



Characterizing tree species richness in indigenous peoples' lands: addressing often-overlooked data limitations and biases



Sven Kock¹✉, Jorge C. Llopis^{1,2}, Miguel Fernández^{3,4}, Mónica Moraes R^{5,6} & Álvaro Fernández-Llamazares^{1,7}

Several studies have characterized species richness distribution in Indigenous Peoples' Lands (IPL) using global biodiversity datasets. However, these datasets contain substantial limitations and biases that may compromise the accuracy of spatial biodiversity assessments. In this study, we characterize tree species richness across IPL in Bolivia based on three different datasets, and then examine how spatial and taxonomic biases shape characterization outcomes. We also assess the institutional provenance of the data. Characterization results closely reflect identified spatial and taxonomic biases. As data generally come from various sources, the scope of the individual collections and our characterization do not align. Most data are hosted by international institutions without information about contributions from Indigenous Peoples. This complicates sensitive interpretation and weakens visibility of Indigenous knowledge systems. Our findings underscore the limitations of relying solely on global biodiversity datasets to map species distributions in IPL and emphasize the need for locally grounded, ethically framed data collection practices.

Biodiversity is rapidly declining at a global scale, driven primarily by land use change, overexploitation, and environmental pollution¹. This decline represents a profound loss of the intrinsic value of nature² and results in a significant reduction of ecosystem services upon which human well-being depend³, making biodiversity loss a central concern in global sustainability efforts⁴. While mainstream conservation approaches have struggled to prevent this continuous degradation⁵, many Indigenous Peoples have been successful at maintaining sustainable ways of living for generations⁶. There is increasing recognition in both scientific and policy circles that Indigenous Peoples play globally important roles in biodiversity conservation^{7,8}. However, Indigenous knowledge systems, which underpin many long-standing sustainable practices, are increasingly subject to considerable pressures, such as suppression, obliteration, assimilation and misappropriation⁹.

In this context, numerous scientific analyses have investigated the spatial relation between Indigenous Peoples' Lands (IPL) and different metrics of species richness and diversity. Many of these analyses have

focused on particular species groups such as mammals, bats or vertebrates, and have characterized the overlap between their distribution ranges and IPL^{10,11}. Some have compared this overlap to that with Protected Areas (PA) and other areas (OA)^{12,13}. Schuster et al.¹⁴ used species ranges to identify the species richness in IPL, PA and OA¹⁴, and Seebens et al.¹⁵ calculated the mean number of alien species in IPL and OA. These studies, which we call “spatial characterizations” in this article, aimed to characterize species composition or distribution in IPL (see refs. 10,11), or compare them to those in other areas (see refs. 12–15).

The mentioned spatial characterizations draw on global biodiversity databases, mostly the global International Union for Conservation of Nature (IUCN) Red List of Threatened Species¹⁶ (hereafter Red List, see refs. 10–14). Seebens et al.¹⁵ relied on the database of the Global Biodiversity Information Facility¹⁷ (hereafter GBIF). However, these databases are not designed to measure spatial species distribution, potentially undermining the accuracy and representativeness of these efforts. In fact, the Red List is a selective effort

¹Institute of Environmental Science and Technology (ICTA-UAB), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain. ²Centre for Development and Environment, University of Bern, Bern, Switzerland. ³Department of Environmental Science and Policy, George Mason University, Fairfax, VA, USA. ⁴FONPLATA Development Bank, Santa Cruz, Bolivia. ⁵Instituto de Ecología, Universidad Mayor de San Andrés, La Paz, Bolivia. ⁶Herbario Nacional de Bolivia, La Paz, Bolivia. ⁷Department of Animal Biology, Plant Biology and Ecology (BABVE), Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain.

✉e-mail: skock@outlook.de

focused exclusively on assessing extinction risk, with no claim to spatial representativeness¹⁸. GBIF represents botanic collections accumulated over time¹⁹, so it cannot account for case-study-specific sample requirements, such as representative sites and species, or temporal regularity²⁰.

In general, it is well known that global inventories of biodiversity are incomplete^{21,22}. Numerous species are still undescribed²³ (Linnean shortfall, after Karl von Linné), and scientific knowledge about their distribution is meager at best, particularly at the local scale required to distinguish between different area management types in spatial characterizations²⁴ (Wallacean shortfall, after Alfred Russel Wallace). See ref. 25 for a detailed description of the Linnean and Wallacean shortfalls, as well as further gaps in our current knowledge of biodiversity. Additionally, biodiversity data are significantly influenced by the preferences and limitations of those involved in collecting it, which makes the data biased²⁶.

In this article, we assess how spatial and taxonomic biases in biodiversity datasets undermine the characterization of tree species richness in IPL, PA, and OA (from here referred to as area types). Specifically, we focus on the case of Bolivia, a country with high biological and cultural diversity^{27–30}, where, according to its national census, more than half of the population self-identifies as Indigenous³⁰.

For our analyses, we first identify the spatial distribution of the different area types in Bolivia and illustrate the estimated tree species richness across its ecoregions³¹.

Then, we conduct spatial characterizations, assessing the distribution of tree species across the three different area types we defined in Bolivia (IPL, PA, and OA). We perform these characterizations using methodologies similar to those in the aforementioned studies (Refs. 10–15). However, unlike these studies, we assess the same study area separately using multiple databases and measures to gain insight into the dependence of the results on such methodological and data choices.

To discuss the presence of biases in the datasets, we visualize the spatial and taxonomic distribution of the data and make them comparable to available ecological descriptions of the biodiversity of Bolivia. In particular, for the spatial distribution, we map species richness across the ecoregions in Bolivia to compare it to the previously mapped estimates by Ibisch et al.³¹. For the taxonomic distribution, we examine the distribution of species across families to compare it with published descriptions of representative plant families in Bolivia.

In addition, we examine which studies generated the data or which institutions hosted them to determine whether their spatial and taxonomic scopes align with those of the spatial characterizations we conducted. We also discuss whether the metadata provided with the datasets can help identify possible limitations and preferences influencing the decision-making in the data generation.

The datasets used in this study are derived from the Red List and the GBIF. We additionally consider a national dataset by the Bolivian Biodiversity Observation Network (hereafter BBON)³². We include this database to ensure that any identified patterns are not an outcome of the mismatch between the global scale of the datasets and the national scale of our assessment. For instance, in ref. 22, the nationally compiled dataset was more detailed than a global one. The Red List provides estimated species ranges based on known occurrences, abiotic features, and expert scientific knowledge³³. GBIF hosts the largest species occurrence database, offering a standardized framework that enables the integration of records from diverse sources and across extended timeframes¹⁷. BBON is the most comprehensive national species occurrence database for Bolivia, combining records from multiple sources, similar to GBIF. Unlike the datasets downloaded from GBIF, the BBON datasets underwent extensive quality control during compilation. Despite these differences, considerable overlap between GBIF and BBON records can be expected.

We focus on tree species, considering that, as of today, they have not been addressed in any spatial characterization related to IPL. Trees are particularly relevant because they are long-lived organisms that often embody enduring ethnobiological relationships, reflecting generations of place-based cultural practices and spiritual knowledge³⁴. In Bolivia,

numerous Indigenous communities maintain rich traditions of tree use and stewardship, including the cultivation of culturally-significant species in homegardens, the use of specific trees in rituals and healing practices, and the transmission of tree-related knowledge across multiple generations^{35–37}. Our hypothesis is that the results of the spatial characterizations are strongly influenced by spatial and taxonomic biases, and that these biases are related to the sources of the datasets.

Results

Spatial characterization

IPL cover 19% of the Bolivian Territory, with another 19% under PA status, and 4% with dual IPL and PA status (from here on IPL-PA). IPL thereby includes both Community Lands of Origin (TCO, according to their name in Spanish) and Indigenous Peasant Territories (TIOC, according to their name in Spanish), where third parties have legally acquired and respected rights³⁸. IPL-PA areas are not independently designated, but represent zones where IPL and PA overlap spatially. Since IPL, PA, and IPL-PA areas overlap in these regions, the combined area type percentages exceed 100% by an amount equal to twice the percentage of IPL-PA. The remaining areas (i.e., OA) cover around 66% of Bolivia (Fig. 1a).

Figure 1b illustrates the estimated tree species richness distribution over the different ecoregions, based on the estimations by Ibisch et al.³¹. According to these estimations, the highest richness can be expected in the northern part of the country, particularly the Southwestern Amazon, and the Yungas (ecoregions 1, and 6). In contrast, the Chaco Cerrado, the Flooded Savannas of the Pantanal, the Gran Chaco, the Mantane Chaco, and the Prepuna (ecoregions 2.4, 3.2, 5, 8, 9, and 10), located in the south and southeast, are estimated to have low tree species richness. Intermediate levels of species richness are estimated for all other Cerrado regions, the Flooded Savannas of the Llanos de Moxos, the Chiquitano Dry Forest, and the Tucumano-Bolivian Forest (ecoregions 2.1–2.3, 3.1, 4, and 7; see Table S2). Ibisch et al.³¹ do not provide estimates for the Puna ecoregions in the southwest, but characterized them as scarcely vegetated (ecoregion 11) or as too dry for any considerable vegetation (ecoregion 12).

For Bolivia, the Red List provides ranges for 45 different tree species, the GBIF dataset 257,999 occurrences of 4,135 distinct tree species, and the BBON dataset 199,263 occurrences of 4142 distinct tree species. Overlapping the species ranges from the Red List shows that most of them are concentrated in the central south (Fig. 2a). Species richness in GBIF and BBON is more evenly distributed across the country. As they are based on observation points and not on ranges, they occur more punctually than those of the Red List, which cover the entire area of Bolivia (see Fig. 2d and 2g). Many of the locations with the highest species richness in GBIF and BBON are in the northwest of the country.

The middle column in Fig. 2 (subfigures b, e, and h) presents the number of distinct species in each area type. The OA area type has the highest values for all datasets. There are fewer species within PA and IPL than in OA, and the fewest are in IPL-PA areas. Notably, these values relate to the area type sizes, as OA cover the highest and IPL-PA areas the lowest share of Bolivia. However, when considering species density (Fig. 2f, i), which is the number of species per area size, IPL-PA have higher values, and OA have lower values than IPL and PA. Based on the Red List data, in PA and OA, higher relative species ranges are measured than in IPL and IPL-PA (Fig. 2c), with differences between area types less pronounced than for species density based on GBIF and BBON. While the highest species density based on GBIF and BBON is more than six times higher than the lowest, the difference between the relative species ranges measured based on the Red List is no more than 30%. Overall, the three different calculations (present species, relative species range, and species density) yield notably different results with no shared apparent pattern. Moreover, the results for both the GBIF and the BBON datasets are very similar.

Spatial distribution of the biodiversity data

Across ecoregions (Fig. 3), the highest number of species ranges in the Red List are found in the Inter-Andean Dry Forests (ecoregion 9), followed by

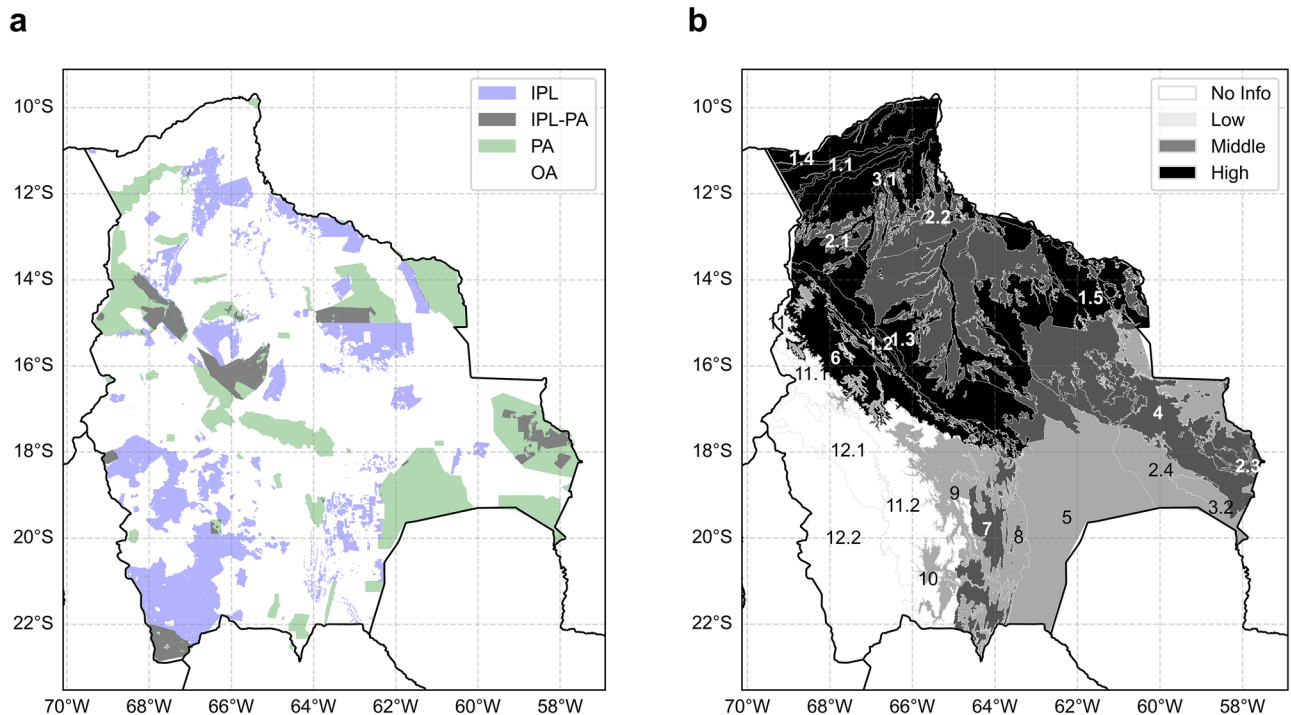


Fig. 1 | **a** Distribution of IPL, IPL-PA, PA, and OA in Bolivia, and **b** Distribution of ecoregions in Bolivia and estimated tree species richness. The ecoregions are 1: Southwestern Amazon, 2: Cerrado, 3: Flooded Savannas, 4 Chiquitano Dry Forest, 5:

Gran Chaco, 6: Yungas, 7 Tucumano–Bolivian Forest, 8: Montane Chaco, 9: Inter-Andean Dry Forests, 10: Prepuna, 11: Northern Puna, 12: Southern Puna. See Table S1 for more details.

the Yungas (ecoregion 6), and the Subhumid Puna (ecoregion 11). Slightly less ranges intersect with the Tucumano Bolivian Forests (ecoregion 7), the Eastern Cordillera (ecoregion 11.3), the montane Chaco (ecoregion 8), and the Gran Chaco (ecoregion 5). In the Amazon, Cerrado, savanna, and dry forest regions northeast of the Bolivian Yungas (ecoregions 1–4), as well as the Southern Dry Puna in the south-east (ecoregion 12), there are only few Red List species ranges.

The distributions of GBIF and BBON are comparable to each other. For both GBIF and BBON, most distinct species are found in the Pre-Andean Amazonian Forests (ecoregion 11.3) and Yungas (ecoregion 6), followed by the other Amazonian ecoregions (1.1, 1.2, 1.4, 1.5), and the Inter-Andean Dry Forests (ecoregion 9). The Pre-, Northern and Southern Puna (ecoregions 10–12), as well as most Cerrado ecoregions (2.1, 2.2, 2.4), stick out with particularly low species richness in the datasets.

Zooming into the ecoregions, both the GBIF and BBON data are very clustered. In the GBIF dataset, 55 of the 11,189 grid cells (0.5% of the total area) concentrate more than 50% of the observations, while around 72% of the grid cells do not contain any observation. The BBON dataset shows a similar pattern, with 49 of 11,189 grid cells (0.4% of the total area) containing more than 50% of the observations, while 73% of the cells do not have any observation. Most of the grid cells with the highest observation counts were located in PA (Table S3), and total observations in PA exceeded those in IPL by a factor of 2.5 (GBIF) and 1.9 (BBON). In general, the highest number of overall observations per area type in GBIF and BBON is in IPL-PA, followed by PA, IPL, and OA (See Table S4).

Taxonomic distribution of the biodiversity data

In the Red List, 21 out of 45 species, or almost half, are Cactaceae (Fig. 4). The other 24 species are distributed across 17 families, with a maximum of three species (< 7% of the 45 species) per family. The families with most species in GBIF and BBON are Fabaceae and Rubiaceae. Together, they contribute almost 20% of more than 4000 distinct species in each dataset, while they are not present in the Red List. The remaining species are distributed across 157 (GBIF) and 159 (BBON) families, with no family

with a share higher than 7%. Notably, only around 0.5% of the species in both GBIF and BBON are Cactaceae, the most abundant family in the Red List.

Assessment of the dataset sources

The Red List data was provided by five institutions (Fig. 5a). A global cactus assessment contributed 21 of the 45 species ranges. Six are associated with the TROPICOS collection generated by the Missouri Botanical Garden (MBG) in cooperation with national institutions (according to the expert knowledge of author M.M.R., although this information is not explicitly documented in the dataset metadata), three are associated with the Herbario Nacional de Bolivia (LPB), and three are associated with a joint effort between the MBG and the LPB. Three more entries are affiliated with Botanic Gardens Conservation International (BGCI), meaning that they were collected by Bolivian institutions with funding by BGCI. For the remaining five entries, no information about the origin is provided. Overall, at least 49% of the data are from international institutions.

The GBIF dataset has occurrence entries from 322 different institutions, with almost 80% of the entries provided by six different herbaria. At least one-third of the herbaria are international (Fig. 5b), particularly from Denmark, Spain, and the US. Around 78% of the entries are based on preserved specimens, 81% of which are provided by TROPICOS, which is associated with the data by all six herbaria, except the University of Aarhus Herbarium in Denmark (AAU). Another 22% of the entries in GBIF are observation data provided by the AAU (90%) and iNaturalist (9.4%). See Table S5 for more details.

The BBON dataset does not include metadata about individual data sources. The authors report that 55.4% of the data included is hosted by institutions based outside Bolivia, and 44.6% by institutions based within the country. Moreover, 93.7% of their data is from voucher specimens and 6.3% from direct expert observations. However, these numbers are not comparable to those for the other datasets, because they are for all species in the entire dataset and not only the tree species we assess.

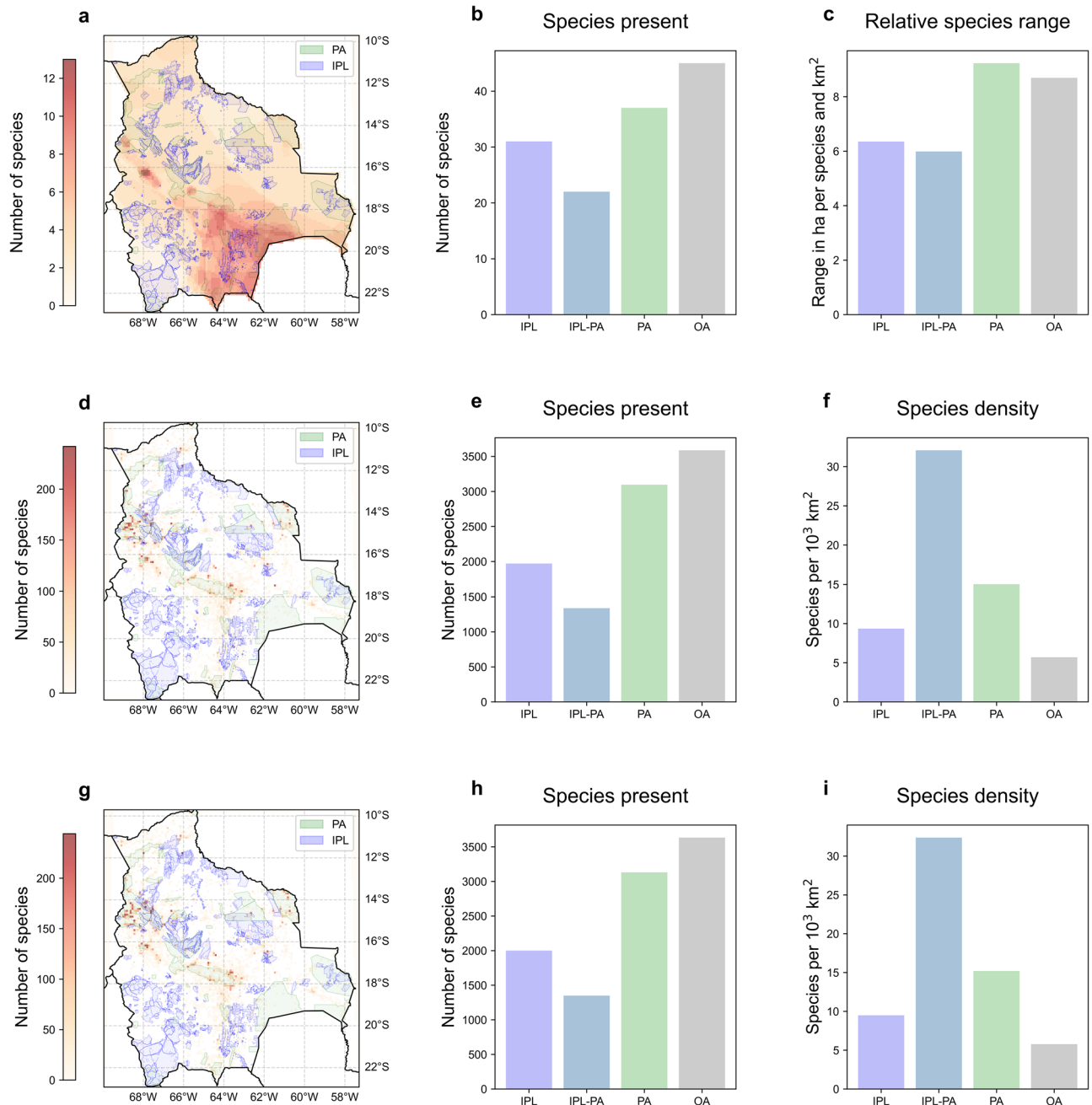


Fig. 2 | Distribution maps (a, d, g), number of species in different area types (b, e, h), relative species range (c), and species density (f, i) based on Red List (top row, a–c), GBIF (middle row d–f), and BBON (bottom row, g–i).

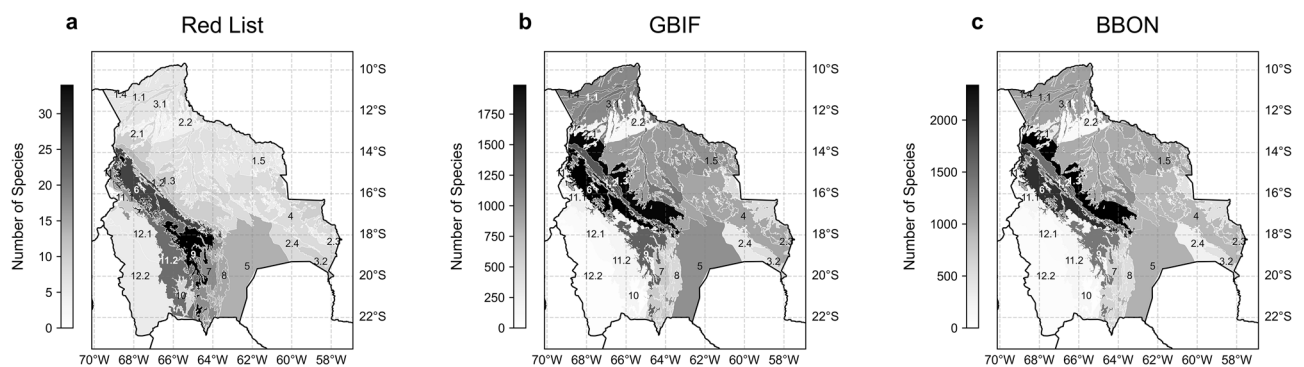


Fig. 3 | Number of species intersecting with the different ecoregions of Bolivia after the (a) Red List, (b) the GBIF, and (c) the BBON datasets.

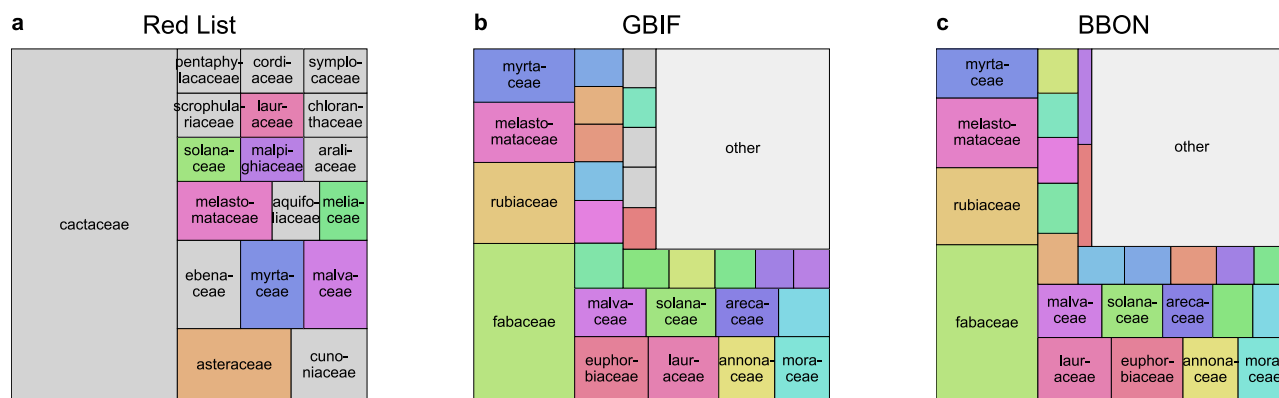
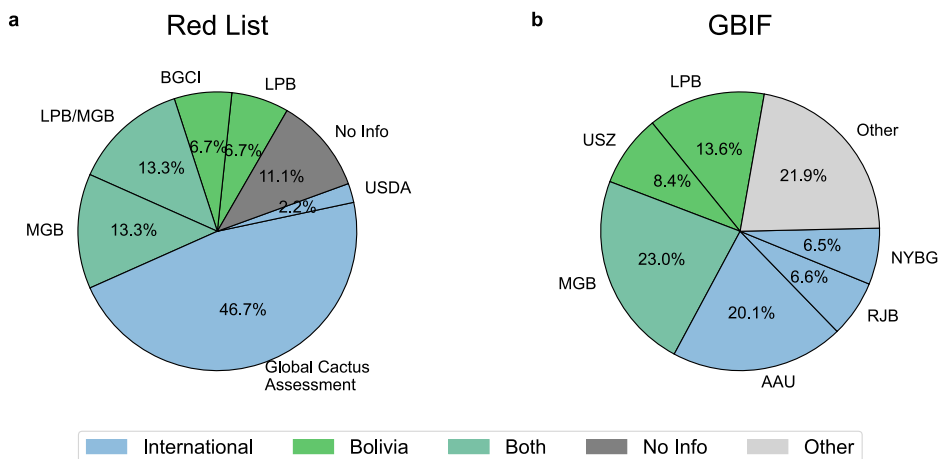


Fig. 4 | Shares of the total number of species in the different families for Red List (left), GBIF (middle), and BBON (right). Families representing less than 1% of all species are grouped under “Other”. Families above 1% and existing in at least two datasets have the same colors.

Fig. 5 | Share of studies and collections providing the data for the Red List (left) and GBIF (right), and their origin. Where MBG is the Missouri Botanical Garden, US, AAU is the University of Aarhus Herbarium, Denmark, LPB is the Herbario Nacional de Bolivia—Universidad Mayor de San Andrés, La Paz, Bolivia, USZ is the Herbario—Universidad Gabriel René Moreno in Santa Cruz, Bolivia, RJB is the Real Jardín Botánico, Madrid, Spain, BGCI is Botanic Gardens Conservation International, NYBG is the New York Botanical Garden, and USDA is the US Department of Agriculture.



Discussion

This study used three different datasets to characterize species distribution across area types within the same study area and generated strikingly different results. This suggests that such characterizations have low robustness across different datasets. As a result, a researcher assessing species distribution based on the Red List may identify different priority areas, as well as different threats and adverse factors affecting species, than another researcher using GBIF or BBON.

In this section, we firstly examine the spatial and taxonomic biases in the representation of biodiversity within the used datasets. Secondly, we discuss structural reasons for these biases and whether they were mitigated by choosing a national-scale instead of a global-scale database. Thirdly, we evaluate how the biases may have impacted the spatial characterization in our case study and other studies in the scientific literature. Finally, we reflect on how the biases encountered, as well as the methodological design of quantitative characterizations, may obscure rather than reveal positive interrelations between Indigenous knowledge systems and biodiversity.

Spatial biases are observable across ecoregions, as none of the distributions completely align with the estimated distribution after Ibisch et al.³¹. According to their estimates, the highest tree species richness occurs in northern Bolivia, particularly within the Southwestern Amazon and the Yungas (ecoregions 1, and 6). However, few Red List species ranges intersect with the Southwestern Amazon (ecoregion 1). Instead, many species overlap with southern and southwestern ecoregions with lower tree species richness, such as the Inter-Andean Dry Forests, or the Chaco ecoregions (9, 8, and 5), and with the scarcely vegetated Northern Puna (ecoregion 11). GBIF and BBON have far less species in most of the Southeastern Amazon (ecoregion

1) than in the Yungas (ecoregion 6), though Ibisch et al.³¹ indicated that the Southeastern Amazon is the species-rich ecoregion of Bolivia (see Table S2). Also, there is a lower species richness in the Tucuman Forest (ecoregion 7) than in the Inter-Andean Dry Forests (ecoregion 9) and the Gran Chaco (ecoregion 5).

There are also spatial biases within ecoregions, since the species richness distribution considerably derives from expectable patterns. In the Red List, the distribution does not reflect characteristic landscape features. For example, valleys in the Central Andean and Dry Puna (ecoregions 4 and 5), which are more vegetated than the surrounding dry environment³¹, do not resembled the species distribution (Fig. 2a). Considering that the Red List ranges are estimations informed by data and expert knowledge³³, the species ranges were likely drawn too coarsely to capture such small-scale features. In GBIF and BBON, there is a strongly uneven distribution due to clustering, an issue that was also identified by Beck et al.³⁹ in GBIF records from Switzerland. Other than in their Switzerland case study, clustering cannot be resolved in Bolivia by sub-sampling because for most of the area, no data are available. Considering that biodiversity is generally better documented in the Global North than in the Global South^{21,26}, this issue likely affects other regions in the Global South as well.

Significant taxonomic biases can be observed in the Red List. Half of the species in the dataset are Cactaceae. However, the family only constitutes 1% of the native tree species in Bolivia⁴⁰. It is characteristic in southwestern ecoregions such as the Dry Chaco (ecoregion 2), but not in the tree-rich ecoregions of the country³¹. The taxonomic distributions in GBIF and BBON are similar to those of a list of native Bolivian tree species by Moraes et al.⁴⁰, indicating a better representativeness.

Examining the sources of the datasets provides insight into the structural nature of the identified biases. All three datasets combine data from various sources, likely merging information with differing spatial and taxonomic scopes¹⁹. This aggregation results in non-representative sampling of both locations and species, which are crucial factors for accurately characterizing regional biodiversity²⁰. For instance, almost half of the Red List species ranges originated in the Global Cactus Assessment by Goettsch et al.⁴¹, which evaluated global cactus threat status. As our spatial characterization targets all tree species rather than cacti alone, the sampled species are unrepresentative for our study (or any study with similar objectives) and help explain the taxonomic bias identified before.

In GBIF, around one-fifth of the entries (those by AAU) correspond to a study by Cámara-Leret et al.⁴², who inventoried palm species in 96 different sites across the Western Amazon to correlate them to the respective biophysical conditions. The GBIF dataset has up to 3400 species records at these observation sites, which contributes to the clustering effect and results in an unevenly high likelihood of detecting elevated species richness there. Most of the sites in Cámara-Leret et al.⁴² were in the Southwestern Amazon (ecoregion 1, 42%) and the Yungas (ecoregion 6, 23%). Therefore, neither their spatial scope nor their taxonomic focus on palm species aligns with the scope of our spatial characterization, which considers all tree species in all Bolivia.

Biases were not mitigated in our case study by selecting a carefully curated national-scale instead of a global-scale database. BBON has a national scope, as opposed to the global scope of GBIF, and around 30% fewer observations than GBIF. It also underwent more extensive data cleaning than the data by GBIF. Nevertheless, the results based on both datasets are similar throughout all analyses. This indicates that size and overall spatial scope of the datasets, as well as data cleaning, played a minor role in mitigating biases. This observation relates to our previous discussion: regardless of the overall spatial scope of the datasets, they all integrate sources with varying scopes. As a result, most of the studies they rely on likely have different scopes than the spatial characterization for which they are used. The observation that the distributions are not affected by the dataset size supports an argument by Chapman et al.²⁶ that collecting more data cannot be expected to mitigate biases if the individual and structural factors shaping data-generation processes remain unchanged.

The identified biases relate to observed deviations in the spatial characterizations. For example, many of the ecoregions with relatively higher species richness than expected in the Red List (particularly ecoregions 7, 8, and 11) have almost no share of IPL-PA areas. This aligns with the low relative species ranges we measured in these areas (Fig. 2c), compared to the species densities based on GBIF and BBON (Fig. 2f and i). The overrepresentation of Cactaceae in the Red List contributed meaningfully to this spatial bias. Furthermore, the disproportionately high number of cluster cells and observations per area in IPL-PA and PA (see Tables S1, S3, and S4) align with high species densities measured for these area types based on GBIF and BBON (Fig. 2f and i).

Biases such as those observed in our analysis have also likely influenced other studies in the literature. References 10–14 examined the proportion of Red List species ranges that overlap with different area types. This measure is similar to the relative species range measure used in this study (Fig. 2c). Reference 15 considered the number of species per area, which is analogous to the species density measure in this case study (Fig. 2f, i).

Interestingly, some of these analyses identified smaller, rather than larger, proportions of species range within IPL than in the overall study area. For instance, ref. 10 reported that half of the species have less than 20% of their ranges in IPL, whereas the overall share of IPL was 28%. Similarly, ref. 11 found that only 22 of the 223 assessed species have an overproportionate share of their ranges in IPL. At first glance, such findings may appear to contradict, rather than support, the discourse that Indigenous Peoples positively contribute to biodiversity. We argue that both the data biases and the overall methodological approach in spatial characterizations may actively disfavor IPL and Indigenous knowledge systems.

Biodiversity data strongly correlate with the human processes underpinning data production. Therefore, observed biases reflect the limitations, priorities, and interests of the individuals and institutions generating the data²⁶. For example, GBIF records are concentrated near roads and cities, and have a significantly higher coverage in temperate biomes than in tropics and montane grasslands⁴³. Data distribution can also reflect socio-political conditions, such as armed conflicts⁴⁴ or racist policies⁴⁵. We found that a considerable share of the data in both the Red List and GBIF has been generated or hosted by non-Bolivian institutions. This mismatch between plant locations and the institutions responsible for data generation and hosting reflects a broader pattern linked to colonial legacies in global specimen collections^{46,47}. Even in monitoring efforts that emphasized the participation of Indigenous Peoples, Indigenous Peoples were rarely involved in decision-making processes such as program initiation, survey design, or objective definition⁴⁸. In addition, even though this is not the case in Bolivia, research permits to sample species in Amazonian countries often do not include authorization to enter IPL¹¹. These factors may contribute to the underrepresentation of IPL areas in these datasets.

Despite these insights, the possibility that (Indigenous) Bolivian scientists actively participated in generating the data used in this case study cannot be ruled out. While the metadata allowed us to identify the studies that generated the data or the institutions that hosted it, it does not provide information that indicates the limitations, priorities, and preferences that may have shaped the observed data distribution. Including such information could improve transparency regarding existing data limitations and allow for more nuanced analyses. Furthermore, recognizing the contributions of Indigenous Peoples is a crucial part of addressing power imbalances in science and environmental management, as it increases visibility and supports more equitable benefit sharing⁴⁹.

One additional aspect that may disadvantage Indigenous Peoples in spatial characterizations is that characterizing species distributions based on global databases is grounded in Western, rather than Indigenous, knowledge systems. While the accumulation of biodiversity data is essential for informing policy within a Western paradigm of global resource management, the standardization and simplification required to integrate data from multiple sources can erase details that are fundamental for understanding ecological systems and interconnected cultural dynamics¹⁹. The environmental stewardship of Indigenous Peoples cannot be comprehensively captured through species counts alone; it must also account for the complex relationships between Indigenous Peoples and their environments⁵⁰. As emphasized by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), recognizing and respecting the contribution of Indigenous knowledge requires, among other things, collaboration and knowledge exchange⁵¹. By contrast, reducing complex socio-ecological interrelationships to simplified datasets is inconsistent with the holistic nature of Indigenous knowledge systems⁵² and may therefore obscure, rather than illuminate, them.

Overall, our findings suggest that the results of spatial characterizations are shaped by the underlying datasets and their inherent spatial and taxonomic biases. These biases relate to likely mismatches between the scopes of the studies contributing to the datasets and that of the spatial characterization. As all datasets combine different sources, they all involve mixed geographic and taxonomic scopes, regardless of whether their overall scope matches that of the characterization. Additionally, the utilized datasets do not reflect the contributions of Indigenous Peoples, complicating accurate interpretation of the spatial characterization results, and weakening the visibility of Indigenous Peoples' contributions.

We recommend that future studies seeking to characterize spatial patterns of species richness use data that are well aligned with the specific objectives of the study, particularly in terms of geographic scope and taxonomic coverage. It is also paramount to account carefully for potential sources of bias in data collection and interpretation. In studies focused on biodiversity within IPL, we underscore the importance of Indigenous participation, not only as a matter of empowerment and ethical engagement, but also to enhance the accuracy, relevance and balance of biodiversity data.

Methodology

Datasets used

Biodiversity data for our analyses were retrieved from three different databases: the IUCN Red List of Threatened Species, the database of the Global Biodiversity Information Facility (GBIF), and the database of the Group on Earth Observations Biodiversity Observation Network (BBON). We used the former two datasets because they have been used in spatial characterizations found in the literature (see refs. 10–15), and the database by BBON to include one dataset that corresponds to the scale of our study area. For the Red List, we retrieved species distribution maps as vector data from Version 6.3 of the IUCN Red List¹⁶ which characterize known and possible present and historical tree species ranges. For GBIF, we retrieved all the occurrence data for Bolivia from the website⁵³ as a cvs-file. For BBON, the complete occurrence dataset originating from Fernández et al.³² was provided as vector data by author M.F. We used the ecosystem classification of Bolivia by Ibisch et al.³¹ as a reference throughout the paper, to compare the observed spatial and taxonomic patterns with expected distributions.

We used vector data indicating the boundaries of Bolivia, PA, and IPL to map the different area types. The Bolivian boundaries were obtained from Natural Data (*naturaldata.com*), the PA boundaries from the World Database on Protected Areas (WDPA)⁵⁴, and the IPL boundaries from the National Agrarian Reform Institute (INRA, after the name in Spanish)³⁸. The global vector dataset of Dinerstein et al.⁵⁵ was used to map the ecoregions of Bolivia.

Data preparation

All preparations and analyses were carried out in Python⁵⁶. A scheme representing the structure of the codes can be found in Figure S1. For GBIF and BBON, we removed all entries lacking either species names or geographic coordinates. We used each latitude and longitude pair provided in the csv-file by GBIF to convert it into vector data. Then, we filtered out all entries outside of the Bolivian borders. We filtered tree species in the Red List, the GBIF, and the BBON datasets by using the tree species list provided by Keppel et al.²¹. This list is based on the GlobalTreeSearch⁵⁷ and the GIFT⁵⁸ databases and is the most complete tree list available, including the scientific names of 58,044 globally known tree species. It is important to note that the list defines trees not based on taxonomy but based on the definition by the Global Tree Specialist Group of the IUCN. This definition includes all woody plants that grow at least two meters tall and have at least one stem with a diameter of five centimeters or more at breast height.

Different from Seebens et al.¹⁵, we did not implement measures to account for spatial biases in the GBIF dataset. We adopt this approach to present the unmodified data distribution. Also, such measures can be contentious, particularly given that important sources of bias, such as systemic inequities or human preference, are challenging to quantify and adjust for ref. 26.

For all five datasets, we extracted entries located within Bolivia using the national boundary of Bolivia. All datasets were projected to WGS 84/UTM zone 20S to enable spatial analyses in meters. For the maps, we reprojected them to the WGS 84 dataset for more comprehensive axis descriptions.

Spatial characterizations

The IPL and PA area types were defined by the datasets adopted. Dual IPL-PA status areas are those where IPL and PA intersect. All other areas were defined as OA. The total areas of Bolivia and each area type were identified through the respective geometries. The relative areas were then calculated by dividing the area type sizes by that of Bolivia.

We characterized the expectable tree species richness across the ecoregions of Bolivia by mapping tree species richness estimations by Ibisch et al.³¹, using their vector data. Ibisch et al.³¹ estimated the trees species richness for most of the ecoregions they characterized, either as ranges or as thresholds (e.g., <50, 250–450, or >1000). We used these estimations to group 18 of the 23 (sub-) ecoregions into three equally sized categories of relative species richness: “low”, “medium”, and “high”. Five subregions were

grouped as “no info” because Ibisch et al.³¹ did not estimate their tree species (see Table S2). Mapping the estimated distribution of tree species provides a baseline for subsequent qualitative evaluation of spatial data biases in the Discussion.

For the species distribution maps, we created one raster layer per dataset indicating the number of species in each pixel of a 10 km² grid. Therefore, we created a separate raster layer for each species, where each grid cell overlapping with the vector data of the species was assigned a value of 1 and each other a value of 0. For the Red List, we directly used the species range entries for this, as each of them represents one species. For GBIF and BBON, we grouped the occurrences into separate sets (tuples), with each set containing all the records for a single species. Finally, we added up the raster layers of the individual species to generate a single layer per dataset.

We determined how many species are present in the different area types by counting the number of species, represented by species range polygons (for the Red List) or species occurrence tuples (for GBIF and BBON), that intersected with the polygons of the area types. As mask layers, we used the IPL and PA boundaries, prepared as previously described. For IPL-PA, we created the mask layer by intersecting IPL and PA. For OA we first combined IPL and PA and then defined the mask layer as all regions of Bolivia that are outside this combination.

The relative species ranges ($range_{rel}$ [ha/species/km²]) based on the Red List in the different area types were determined by summing the species range areas ($range_i$ [m²]) of all species i within the defined mask layers. Then, we divided this total species range by the area of the respective mask layer (A , [m²]) and the total number of species (n), and multiplied everything by 10² to convert the unit to hectare per species and square kilometer (see Eq. 1).

$$range_{rel} = \left(\sum_i range_i \right) / n / A * 10^2 \quad (1)$$

We calculated the species density (ρ_{sp} [number/1000 km²]) based on GBIF and BBON, defined as the number of distinct species per thousand square kilometers by extracting all occurrence points within each of the mask layers. Then, we counted the number of distinct species (n_{sp} [number]) present in each area. Finally, we divided n_{sp} by A to obtain ρ_{sp} and multiplied the result by 10⁹ to convert the unit to number per 1000 km², as shown in Eq. 2.

$$\rho_{sp} = n_{sp} / A * 10^9 \quad (2)$$

Spatial distribution

We mapped the spatial distribution of the species richness across the different ecoregions of Bolivia to compare it to the estimated tree richness distribution by Ibisch et al.³¹. Therefore, we measured how many species ranges of the Red List and how many occurrences of distinct species from GBIF and BBON are within each ecoregion of Bolivia. For the Red List, we counted the intersections between each ecoregion and all species range polygons. For GBIF and BBON, we extracted all occurrence points located within each ecoregion and then counted the number of distinct species.

A raster layer representing the spatial distribution of the observation numbers was created for both GBIF and BBON to show the clustering of the datasets. As these raster layers are rectangular, empty cells could either be cells outside Bolivia or in areas within Bolivia with no observations. To account for this, we identified the number of grid cells in the raster layer corresponding to the shape of Bolivia by masking it to the borders of Bolivia. The minimum number of cells having half of the observations was found iteratively based on a sorted list of cells derived from these raster layers.

Taxonomic distribution

We evaluated the taxonomic distribution of species in the datasets by organizing the species by their families. For GBIF and BBON, we first grouped all entries by species and then by family. For the Red List, we

grouped entries directly by family, as each entry represented a distinct species. To measure the average difference in family distribution between (i) the Red List and GBIF, (ii) the Red List and BBON, and (iii) BBON and GBIF, we first identified the families shared by each pair of datasets. We then filtered each dataset to include those shared families only and sorted them alphabetically to ensure matching order. Finally, we calculated the ratio of the higher to the lower percentage for each family and identified the mean.

Identifying the sources

We used the metadata provided by each dataset to gather information about the data sources. For the Red List, we used the source, citation, and compiler name, and for GBIF, we used the institution and collection. Red List entries were grouped by source, and similar descriptions were merged when they appeared to originate from the same institutions or studies. Based on this grouping, we researched the associated sources and used the expert knowledge of M.M.R. to determine whether they were affiliated with Bolivian institutions. For the GBIF dataset, we automated the process to count the number of entries per institution, per collection, and per collection-institution-pair. For the visualization in Fig. 4, we applied a threshold of 10,000 entries and grouped institutions contributing fewer as “other”. As with the Red List, we used all available metadata to identify the origin of institutions or studies and determined whether they were Bolivian.

Data availability

All datasets used in this study, except for those of the Group on Earth Observations Biodiversity Observation Network³² and the spatial data by Ibisch et al.³¹, are publicly available and can be obtained from the referenced sources. We provide an explanation in the Supplementary Information on how to integrate these datasets into our code structure to reproduce the assessments. All results discussed in the article are shared either in the main manuscript or in the Supplementary Information, in some cases with thresholds applied to ensure clarity. They can be fully reproduced using our code.

Code availability

All code used in this study is available on Zenodo: <https://doi.org/10.5281/zenodo.18467099>.

Received: 4 February 2026; Accepted: 10 August 2026;

Published online: 22 August 2026

References

1. Jaureguiberry, P. et al. The direct drivers of recent global anthropogenic biodiversity loss. *Sci. Adv.* **8**, eabm9982 (2022).
2. Cardinale, B. J. et al. Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
3. Díaz, S. et al. Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
4. Arora, N. K. & Mishra, I. Life on land: progress, outcomes and future directions to achieve the targets of SDG 15. *Environ. Sustain.* **7**, 369–375 (2024).
5. Secretariat of the Convention on Biological Diversity. *Global Biodiversity Outlook 5* <https://www.cbd.int/gbo/gbo5/publication/gbo-5-en.pdf> (2020).
6. Gadgil, M., Berkes, F. & Folke, C. Indigenous knowledge for biodiversity conservation. *Ambio* **22**, 151–156 (1993).
7. Maffi, L. Linguistic, cultural, and biological diversity. *Annu. Rev. Anthropol.* **34**, 599–617 (2005).
8. IPBES. *Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services* <https://doi.org/10.5281/zenodo.3831673> (2019).
9. Fernández-Llamazares, Á et al. Scientists' Warning to Humanity on Threats to Indigenous and Local Knowledge Systems. *J. Ethnobiol.* **41**, 144–169 (2021).
10. O'Bryan, C. J. et al. The importance of Indigenous Peoples' lands for the conservation of terrestrial mammals. *Conserv. Biol.* **35**, 1002–1008 (2021).
11. Fernández-Llamazares, Á et al. The importance of Indigenous Territories for conserving bat diversity across the Amazon biome. *Perspect. Ecol. Conserv.* **19**, 10–20 (2021).
12. Sze, J. S. et al. Indigenous Peoples' Lands are critical for safeguarding vertebrate diversity across the tropics. *Glob. Change Biol.* **30**, e16981 (2024).
13. Estrada, A. et al. Global importance of indigenous peoples, their lands, and knowledge systems for saving the world's primates from extinction. *Sci. Adv.* **8**, eabn2927 (2022).
14. Schuster, R., Germain, R. R., Bennett, J. R., Reo, N. J. & Arcese, P. Vertebrate biodiversity on indigenous-managed lands in Australia, Brazil, and Canada equals that in protected areas. *Environ. Sci. Policy* **101**, 1–6 (2019).
15. Seebens, H. et al. Biological invasions on Indigenous peoples' lands. *Nat. Sustain.* **7**, 737–746 (2024).
16. IUCN. The IUCN Red List of Threatened Species. (2022).
17. GBIF Secretariat. What is GBIF? <https://www.gbif.org/what-is-gbif>.
18. *Guidelines for the Application of IUCN Red List of Ecosystems Categories and Criteria: Version 2.0.* (IUCN, International Union for Conservation of Nature and Natural Resources, Gland, Switzerland, 2024).
19. Devictor, V. & Bensaude-Vincent, B. From ecological records to big data: the invention of global biodiversity. *Hist. Philos. Life Sci.* **38**, (2016).
20. Buckland, S. T. & Johnston, A. Monitoring the biodiversity of regions: Key principles and possible pitfalls. *Biol. Conserv.* **214**, 23–34 (2017).
21. Keppel, G. et al. Synthesizing tree biodiversity data to understand global patterns and processes of vegetation. *J. Veg. Sci.* **32**, e13021 (2021).
22. Mugal, M. A. et al. Bridging global knowledge gaps in biodiversity databases: a comprehensive data synthesis on tree diversity of India. *Biodivers. Conserv.* **32**, 3089–3107 (2023).
23. Costello, M. J., May, R. M. & Stork, N. E. Can we name Earth's species before they go extinct? *Science* **339**, 413–416 (2013).
24. Lomolino, M. V. & others. Conservation biogeography. *Front. Biogeogr. Dir. Geogr. Nat.* **293**, 293–296 (2004).
25. Hortal, J. et al. Seven shortfalls that beset large-scale knowledge of biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 523–549 (2015).
26. Chapman, M. et al. Biodiversity monitoring for a just planetary future. *Science* **383**, 34–36 (2024).
27. Ibisch, P. L. Bolivia is a megadiversity country and a developing country. in *Biodiversity: A Challenge for Development Research and Policy* (eds Barthlott, W., Winiger, M. & Biedinger, N.) 213–241 (Springer, Berlin, Heidelberg, 2001) https://doi.org/10.1007/978-3-662-06071-1_15.
28. Ibisch, P. L. Biodiversity Conservation in Bolivia: History, Trends and Challenges. in *Environmental Issues in Latin America and the Caribbean* (eds Romero, A. & West, S. E.) 55–71 (Springer Netherlands, Dordrecht, 2005) https://doi.org/10.1007/1-4020-3774-0_3.
29. Moraes, R. M. & Sarmiento, J. Biodiversity in Bolivia. in *Global Biodiversity* (Apple Academic Press, 2018).
30. Gigler, B.-S. *Poverty, Inequality and Human Development of Indigenous Peoples in Bolivia*. (2009).
31. Ibisch, P., Beck, S., Gerkmann, B. & Carretero, A. Ecoregiones y ecosistemas. *Biodivers. La Riqueza de Bolivia* **47**, 88 (2003).
32. Fernández, M. et al. Challenges and opportunities for the Bolivian Biodiversity Observation Network. *Biodiversity* **16**, 86–98 (2015).
33. IUCN SSC Red List Technical Working Group. *Mapping Standards and Data Quality for the IUCN Red List Spatial Data* <https://www.iucnredlist.org/resources/mappingstandards> (2024).
34. Chambers, J. Q., Higuchi, N. & Schimel, J. P. Ancient trees in Amazonia. *Nature* **391**, 135–136 (1998).

35. Guèze, M. et al. Are ecologically important tree species the most useful? A case study from indigenous people in the Bolivian Amazon. *Econ. Bot.* **68**, 1–15 (2014).
36. Moraes R. M. *Flora de palmeras de Bolivia*. (Herbario Nacional de Bolivia, Instituto de Ecología, Universidad Mayor de San Andrés, La Paz, Bolivia, 2020).
37. DeWalt, S. J., Bourdy, G., ChÁvez De Michel, L. R. & Quenevo, C. Ethnobotany of the Tacana: Quantitative inventories of two permanent plots of Northwestern Bolivia. *Econ. Bot.* **53**, 237–260 (1999).
38. INRA. Tierras Comunitarias de Origen y Territorio Indígena Originario Campesino a nivel Nacional—Gestión 2018. Instituto Nacional de Reforma Agraria (2018).
39. Beck, J., Böller, M., Erhardt, A. & Schwanghart, W. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecol. Inform.* **19**, 10–15 (2014).
40. Moraes, R. M., Zenteno-Ruiz, F. S., & Fuentes, C. A. F. Árboles de Bolivia: Actualización e implicaciones del conocimiento. **13**, 1–90 (2017).
41. Goettsch, B. et al. High proportion of cactus species threatened with extinction. *Nat. Plants* **1**, 1–7 (2015).
42. Cámara-Leret, R., Tuomisto, H., Ruokolainen, K., Balslev, H. & Munch Kristiansen, S. Modelling responses of western Amazonian palms to soil nutrients. *J. Ecol.* **105**, 367–381 (2017).
43. Hughes, A. C. et al. Sampling biases shape our view of the natural world. *Ecography* **44**, 1259–1269 (2021).
44. Zizka, A. et al. Bio-Dem, a tool to explore the relationship between biodiversity data availability and socio-political conditions in time and space. *J. Biogeogr.* **48**, 2715–2726 (2021).
45. Ellis-Soto, D., Chapman, M. & Locke, D. H. Historical redlining is associated with increasing geographical disparities in bird biodiversity sampling in the United States. *Nat. Hum. Behav.* **7**, 1869–1877 (2023).
46. Faxon, H. & Chapman, M. Beyond spatial bias: Understanding the colonial legacies and contemporary social forces shaping biodiversity data. *Environ. Res. Lett* <https://doi.org/10.1088/1748-9326/add6b6>. (2025).
47. Park, D. S. et al. The colonial legacy of herbaria. *Nat. Hum. Behav.* **7**, 1059–1068 (2023).
48. Thompson, K.-L., Lantz, T. C. & Ban, N. C. A review of Indigenous knowledge and participation in environmental monitoring. *Ecol. Soc.* **25**, art10 (2020).
49. Reyes-García, V. et al. Data sovereignty in community-based environmental monitoring: toward equitable environmental data governance. *BioScience* **72**, 714–717 (2022).
50. Fernández-Llamazares, Á et al. No basis for claim that 80% of biodiversity is found in Indigenous territories. *Nature* **633**, 32–35 (2024).
51. Hill, R. et al. Working with Indigenous, local and scientific knowledge in assessments of nature and nature's linkages with people. *Curr. Opin. Environ. Sustain.* **43**, 8–20 (2020).
52. Tengö, M., Brondizio, E. S., Elmqvist, T., Malmer, P. & Spierenburg, M. Connecting diverse knowledge systems for enhanced ecosystem governance: the multiple evidence base approach. *AMBIO* **43**, 579–591 (2014).
53. GBIF. GBIF Occurrence Download. (2025).
54. UNEP-WCMC and IUCN. Protected Planet: The World Database on Protected Areas (WDPA) [Online]. UNEP-WCMC and IUCN (2025).
55. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).
56. Van Rossum, G. & Fred L. D. Jr. Python 3 Reference Manual. Python Software Foundation (2009).
57. Beech, E., Rivers, M., Oldfield, S. & Smith, P. P. GlobalTreeSearch: The first complete global database of tree species and country distributions. *J. Sustain.* **36**, 454–489 (2017).
58. Weigelt, P., König, C. & Kreft, H. G. I. F. T. A Global Inventory of Floras and Traits for macroecology and biogeography. *J. Biogeogr.* **47**, 16–43 (2020).

Acknowledgements

We thank Jocelyne Sze and Jordi Cristobal Rosello for inspiring discussions in earlier stages of this work. This study has received financial support from an ERC Starting Grant from the European Union (ERC, IEKCHANGES, 101117423) to Á.F.-L. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union and the European Research Council Executive Agency. Neither the European Union nor the grant authority can be held responsible for them. Á.F.-L. was further supported by a Ramón y Cajal research grant from the Spanish Ministry of Science and Innovation (RYC2021-034198-I). This work contributes to ICTA-UAB 'María de Maeztu' Programme for Units of Excellence of the Spanish Ministry of Science and Innovation (CEX2024-001506-M/funded by MICIU/AEI/ 10.13039/501100011033).

Author contributions

S.K. conceptualized the study, curated the data, developed the methodology and software, conducted the formal analysis and investigation, and wrote the original draft. J.C.L. contributed to reviewing and editing the manuscript. M.F. provided data and contributed to reviewing and editing the manuscript. M.M.R. contributed to reviewing and editing the manuscript. Á.F.-L. conceptualized the study and contributed to reviewing and editing the manuscript. All authors have read and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44185-026-00153-6>.

Correspondence and requests for materials should be addressed to Sven Kock.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.