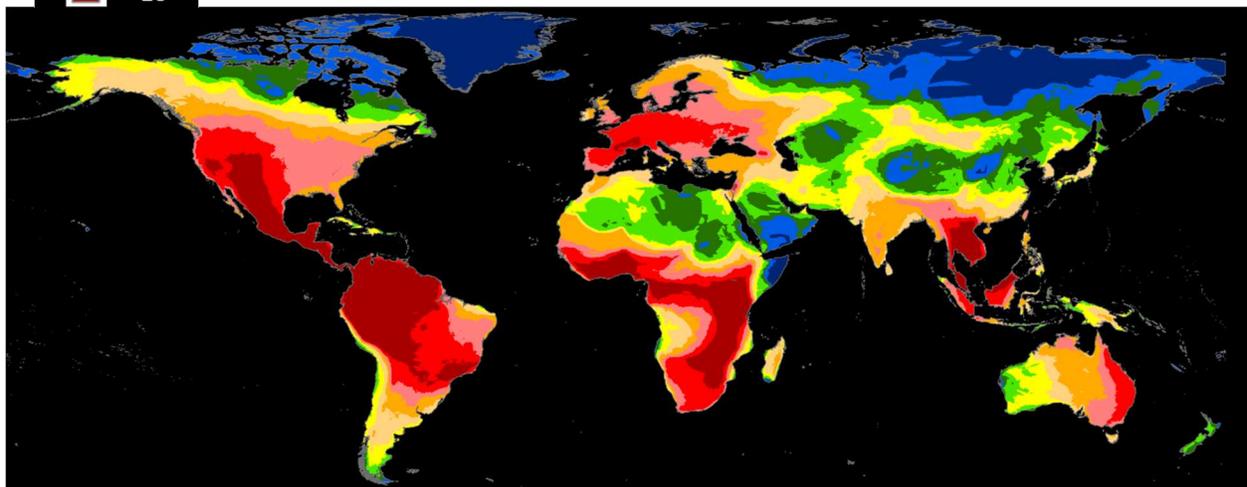258   and South Asia (also see Annex Figure 1).
259

260 **Figure 2. Matched mammals species densities: Marsh et al. (2022) versus GBIF occurrence**
261 **reports**

262

Marsh et al. (2022)
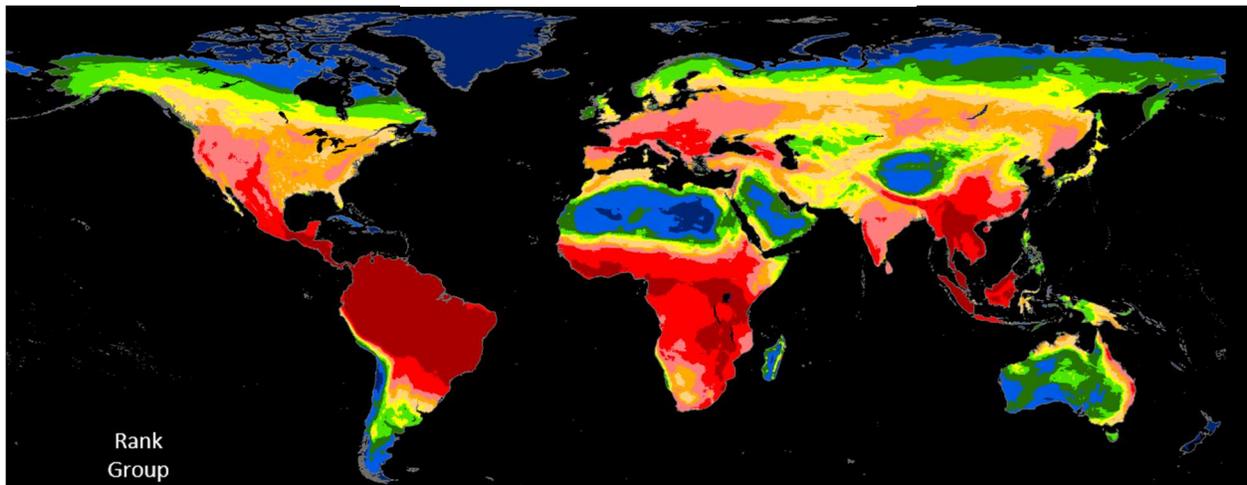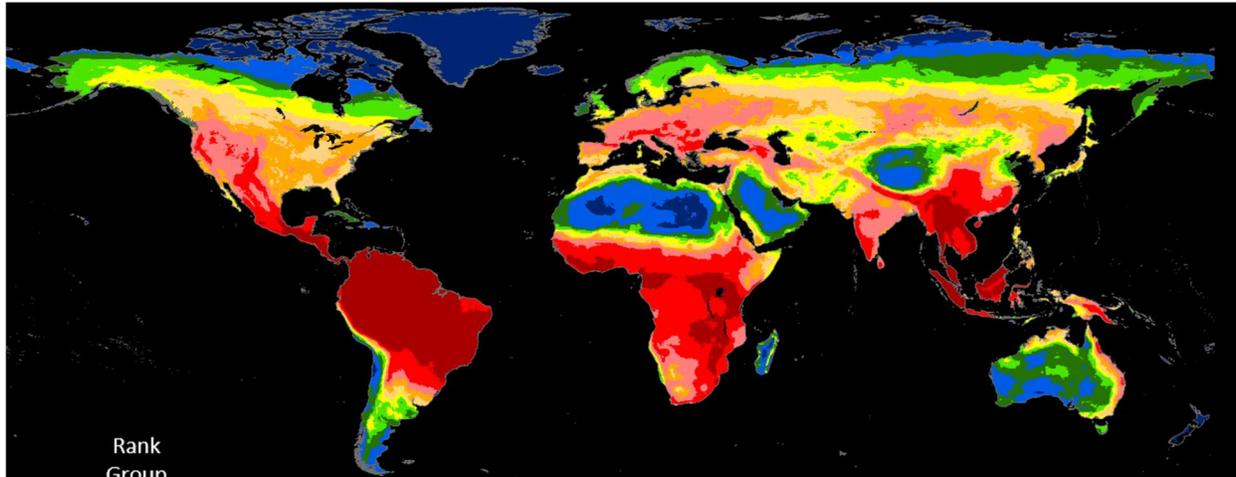

GBIF

263

264
265

11

266

267 Technical differences can explain divergences in the two patterns. For example, the Marsh et al.
268 (2022) maps estimate native ranges of global mammals, taking into account recorded historical
269 occurrences and biogeographical factors that correlate with the range of each mammal species. By
270 contrast, the GBIF mammal maps bound the areas where species occurrences have been reported
271 since 1970. Regions where GBIF ranks are higher than Marsh et al. (2022) ranks have many
272 species with reported occurrences beyond their estimated native ranges; regions with lower GBIF
273 ranks have occurrence reports clustered in subareas within native ranges. This difference could
274 reflect underreporting for GBIF species in lower-ranked areas, although many higher-ranked areas
275 appear similarly disadvantaged for species observation. In our view, the more plausible
276 explanation is that lower-ranked regions are populated by many species whose ranges have
277 contracted over time. The ongoing accumulation of GBIF species occurrence reports should help
278 to resolve this issue.

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

287

288

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

Marsh et al. (2022)

GBIF

292

293

294
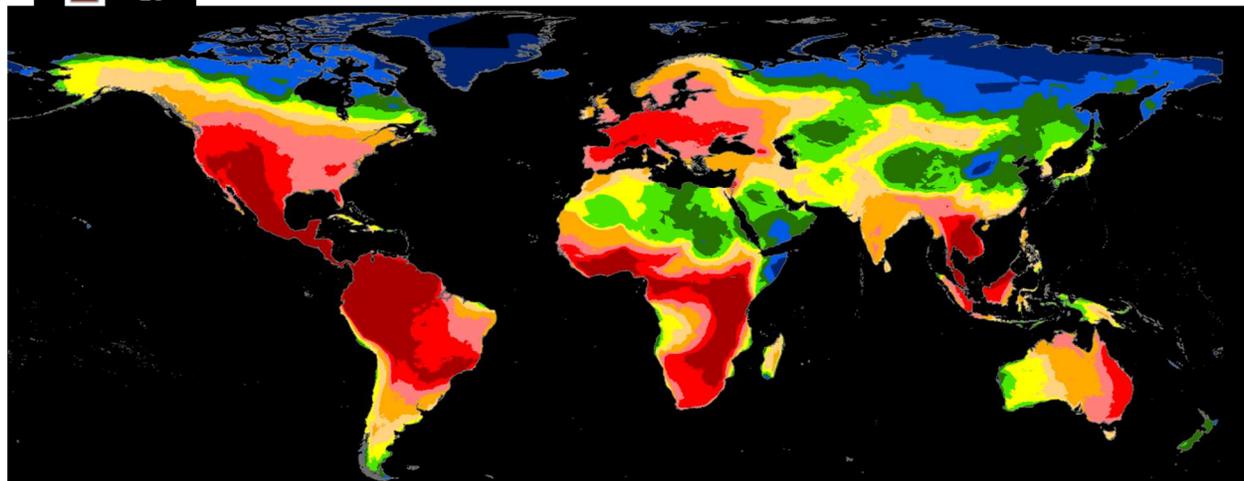
3.3.2 Ants

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 (also see Annex Figure 3).

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.
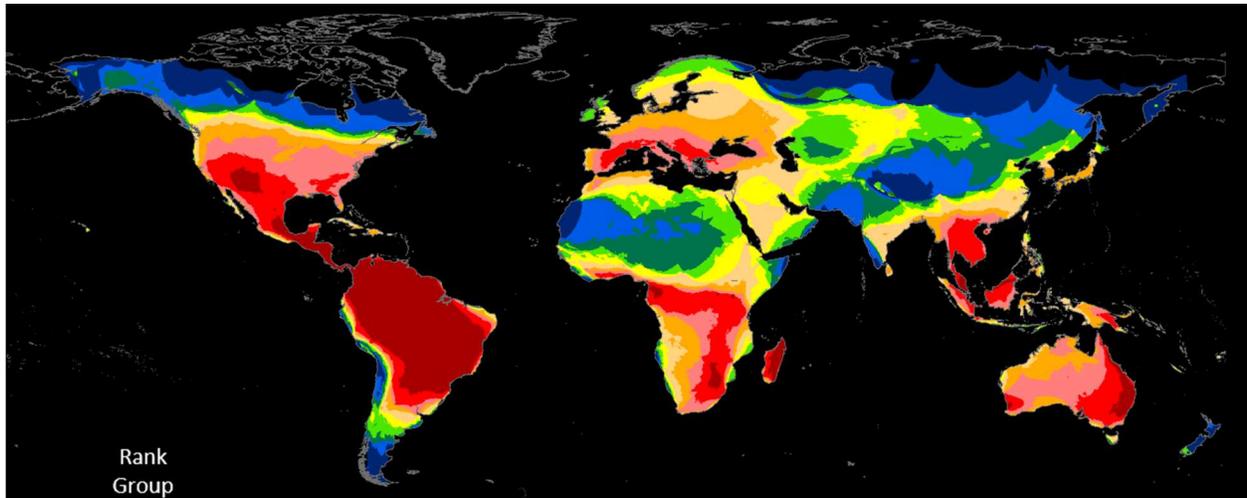
322

323

13

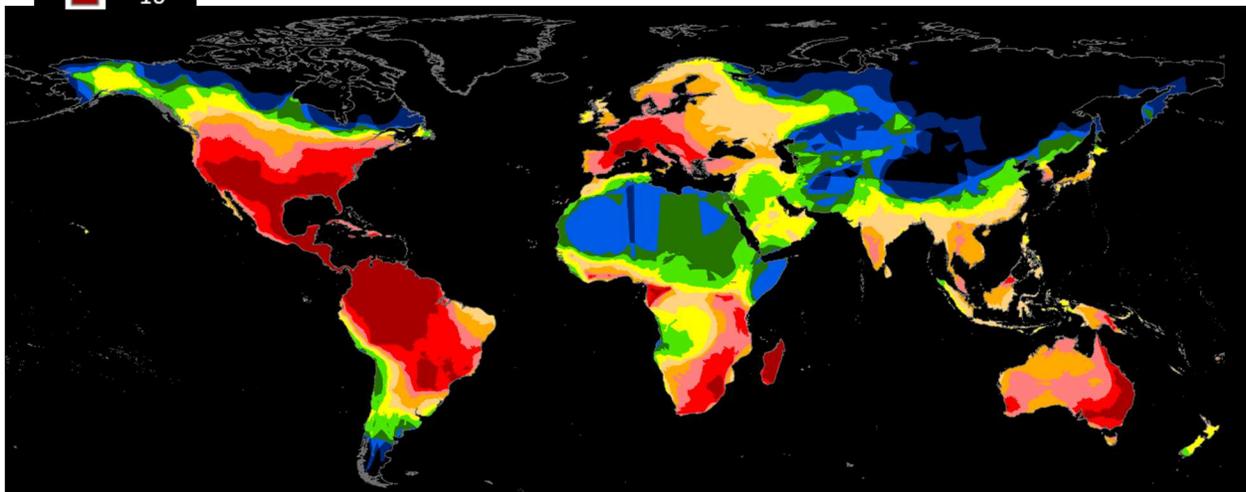**Figure 4. Matched ant species densities: Kass et al. (2022) versus GBIF occurrence reports**

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 (also see Annex Figure 4).
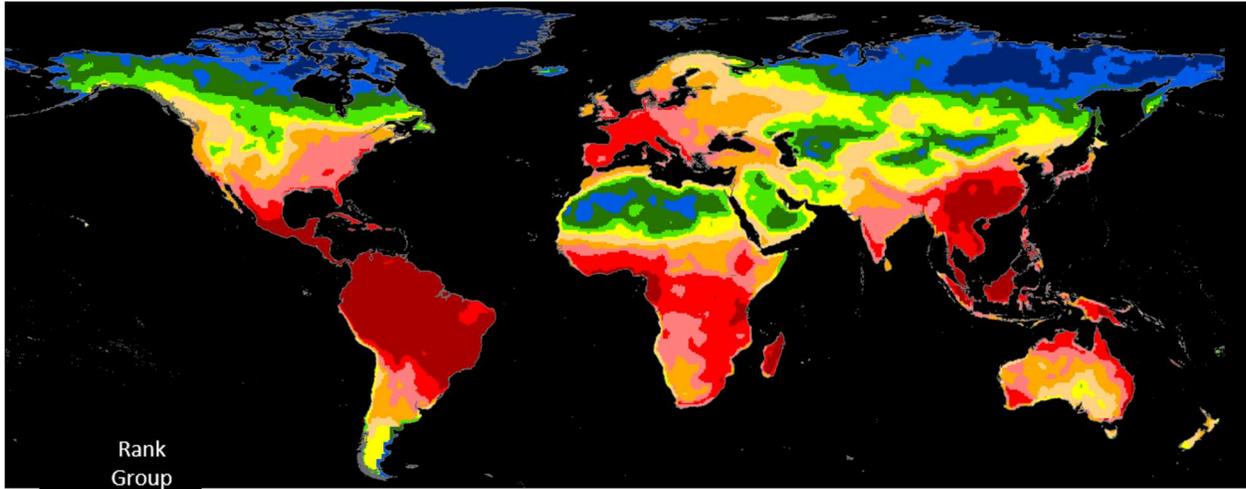
Figure 5. Matched vascular-plant species densities: Borgelt et al. (2022) versus GBIF occurrence reports
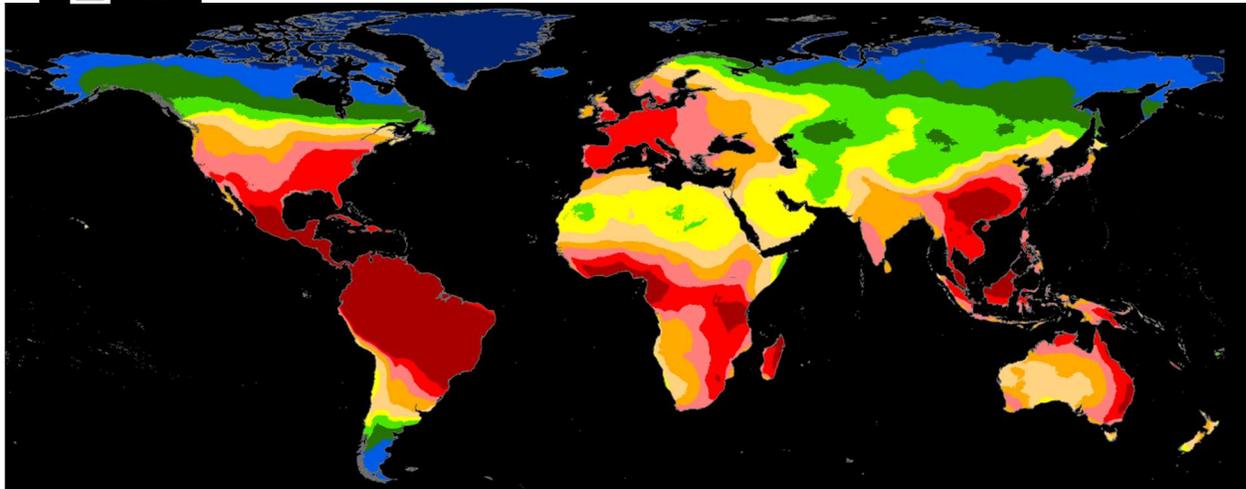
362
363
364

Borgelt et al. (2022)

365

366
367
368
369

370   *3.4   Summing up*

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

386   # 4. Priority-setting applications

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

399   *4.1 Endemic species distribution by group and country*

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

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

17

410   10 status to Australia, United States, Brazil, Mexico, South Africa, China, New Zealand,
411   Madagascar, Japan, and Costa Rica. Even for the top 30 countries, endemicity varies enormously
412   by species group. In terms of vertebrates, for example, 62 percent are endemic in Australia versus
413   only 3 percent in the United Kingdom. For plants, the endemicity in Madagascar, Australia, and
414   New Zealand is extremely high, at 89.6 percent, 88.1 percent, and 84.4 percent, respectively. For
415   arthropods, the maximum endemicity is even higher in New Zealand and Australia, at 92.8 percent
416   and 91.2 percent, respectively. Mollusks, fungi, and other species more closely resemble
417   arthropods and vascular plants in the relative compactness of their ranges.

418

419

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

Country	Vertebrates	Arthropods	Mollusks	Other animals	Vascular plants	Fungi	Other species
					Endemic 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

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

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

443 **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

444
445 We believe that an upper bound of 25 km x 25 km on critical scale for small-range species is
446 appropriately conservative. The small-range species count increases to 117,946 at 50 km x 50 km
447 or less, 170,081 at 100 km x 100 km or less, and 246,172 at 200 km x 200 km. From a policy
448 perspective, the feasibility and sustainability of species protection tend to decline as the number
449 of species protected increases. Since even the 25 km x 25 km limit qualifies nearly 85,310 species
450 as having a small occurrence region, we retain it here, recognizing that other analyses may well
451 opt for higher limits.

452 Using GIS overlays of GBIF maps and country boundaries, we count species with small occurrence
453 regions by country, finding that their international distribution is skewed. The top 30 countries
454 account for 75.5 percent of them (64,443 out of 85,310 species). Our overall results assign top 10
455 status to Australia, United States, Brazil, Mexico, France, South Africa, Costa Rica, China,
456 Colombia, and Japan. Australia leads with 8,673 species, followed by the United States (7,791),
457 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

Country	Species (%)						
	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**

22

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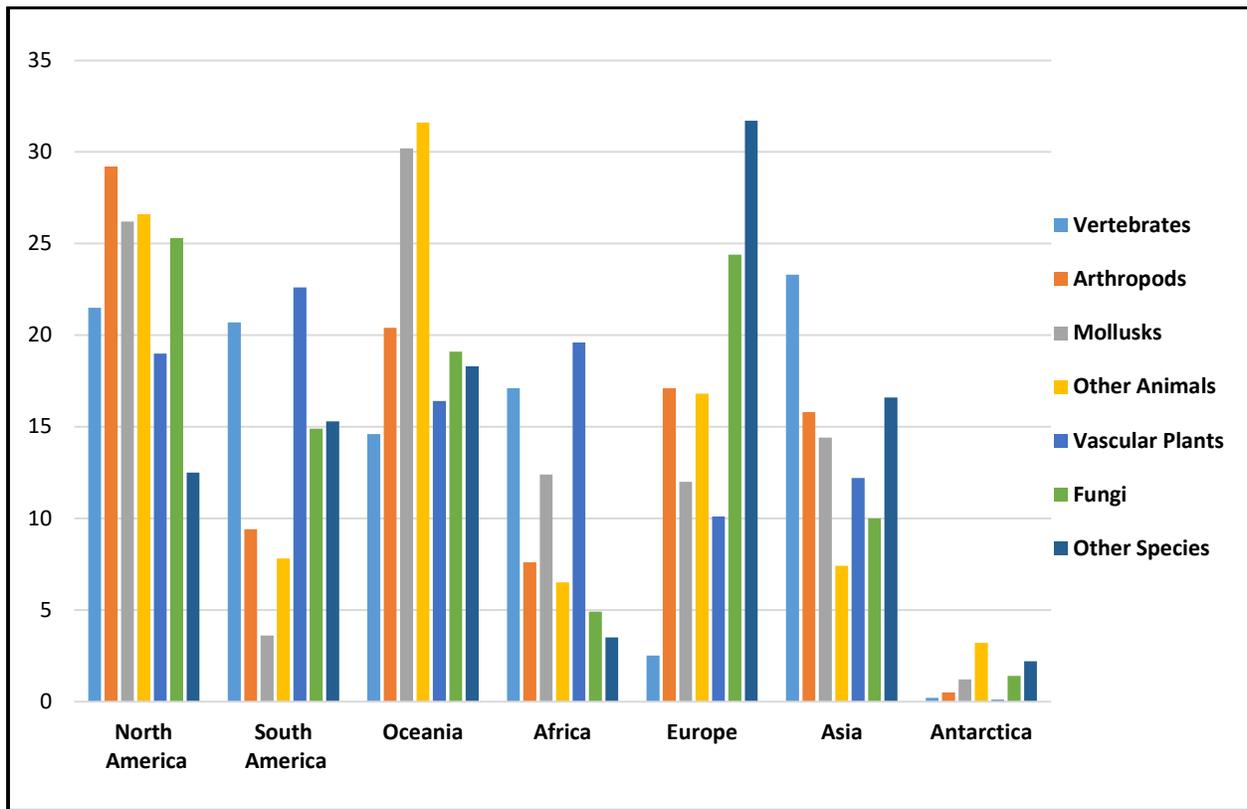
Among endemic species with small occurrence regions, the largest share is found in Oceania (28 percent), followed by North America (23 percent). Four regions are in the mid-range—South America (15 percent), Asia (13 percent), Africa (11 percent), and Europe (9 percent)—and Antarctica has small representation,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

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

539 **Figure 7. Percent distribution of endemic species with small occurrence regions, by income**
540 **class**
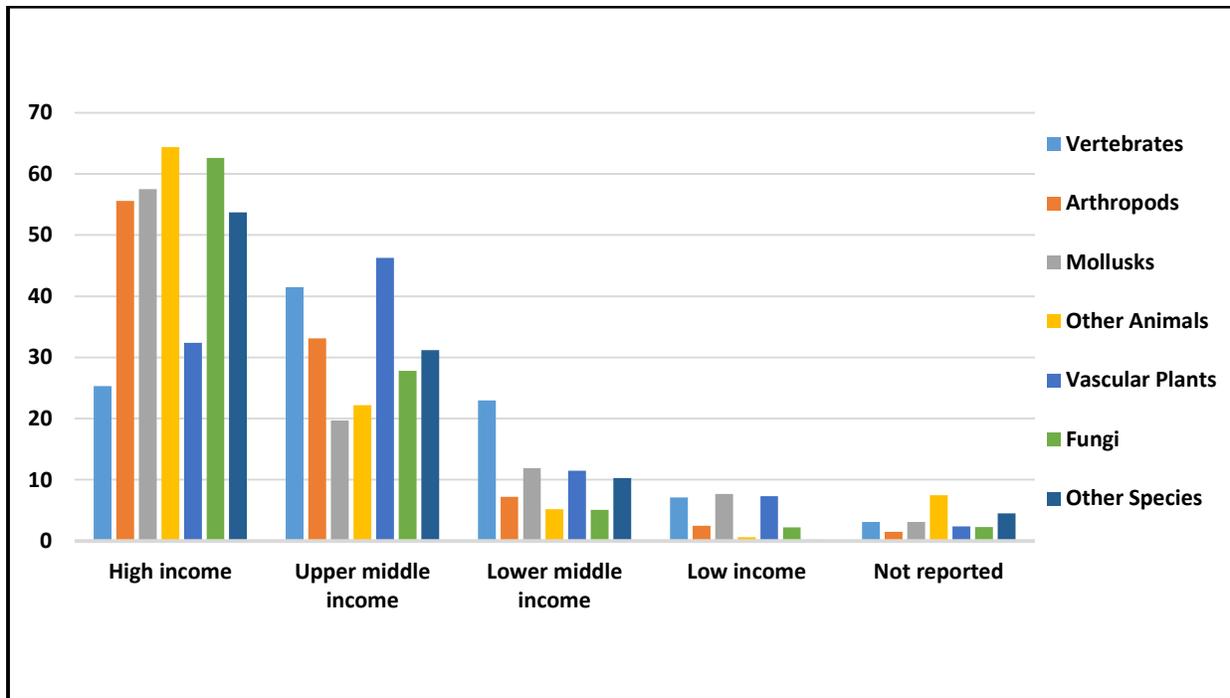
541
542
543 It should be noted that the maps constructed with processed data also provide opportunities for
544 understanding the geographic distribution of the species within countries.

545 5. Discussion

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

558 of previously unmapped taxa—particularly invertebrates, fungi, and plants—which substantially
559 broadens the taxonomic and spatial scope of global biodiversity representation.

560 *5.1 Expanding the biodiversity evidence base*

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

571 *5.2 Addressing taxonomic and geographic bias*

572 A major contribution of this work lies in filling taxonomic gaps that have historically constrained
573 global biodiversity assessments. Existing global datasets emphasize vertebrates, overlooking
574 invertebrates and fungi that together comprise the majority of known species (Cardoso et al.,
575 2011). The integration of over 213,000 arthropods, 38,000 fungi, and 24,000 other animal taxa
576 demonstrates the feasibility of incorporating these groups systematically into global-scale
577 analyses. While GBIF data remain spatially biased toward accessible and high-income regions
578 (Yesson et al., 2007), ongoing growth in citizen science platforms and institutional data sharing is
579 steadily improving geographic representativeness. As observation density increases, particularly
580 across tropical and remote regions, boundary estimation will gain precision, enabling more
581 accurate identification of small-range and data-poor species.

582 *5.3 Implications for conservation priority setting*

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

595 *5.4 Policy relevance and the Global Biodiversity Framework*

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

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

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

633 *5.5 Limitations and future directions*

634 Despite these advances, several limitations warrant acknowledgment. First, GBIF occurrence data
635 remain uneven in spatial and temporal coverage, with collection bias favoring regions with
636 established research infrastructure, accessibility and funding (Beck et al., 2022; Boakes et al.,
637 2010; Hickisch et al. 2019; Faxon and Chapman 2025). Hortal et al. (2015) discuss shortfalls of

biodiversity data including taxonomic (Linnean), distribution (Wallacean), and abundance (Prestonian). First, taxonomic challenges can persist because research effort is often linked to a disproportionate concentration on early-described taxa, as in the case of 10,000 reptile species examined by Guedes et al. (2023). In a study of the Eastern Arc Mountains in Tanzania, Ahrends et al. (2011) find suggestive evidence of biodiversity patterns linked with investment in inventories. Second, accurate species occurrence records are needed to improve the understanding of species distribution mitigating the "Wallacean Shortfall." ~~Second~~Third, presence-only data do not capture species abundance, which limits inference about population viability within mapped boundaries. ~~Third~~Fourth, 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, sampling, latitudinal taxonomic gradient and temporal updating to capture dynamic species–environment interactions (Elith & Leathwick, 2009; Diniz Filho et al. 2023).

Although GBIF strengthens the data ecosystem providing a common repository for data collections from scientists, agencies, community science and museums (Heberling et al., 2021), further investment would help improve GBIF completeness by building local digitization and curation capacity and by integrating diverse data streams. Empirical work in the Amazon illustrates both the promise and challenges of current approaches: de Araujo et al. (2022) find that combining global and national databases can increase biodiversity knowledge and reduce inventory gaps. Sustained efforts to digitize, publish, and improve data quality are essential for keeping biodiversity data efficient, effective, and relevant (Ball-Damerow et al. 2019). Targeted capacity-building programs, such as GBIF's Biodiversity Information for Development (BID), support data mobilization, training, and publisher engagement to improve the coverage and usability of biodiversity data across world (GBIF, 2025).

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

677 records and R software for data processing. The associated scripts are available from the authors
678 upon request.

679 *Ethics approval and consent to participate*: Not applicable

680 *Consent for publication:* Not applicable

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

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

684

Author contributions

686 **Susmita Dasgupt**a: Conceptualization, Funding Acquisition, Project Supervision, and Writing
687 original draft.

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

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

691

Acknowledgements

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

702

References

704 Ahrends, A., Burgess, N. D., Gereau, R. E., Marchant, R., Bulling, M. T., Lovett, J. C., ... &
705 Rahbek, C. Funding begets biodiversity. *Diversity and Distributions*, *17*(2), 191-200, 2011.

706 Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., ... &
707 Guralnick, R. P. Research applications of primary biodiversity databases in the digital age. *PloS*
708 *one*, *14*(9), e0215794, 2019. https://doi.org/10.1371/journal.pone.0215794

709 Blankespoor B, Dasgupta S, Wheeler D. Bridging Conflicts and Biodiversity Protection: The
710 Critical Role of Reliable and Comparable Data. World Bank Policy Research Working Paper
711 11076; 2025 February. https://documents.worldbank.org/en/publication/documents-
712 reports/documentdetail/099415202272518792

713 Blankespoor, B., Dasgupta, S., Wheeler, D., Jeuken, A., van Ginkel, K., Hill, K., and Hirschfeld,
714 D.: Linking sea-level research with local planning and adaptation needs, Nature Climate Change,
715 13(8), 760–763, 2023. https://doi.org/10.1038/s41558-023-01789-8

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

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

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

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

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

729 Burgin, C., Wilson, D., Mittermeier, R., Rylands, A., Lacher, T., and Sechrest, W.: Illustrated
730 Checklist of the Mammals of the World (Vol. 2, Eulipotyphla to Carnivora), Lynx Edicions,
731 2020b. ISBN: 978-84-16728-36-7

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

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

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

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

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

De Araujo, M. L., Quaresma, A. C., & Ramos, F. N. GBIF information is not enough: national database improves the inventory completeness of Amazonian epiphytes. *Biodiversity and Conservation*, *31*(11), 2797-2815, 2022. https://doi.org/10.1007/s10531-022-02458-x

Diniz Filho, J. A. F., Jardim, L., Guedes, J. J., Meyer, L., Stropp, J., Frateles, L. E. F., ... and Hortal, J. Macroecological links between the Linnean, Wallacean, and Darwinian shortfalls. *Frontiers of Biogeography*, *15*(2), 2023. https://doi.org/10.21425/F5FBG59566

Elith, J., and 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

Faxon, H., & Chapman, M. Beyond spatial bias: understanding the colonial legacies and contemporary social forces shaping biodiversity data. *Environmental Research Letters*, *20*(6), 064053, 2025. http://doi.org/10.1088/1748-9326/add6b6

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.

GBIF. BID: Biodiversity Information for Development, 2025. Accessed 2025-12-08. https://www.gbif.org/programme/82243/bid-biodiversity-information-for-development

Guedes, J. J., Moura, M. R., & Alexandre F. Diniz-Filho, J. (2023). Species out of sight: elucidating the determinants of research effort in global reptiles. *Ecography*, (3), e06491. 2023. https://doi.org/10.1111/ecog.06491

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

Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B., & Schigel, D. Data integration enables global biodiversity synthesis. *Proceedings of the National Academy of Sciences*, *118*(6), e2018093118, 2021. https://doi.org/10.1073/pnas.2018093118

Hickisch, R., Hodgetts, T., Johnson, P. J., Sillero-Zubiri, C., Tockner, K., & Macdonald, D. W. (2019). Effects of publication bias on conservation planning. *Conservation Biology*, *33*(5), 1151-1163. https://doi.org/10.1111/cobi.13326

Hortal, J., De Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual review of ecology, evolution, and systematics*, *46*(1), 523-549, 2015. https://doi.org/10.1146/annurev-ecolsys-112414-054400

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

814 Meyer, C., Weigelt, P., & Kreft, H.: Multidimensional biases, gaps and uncertainties in global
815 plant occurrence information. *Ecology Letters, 19*(8), 992–1006.
816 2016. https://doi.org/10.1111/ele.12624

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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