

# Protected Area Networks Do Not Represent Unseen Diversity

Angel Delso (✉ [angeldelso@hotmail.com](mailto:angeldelso@hotmail.com))

Menéndez Pelayo International University

Jesús Muñoz

Real Jardín Botánico

Javier Fajardo

World Conservation Monitoring Centre

---

## Research Article

**Keywords:** prioritization, unseen diversity, representativeness, protected areas, conservation, systematic conservation planning

**Posted Date:** January 20th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-145219/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Most existing protected area networks are usually biased to protect charismatic species or showy landscapes. We hypothesized that conservation networks designed including unseen diversity – groups usually species-rich but consisting of inconspicuous taxa, or affected by knowledge gaps – would be more efficient than networks ignoring those groups. To test this hypothesis, we created species distribution models for 3,006 species of arthropods and determined which were represented in three networks of different size and biogeographic origin. We assessed the efficiency of each network using spatial prioritization to measure its *completeness* – increment needed to achieve conservation targets – and *specificity* – how much overlap the priority areas based on unseen diversity with existing networks. We find that representativeness of unseen diversity in existing protected areas – *extrinsic representativeness* – was low, as ~40% of unseen diversity species were unprotected. We also find that existing networks should be expanded by an additional ~26-46% of their current area to complete targets, and that specificity can be as low as 8.8%, meaning that existing networks are not efficient to conserve unseen diversity. We conclude that information on unseen diversity must be included in systematic conservation planning approaches to design more efficient and ecologically representative protected areas.

# Introduction

Biodiversity crisis is now beyond discussion <sup>1,2</sup>. Around one million of all described species (25% of all the assessed species) has been estimated to be threatened <sup>3</sup>, a striking figure that emphasizes that increasing conservation efforts are crucially needed. One of the most successful approaches to stop species loss and ecosystem degradation is the declaration of Protected Areas (PA) <sup>4</sup>. In recognition for this, nations worldwide convened the target to protect at least 17% of all lands and 10% of seascapes by 2020 through ecologically representative systems of PA <sup>5</sup>. Spurred by that goal, by 2018 PA networks reached a global coverage of 14.9% of terrestrial lands <sup>6</sup>. However, there is mounting concern that satisfactory biodiversity outcomes will not be achieved by 2020 in countries and regions that will not meet the goal, but also in many that will do it <sup>6-8</sup>. This stresses the importance of networks being ecologically representative and illustrates the complex endeavor of designating new PA, which must reconcile opposing views under stressful shortages of funds and time. This concern has been raised in the design of targets within the post-2020 biodiversity framework <sup>9</sup>, arguing that new targets should focus on increasing the effectiveness of biodiversity conservation, emphasizing the value of lands selected for preservation <sup>10</sup> to promote quality areas that enhance network properties such as representation.

Since the amount of area that can be protected is limited, it is crucial that the sites selected for conservation are representative of the three elements of biodiversity, genetic, species, and ecosystem diversity, and efficient <sup>11-13</sup>. However, most existing PA were designed following criteria different to representativeness, such as aesthetic, economic and socio-political criteria <sup>11</sup>. As a result, the global PA

network is biased towards certain biomes and ecoregions<sup>4</sup>, overrepresenting high elevation or remote places where conflict with alternative land uses is low<sup>14</sup>, and resulting in PA networks of general doubtful effectiveness with regard to conservation of biodiversity representativeness e.g.,<sup>15,16</sup>. More recently, the scope and design of PA networks in the last few decades have shifted towards more scientifically sound objectives, including representativeness<sup>17</sup>. A major milestone in the incorporation of biodiversity representativeness in conservation planning is represented by the mainstreaming and formalization of the principles of Systematic Conservation Planning<sup>12</sup>, a framework aiming at promoting efficient conservation decision making based in explicitly declared scientific targets.

When representativeness has been an explicit criterion on PA designation, it has been usually based on the most studied or charismatic taxonomic groups, such as mammals, birds or vascular plants, or coarse filter biodiversity surrogates such as ecosystems<sup>11,18,19</sup>. As a result, PA networks representativeness is biased towards taxa that usually are among the least diverse<sup>19</sup>, and which is worse, are poor surrogates for representativeness of highly diverse but less showy or less studied taxa, such as arthropods, mollusks or annelids<sup>18,20-22</sup>. The latter groups, among many others, account for most of the “unseen” or “hidden” diversity, and are rarely, if at all, included in conservation decisions<sup>23,24</sup>, breaching the criterion of representativeness. This problem is compounded if we consider that extinction rates are higher among the groups that contribute more to the unseen diversity<sup>25-28</sup>. Nonetheless, these taxa are important in conservation planning because they represent the majority of the diversity in any given area, inhabit all types of environments, and fulfil essential functions for the maintenance of many ecosystem services of immense economic value for humans. Indeed, it has been estimated that the annual economic value of pollination, dung burial, pest control and recreation services provided just by the insects annually represent \$57 billion in USA<sup>29</sup>, and other authors have estimated that at the national scale, the pollination service provided by insects range from 1–16% of the market value of agricultural production<sup>30</sup>. And only to mention another crucial nature’s contribution to people, although challenging to value, the diversity of saprotrophic fungi and soil invertebrates (e.g., nematodes, mites, collembolans, annelids, myriapods, etc.), all of them unseen diversity elements, as well as their intricate interactions, are crucial for the maintenance of soil fertility<sup>31</sup>.

At least to a certain extent, the reason why unseen diversity has been excluded from conservation planning is the knowledge gap for not-so charismatic groups. Two types of shortfalls affect this gap and hinder the ability of planners to include these groups in conservation planning: there are still many undescribed species –the “Linnean shortfall”–, and knowledge about the distribution of many species is patchy –the “Wallacean shortfall”<sup>32</sup>. However, spatial information for the groups of the unseen diversity is growing (i.e. insect records added to GBIF per year grew from 1.6 million in year 2000 to 5.8 million in year 2014, cf. [www.gbif.org](http://www.gbif.org)). Such raw distribution data can be used in systematic conservation planning as they are, despite their limitations, but they are even more valuable when combined with the use of species distribution models SDM,<sup>33</sup>. SDM relate georeferenced observations of well-identified individuals, although limited in number, to relevant ecological predictors, to produce suitability maps that show where

the species might be found, which combined with actual species distribution will reduce the existent information gap.

Once there is enough information of species' distributions, either actual data points or SDM, it is possible to test the ability of PA networks to represent different elements of biodiversity. This type of analysis typically start by conducting a representation analysis, where conservation features are classified as either represented or not represented by a given PA network based on the overlap among their distributions<sup>12</sup>. Here, we define *intrinsic representativeness* as the degree of representation of those elements of biodiversity that were considered by planners when designing a given network. As the notion of conservation planning with the aim of representing the elements of biodiversity is relatively new, and acknowledging that it has not been considered during the definition of many existing PA, we extend this concept of *intrinsic representativeness* to a vaguer notion that includes elements of biodiversity commonly represented in conservation planning, typically terrestrial vertebrates and charismatic or umbrella species. In fact, these represent the lion's share of the funds dedicated to biodiversity conservation<sup>34</sup>. The question is whether the so generated PA also maximize the representativeness of other elements such as the unseen diversity, typically not considered in the design process. We therefore define *extrinsic representativeness* as the degree of representation of these groups. In addition to this, it can be important to inquire whether a PA network is efficient representing a group of species such as the unseen diversity. Systematic conservation planning offers several methods to identify clusters of priority conservation areas that optimize representativeness for a given group of species (i.e. maximize the percentage of the species' distributions included, while minimizing costs). These methods could be used to estimate specific shortfalls of current PA networks through exploring which areas would be required as an addition, or could serve to identify group-specific optimal representation areas, which can be contrasted against actual PA to assess their specificity by determining the degree at which both of them overlap.

The objective of this study is thus to assess the potential of existing PA networks to correctly represent the unseen diversity. Specifically, we test (1) the **extrinsic representativeness** (i.e. of taxa not used for designation) of existing PA networks, and (2) the overall **efficiency** of the existing PA from two perspectives: (i) *representation completeness*, where we explore how much area would be needed expanding current PA to represent the unseen diversity; and (ii) *representation specificity*, where we test the match between the most efficient conservation areas designed using exemplary groups of unseen diversity and the existing PA networks.

Our hypothesis is that the almost-exclusive use of charismatic groups in systematic conservation planning does not result in truly representative PA networks, but rather it is crucially important to use unseen diversity groups to increase both intrinsic and extrinsic representativeness in PA networks.

## Results

The extrinsic representativeness of existing PA was ~ 60% for the three testing areas, with slight differences among countries (Fig. 1). Average extrinsic representativeness was higher for Costa Rica (62.28%), followed by Mexico (60.47%), and USA (56%). Average representativeness also varied across taxa by countries (Fig. 1): 56% (Hymenoptera), 54% (Lepidoptera) and 42% (Coleoptera). Coleoptera was the worst represented order in all sites, and also the one with more neglected species. The best represented order in USA and Costa Rica was Lepidoptera, while Hymenoptera was the best represented in Mexico, with more than 75% of the analyzed species satisfactorily included by PA. Neglected species were fewest in Costa Rica (< 5%), but increased to 25% of the Coleoptera and 17% of the Lepidoptera in Mexico, and ~ 20% for the three taxa in USA.

Efficiency of the existing PA networks also differed by country and taxa for the two perspectives evaluated (Table 1). *Representation completeness* was lowest for Mexico, which would require an expansion of 46.34% its current PA extent to achieve unseen diversity targets. For Costa Rica and USA completeness was higher, although both need an increase of around a third of their existing PA coverage (26.28 and 29.82%, respectively) to fully represent unseen diversity if existing PA are included in the proposed solution (Fig. 2).

*Representation specificity* was highest for Costa Rica, with ~ 50% of the priority areas for unseen diversity overlapping with the existing network, but dropped to 20.1% for Mexico and plummeted to 8.8% for USA.

Table 1

Extrinsic representativeness, completeness, and overlap of priority areas for unseen diversity with current PA, or representation specificity.

		Costa Rica	Mexico	USA
<b>Extrinsic representativeness</b>	unseen diversity targets achieved by the existing protected area network (%)	62.28	60.47	56
<b>Efficiency</b>	<b>Completeness</b> (%)	73.72	53.66	70.18
	<b>Specificity</b> (% overlap with existing PA)	49.88	20.1	8.8

## Discussion

This study provides the first wide scale attempt to assess unseen diversity representativeness and PA representativeness efficiency. Using three PA networks of varied characteristics, we demonstrated considerable gaps in existing PA to protect highly diverse taxa usually ignored in systematic conservation planning and a limited efficiency of PA representing this diversity. Almost half of the analyzed species were not represented at all in existing PA networks, and around 15% of them were neglected, demonstrating that extrinsic representativeness is deficient for unseen diversity. To improve extrinsic

representativeness (Fig. 2), we showed that PA networks need to incorporate large amounts of area, or to change the PA network configuration substantially, which is likely the result of unseen diversity hardly playing a role in former PA planning and designation.

Our finding that about 40% of the unseen diversity species analyzed are not represented at all in current PA networks had already been demonstrated by previous studies focused on other unseen diversity groups. D'Amen, et al. <sup>35</sup> found that 87% of the studied saproxylic beetles were not represented in the Italian PA network, Martín-Piera <sup>36</sup> reported that 5 endemic dung beetle species were underrepresented in Spain, and a previous study with dung beetles showed conservation gaps in Costa Rica <sup>23</sup>. In terms of funding, a recent study has shown that European Union conservation funds are strongly biased to charismatic species, with birds and mammals alone accounting for 72% of species and 75% of the total budget, and that even the two species receiving the most money (brown bear and grey wolf) are not threatened according to the IUCN <sup>34</sup>. This lack of representativeness coverage by existing PA networks also exists for other taxa, but not as severely as for unseen diversity groups. For example, a study in Costa Rica showed that just 25% of the conservation targets for mammal species were not achieved by its PA network, even though the authors used more demanding targets than we did <sup>37</sup>, and only 18% of Mexican mammals were unprotected by current PA <sup>38</sup>. This demonstrates that current PA networks, although most of the times not explicitly defined based on charismatic taxa, do a better job at protecting these groups than they do in the case for less showy groups. Relying more or less exclusively on well-known and charismatic species introduces an additional bias for poorly known systems for which species' data are scarce. This is the case of terrestrial ecosystems in poorly studied regions such as humid tropical forests, but also for "next door" ecosystems which are almost completely composed of unseen diversity groups, such as fresh waters, often neglected in conservation plans despite the existence of thorough systematic conservation planning studies focused on them <sup>39</sup>.

These differences reinforce the idea that conservation gaps are higher for species not included in systematic conservation planning analyses, and also challenge the capacity of charismatic groups to truly represent unseen diversity distribution and richness patterns, as shown by Escalante, et al. <sup>40</sup>. Although our study is not a direct comparison of representation among seen and unseen diversity, our results support the existence of this bias, as suggested by previous research. Spatial distributions of vertebrate and ant diversities do not match in Florida <sup>20</sup>, mosses in Europe show a reversed latitudinal richness pattern as compared to vascular plants <sup>41</sup>, freshwater invertebrates in the Amazon of Ecuador peak at higher elevations than vertebrates <sup>42</sup>, and beetle and polypore species richness is poorly related to bird species richness in boreal forests <sup>43</sup>. There is thus mounting evidence that it is essential to include unseen diversity in systematic conservation planning approaches for PA to be truly representative of biodiversity. If distribution data of inconspicuous taxa is not available or poor, close surrogates should be included, such as better known species of the same family or genus <sup>44</sup>, but not species of taxonomically unrelated groups.

Our methods compound a novel framework to assess the efficiency of PA networks when representing subsets of biodiversity, here focused on what we have termed extrinsic representativeness. The framework is conceived around two complementary spatial analyses involving the identification of optimal areas for the representation of certain target group of conservation features, which can be obtained with the help of a site-selection algorithm. The first analysis (*completeness*) aims at assessing how far is a network of PA from complete representation of the target group. To assess this, we start by identifying the optimal areas needed to expand currently standing PA to achieve full representation of the targets, and we follow by calculating the magnitude of the required expansion. This allows for identifying PA networks that are closer to completion (in this particular aspect). The gap to completion of a PA network will be the result of a combination of factors. The efficiency of the network representing the target group is an important one, but other factors related to biodiversity such as species richness and turnover, as well as current PA extension are also relevant. For instance, regions with high species richness and beta diversity will potentially need larger expansions. These regions are known to require protecting large extensions, because its biodiversity has, in general, few overlaps due to the high replacement, leaving out few complementary opportunities<sup>45,46</sup>.

The second spatial analysis (*specificity*) evaluates the network's efficiency representing the target group by measuring how well captured are the optimal areas for the protection of its elements in PA. The idea is that highly congruent networks are indicative of high specificity in the representation of the target group, whereas a general spatial mismatch indicates poor specificity. This method has been commonly applied in the literature to evaluate PA representation efficiency for whole countries or regions<sup>47,48</sup>. However, these analyses typically target at maximizing the representation of common taxonomic groups such as terrestrial vertebrate species for which comparably more spatial data is available. Here, we propose adapting this analysis to other taxa such as the unseen diversity to assess efficiency capturing extrinsic representativeness. In theory, the efficiency representing groups of species will depend largely on how distinctive is the distribution pattern of the evaluated group compared to the pattern of those elements that have been historically more influential on PA declaration. In this context, it could be expected that conducting this analysis for taxonomic groups or conservation features with very unique distributions (e.g. freshwater diversity, uncharismatic species, soil biodiversity or agrodiversity, among others) have potential to reveal low specificity levels of extrinsic representation. Lastly, considering simultaneously *completeness* and *specificity*, combined with standard representation gap analysis, provides the means for a comprehensive evaluation of the efficiency of PA representing a group of conservation features. For instance, in our results, the gap to completion of the networks in Costa Rica and USA is similar, as both require an equivalent expansion in percent terms (Table 1). However, it should not be assumed that the PA network of Costa Rica is less efficient than that of USA because it still needs a large expansion when its network is already double in size in comparison (Table 2). Indeed, the specificity analysis indicates that PA in Costa Rica are more specific to represent unseen diversity (Table 1). Therefore, the proposed framework is able to diagnose different types of shortfalls: PA in the USA are inefficient representing the unseen diversity because they are largely unspecific to this objective and because they fall short in extent.

Contrastingly, the network in Costa Rica was the most specific, but its biodiversity is so numerous that it still needs a considerable expansion to efficiently represent all the evaluated species.

Unseen diversity plays important roles in ecosystem maintenance as well as services to society, like nutrient cycling, primary production, soil formation, habitat provision or pollination. We defend that unseen diversity should be explicitly considered to designate more representative but also more functional PA networks, as well as to help designing more efficient restoration strategies. We acknowledge that information is poor or at coarse scale for many of the groups providing these services (e.g., fine-grained distributions of pollinators), but given its importance to nature and society, funding programs to close the taxonomic gap are worth the investment<sup>49–53</sup>. However, until those data are available, we argue that SDM, despite limitations, represent a valuable source of information to overcome data gaps and inform conservation planning. Notwithstanding, a few methodological considerations must be considered. Models represent the potential distribution of species and there are several factors that are not fully reflected in model parameters<sup>54</sup>. The most important would be: (1) evolutionary (geographical barriers and speciation processes), (2) ecological (biotic interactions) and (3) anthropogenic (habitat degradation). This generate overprediction, especially in areas and groups with a low sampling effort<sup>55,56</sup>. We have however minimized potential problems by simulating the effect of geographic barriers<sup>57</sup> and excluding from priority areas sites where ecosystems have been degraded, to ensure that only areas that maintain a certain ecological integrity are considered.

In conclusion, this study assesses the representativeness and representation efficiency of three insect groups, acknowledging they are only a minimal part of the unseen diversity. We hypothesize that the same results found in this study will be found with almost any other taxonomic group of unseen diversity, and for any geographical area and resolution. Therefore, we emphasize the need for increasing efforts at describing and sampling these other 'biodiversities', as well as for mainstreaming their inclusion in future conservation planning.

## Methods

**Testing areas.** We used three countries as testing areas to evaluate representativeness of the unseen diversity in their PA networks: Costa Rica, Mexico, and USA. These countries meet three criteria that allow exploring a richer portfolio of contexts relevant to the research question: (1) high availability of georeferenced data on megadiverse taxa, something uncommon for most countries, (2) follow a latitudinal gradient from tropical to temperate climates associated with a decrease in species diversity, and (3) PA networks have different sizes (Table 2).

**Unseen diversity taxa selection.** Deciding about what taxa of unseen diversity could be used was difficult. Almost by definition, unseen diversity groups are at most poorly represented in public databases (e.g. GBIF), and thus the challenge was to find highly diverse taxa with information on distribution for all of the testing areas. For some groups (e.g., bryophytes) the information was good for an area, but too poor for other(s), and information for most groups was poor worldwide. We explored insects as suitable

candidates, as they include 5–6 of the 11 millions of described animals <sup>19</sup>, are found in almost any terrestrial ecosystem worldwide, and they provide fundamental ecosystem services <sup>29</sup>. Although it is estimated that less than 30% of existent insect species have been formally described <sup>58</sup>, certain insect orders are reasonably well known as well as represented in public databases. Moreover, studies have shown that higher taxa insect richness, such as genera or families, are good surrogates for total insect diversity at species level <sup>44</sup>. Thus, we concluded that insect might be appropriate to guide conservation assessments and suitable for testing extrinsic representativeness coverage in existing PA networks. Finally, insect diversity is reasonably well known for the three testing areas, and they follow the latitudinal gradient criteria we seek (Table 2).

Table 2  
Comparative metrics of the three testing areas.

Sources for insect species richness: Costa Rica <sup>59</sup>, Mexico <sup>60</sup> and USA <sup>61</sup>.

	Costa Rica	Mexico	USA
Current Protected Area (% of total area)	27.60%	14.16%	12.99%
Area (km <sup>2</sup> )	51,100	1.97 * 10 <sup>6</sup>	9.83 * 10 <sup>6</sup>
Insect richness (Described / Estimated)	68,500 / 365,000	47,800 / 97,000	91,000 / 164,000
Number of described insect species / total area	1.341	0.024	0.009

We evaluated representation of 1,002 species for each testing area. Species were randomly selected from three of the most diverse orders of insects, Coleoptera, Hymenoptera and Lepidoptera, excluding exotic species listed in national catalogues <sup>62,63</sup> and the IUCN online database (iucngisd.org). These orders were selected because they have the largest and most comprehensive availability of georeferenced data among insect groups, and because they are diverse, allowing us for representing the greatest heterogeneity of niches and ecoregions. For each order, we downloaded occurrence data from different sources (see below), and discarded species with less than 15 unique occurrences, as these would render poor-performing distribution models <sup>64</sup>. From the remaining pool of species, 334 were randomly selected for each order, adding up to a total of 1,002 species per study area (Supplementary Table 1). In recognition that data availability for the chosen taxonomic classes and countries is highly heterogeneous, we chose to use a random, constant and relatively high number of species across testing areas to counterbalance the impact of biases in our results.

**Species data.** We built a database of georeferenced occurrence points, obtained mainly from the Global Biodiversity Information Facility, for Costa Rica (DOI: 10.15468/dl.zropvg), Mexico (DOI: 10.15468/dl.kgaeqh) and USA (DOI: 10.15468/dl.7vadxq), including records within a buffer of 100 km around the border of each country to avoid artifacts in the SDM related to political borders. The database was complemented with additional records from diverse sources: Heliconiine <sup>65</sup>, insect occurrences from

the Mexican “Datos Abiertos” database (datos.gob.mx), a database for dung beetles from Costa Rica and Mexico <sup>66</sup>, and data for the USA accessed through Biodiversity Information Serving Our Nation (BISON, bison.usgs.gov). To minimize the number of wrong records which typically occur in open databases <sup>67</sup>, we used the package CoordinateCleaner <sup>68</sup> on the R 3.4 environment <sup>69</sup> to remove records located at country centroid, natural history museums or research facilities, with a coordinate uncertainty higher than 1 km, duplicates with identical coordinate values, or records unidentified at the species level.

**Species Distribution Models.** We built SDM to estimate the macroclimatic niche of the species <sup>33</sup> using the ‘biomod2’ package <sup>70</sup> in R environment. We generated models as ensembles of three different techniques considered to have higher prediction accuracy <sup>71</sup>: Generalized Boosted Models <sup>72</sup>, Random Forests <sup>73</sup>, and Maxent <sup>74</sup>. An ensemble approach using these three techniques was preferred to avoid problems related to the selection of a single modelling algorithm, which can influence results <sup>75</sup>. A total of 21 variables were considered as potential predictors, the 19 climatic from Worldclim 2.0 <sup>76</sup>, and 2 related to vegetation: the Normalized Difference Vegetation Index (NDVI) calculated from MODIS (modis.gsfc.nasa.gov) by averaging data from 10-day periods, and the height of the tree canopy <sup>77</sup>. To eliminate multicollinearity between predictors, we calculated the Variance Inflation Factor using the package ‘VIF’ <sup>78</sup>, discarding variables with VIF equal or higher than 10. When possible, we decided which variable was kept considering its ecological relevance for the study group. Eight variables remained in the final set of predictors for each country (Table 3). Due to computational limitations, variables resolution was proportional to the area of the country of study: 1 km for Costa Rica, 5 km for Mexico and 10 km for USA.

Table 3

Variables used in the species distribution models for each country. In parenthesis the code of Worldclim variables ([www.worldclim.org](http://www.worldclim.org))

	Costa Rica	Mexico	USA
<b>Vegetation</b>	NDVI	NDVI	NDVI
	Canopy height	Canopy height	Canopy height
<b>Temperature</b>	Annual mean (bio 1)	Annual mean (bio 1)	Annual mean (bio 1)
	Seasonality (bio 4)	Mean diurnal range (bio 2)	Mean diurnal range (bio 2)
	Annual range (bio 7)	Isothermality (bio 3)	Seasonality (bio 4)
<b>Precipitation</b>	Annual mean (bio 12)	Annual mean (bio 12)	Annual mean (bio 12)
	Mean wettest month (bio 13)	Mean wettest month (bio 13)	Mean wettest month (bio 13)
	Mean driest month (bio 14)	Mean driest month (bio 14)	Seasonality (bio 15)

We built SDM using ten thousand background points randomly selected from within occurrence records of other species of the same order, a method that has shown potential to reduce biases resulting from non-systematic occurrence sampling<sup>79</sup>. For each species, models were fitted with 70% of the presence data and validated with the remaining 30%. To increase robustness, we run 10 replicates for each modelling technique and the resulting mean was used. A final ensemble model was obtained for each species by a weighted averaging of models in each replica. Weights were calculated from each model TSS, using only those with  $TSS > 0.7$ <sup>80</sup>. Only species whose ensemble model had a  $TSS > 0.7$  were kept; species with lower TSS values were discarded. When this happened, a new species was analyzed keeping the final number of analyzed species at 1,002 per country (334 per order).

To limit extrapolations of the model too far from species known occurrences, we modified model outputs to restrict them to areas nearby presences using an exponential decay function as the approach in<sup>57</sup>. This method is intended at reduce overprediction at areas far away from known occurrences, simulating the effect of geographic barriers and species dispersal limitations.

Finally, continuous models were transformed into binary maps (presence/absence) using maximum TSS as threshold.

**Representativeness assessment.** To assess representativeness of PA networks, a conservation target must be defined for each conservation feature (here species). The conservation target is the proportion of distribution area of a given conservation feature that should be included in the PA network in order to be considered represented. In the case of species, it is considered to be the minimum fraction of the distribution area necessary for it to thrive<sup>46</sup>. Here, targets were set inversely proportional to the distribution area size of each species, i.e. species with narrower distributions got higher targets than widespread species. Building on the approach in<sup>81</sup>, conservation targets ranged between 80% for species with a distribution area less than 100 km<sup>2</sup>, and 5% for those with ranges larger than 50,000 km<sup>2</sup>. The lower threshold was set following the B1 criteria (geographical distribution threshold) to declare a species as critically endangered<sup>82</sup>. Conservation targets of species between those thresholds were scaled using a loglinear function<sup>83</sup>.

Once species were assigned a target, we evaluated its achievement in current PA by comparison with the amount of SDM included in PA. Species with a distribution inside PA greater than their target were considered as represented. We classified species not achieving targets in two categories: *underrepresented*, when an area equal to 50–99% of their target was included in the solution, and *neglected*, when the area included in the solution was less than 50% of its target.

**Efficiency assessment.** We tested the efficiency of PA networks to represent unseen diversity from two perspectives: *representation completeness*, to measure the amount of area to be added to the existing PA network to represent all species, and *representation specificity* to measure the overlap between priority areas newly generated to optimize unseen diversity representation and the existing PA network.

The proposed framework to test efficiency requires contrasting current PA networks against “optimal” areas, that were obtained using the R package ‘prioritizr’ 4.1.1 <sup>84</sup>. Prioritizr uses the integer linear programming solver Gurobi Optimizer <sup>85</sup> to find the combination of sites that maximize the number of conservation targets achieved while minimizing costs (e.g. the amount of area or the cost of sites).

**Representation completeness.** To assess networks’ representation completeness, we executed Prioritizr and forced existing PA into the solution, so the algorithm selects additional complementary areas as necessary until all species’ targets were achieved. These solutions can inform about how much additional area is needed if targets are not met, allowing for measuring the efficiency of networks as a function of the amount of area needed to complete the species representation sought.

**Representation specificity.** To assess networks’ representation specificity, a second prioritization was conducted, but with existing PA not forced into the solution. With this setup, the algorithm identifies the optimal set of sites that achieve targets without any prior constraint, which implies that current PA (or portions of them) might or might not be part of the solution. The overlap between current PA and resulting areas allows measuring how existing PA networks are efficient at representing unseen diversity.

## Declarations

### Acknowledgements

We thank the "Centro de Cálculo Científico, SGAI" of the Spanish National Research Council (CSIC, Spain) for allowing us to use the cluster “TRUENO” to create the species distribution models.

### Author Contributions

ÁD, JF and JM conceived the idea and designed the project. ÁD collated the datasets and performed the analysis. ÁD, JF and JM wrote the manuscript.

**Competing Interests:** The authors declare no competing interests.

## References

1. Butchart, S. H. M. *et al.* Global biodiversity: Indicators of recent declines. *Science*. **328**, 1164–1168 <https://doi.org/10.1126/science.1187512> (2010).
2. Ripple, W. J. *et al.* Are we eating the world's megafauna to extinction? *Conservation Letters*, e12627, doi:10.1111/conl.12627 (2019).
3. IPBES. *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Advance Unedited Version.* (Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2019).

4. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. The performance and potential of protected areas. *Nature*. **515**, 67–73 <https://doi.org/10.1038/nature13947> (2014).
5. Convention on Biological Diversity. *Strategic plan for biodiversity 2011–2020 and the Aichi targets*, (2010).
6. UNEP-WCMC, IUCN & NGS. (eds UNEP-WCMC, IUCN, & (Cambridge UK; NGS) & Gland, Switzerland; and Washington, D.C., USA, 2018).
7. Tittensor, D. P. *et al.* A mid-term analysis of progress toward international biodiversity targets. *Science*. **346**, 241–244 <https://doi.org/10.1126/science.1257484> (2014).
8. Butchart, S. H. M. *et al.* Shortfalls and solutions for meeting national and global conservation area targets. *Conservation Letters*. **8**, 329–337 <https://doi.org/10.1111/conl.12158> (2015).
9. Visconti, P. *et al.* Protected Area Targets Post-2020. *Science*. **364**, 239–241 <https://doi.org/10.1126/science.aav6886> (2019).
10. Erdelen, W. R. Shaping the Fate of Life on Earth: The Post-2020 Global Biodiversity Framework. *Global Policy*. **11**, 347–359 <https://doi.org/10.1111/1758-5899.12773> (2020).
11. Possingham, H. P., Wilson, K. A., Andelman, S. J. & Vynne, C. H. in *Principles of Conservation Biology* (eds M. J. Groom, G. K. Meffe, & C. R. Carroll) Ch. 14, 507–549(2006).
12. Margules, C. R. & Pressey, R. L. Systematic conservation planning. *Nature*. **405**, 243–252 (2000).
13. Thomassen, H. A. *et al.* Mapping evolutionary process: a multi-taxa approach to conservation prioritization. *Evol. Appl.* **4**, 397–413 <https://doi.org/10.1111/j.1752-4571.2010.00172.x> (2011).
14. Joppa, L. N. & Pfaff, A. High and far: biases in the location of protected areas. *PLoS One*. **4**, 1–6 <https://doi.org/10.1371/journal.pone.0008273> (2009).
15. Jenkins, C. N., Houtan, K. S. V., Pimm, S. L. & Sexton, J. O. US protected lands mismatch biodiversity priorities. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5081–5086, doi:10.1073/pnas.1418034112 (2015).
16. Fajardo, J., Lessmann, J., Bonaccorso, E., Devenish, C. & Muñoz, J. Combined use of Systematic Conservation Planning, Species Distribution Modelling, and Connectivity Analysis reveals severe conservation gaps in a megadiverse country (Peru). *PLoS ONE*. **9**, e114367 <https://doi.org/10.1371/journal.pone.0114367> (2014).
17. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. The performance and potential of protected areas. *Nature*. **515**, 67–73 <https://doi.org/10.1038/nature13947> (2014).
18. Rodrigues, A. S. L. & Brooks, T. M. Shortcuts for biodiversity conservation planning: the effectiveness of surrogates. *Annual Review of Ecology Evolution and Systematics*. **38**, 713–737 <https://doi.org/10.1146/annurev.ecolsys.38.091206.095737> (2007).
19. Pimm, S. L. *et al.* The biodiversity of species and their rates of extinction, distribution, and protection. *Science*. **344**, 9871246752 <https://doi.org/10.1126/science.1246752> (2014).
20. Allen, C. R., Pearlstine, L. G., Wojcik, D. P. & Kitchens, W. M. The spatial distribution of diversity between disparate taxa: Spatial correspondence between mammals and ants across South Florida, USA. *Landscape Ecol.* **16**, 453–464 (2001).

21. Shokri, M. R., Gladstone, W. & Kepert, A. Annelids, arthropods or molluscs are suitable as surrogate taxa for selecting conservation reserves in estuaries. *Biodiversity and Conservation*. **18**, 1117–1130 (2009).
22. Kremen, C. *et al.* Terrestrial arthropod assemblages: their use in conservation planning. *Conserv. Biol.* **7**, 796–808 (1993).
23. Kohlmann, B., Solís, Á., Elle, O., Soto, X. & Russo, R. Biodiversity, conservation, and hotspot atlas of Costa Rica: a dung beetle perspective (Coleoptera: Scarabaeidae: Scarabaeinae). *Zootaxa*. **1457**, 1–34 <https://doi.org/10.11646/zootaxa.1457.1.1> (2007).
24. Chefaoui, R. M., Hortal, J. & Lobo, J. M. Potential distribution modelling, niche characterization and conservation status assessment using GIS tools: a case study of Iberian Copris species. *Biol. Conservation*. **122**, 327–338 <https://doi.org/10.1016/j.biocon.2004.08.005> (2005).
25. Eisenhauer, N., Bonn, A. & Guerra, C. A. Recognizing the quiet extinction of invertebrates. *Nature Communications*. **10**, 50 <https://doi.org/10.1038/s41467-018-07916-1> (2019).
26. Mckinney, M. L. High rates of extinction and threat in poorly studied taxa. *Conserv. Biol.* **13**, 1273–1281 <https://doi.org/10.2307/2641951> (1999).
27. Sánchez-Bayo, F. & Wyckhuys, K. A. G. Worldwide decline of the entomofauna: A review of its drivers. *Biol. Conservation*. **232**, 8–27 <https://doi.org/10.1016/j.biocon.2019.01.020> (2019).
28. Hallmann, C. A. *et al.* More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS One*. **12**, e0185809 <https://doi.org/10.1371/journal.pone.0185809> (2017).
29. Losey, J. E. & Vaughan, M. The Economic Value of Ecological Services Provided by Insects. *BioScience*. **56**, 311–323 [https://doi.org/10.1641/0006-3568\(2006\)56\[311:TEVOES\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)56[311:TEVOES]2.0.CO;2) (2006).
30. Hein, L. The economic value of the Pollination Service, a review across scales. *The Open Ecology Journal*. **2**, 74–82 (2009).
31. Briones, M. J. I. Soil fauna and soil functions: a jigsaw puzzle. *Frontiers in Environmental Science*. **2**, 7 <https://doi.org/10.3389/fenvs.2014.00007> (2014).
32. Cardoso, P., Erwin, T. L., Borges, P. A. V. & New, T. R. The seven impediments in invertebrate conservation and how to overcome them. *Biol. Conservation*. **144**, 2647–2655 <https://doi.org/10.1016/j.biocon.2011.07.024> (2011).
33. Guisan, A. & Zimmermann, N. E. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**, 147–186 (2000).
34. Mammola, S. *et al.* Towards a taxonomically unbiased European Union biodiversity strategy for 2030. *Proceedings of the Royal Society B: Biological Sciences* **287**, 20202166, [doi:10.1098/rspb.2020.2166](https://doi.org/10.1098/rspb.2020.2166) (2020).
35. D'Amen, M. *et al.* Protected areas and insect conservation: questioning the effectiveness of Natura 2000 network for saproxylic beetles in Italy. *Anim. Conserv.* **16**, 370–378 <https://doi.org/10.1111/acv.12016> (2013).

36. Martín-Piera, F. Area networks for conserving Iberian insects: A case study of dung beetles (col., Scarabaeoidea). *J. Insect Conserv.* **5**, 233–252 (2001).
37. Gonzalez-Maya, J. F., Viquez, R. L., Belant, J. L. & Ceballos, G. Effectiveness of protected areas for representing species and populations of terrestrial mammals in Costa Rica. *PLoS One.* **10**, e0124480 <https://doi.org/10.1371/journal.pone.0124480> (2015).
38. Ceballos, G. Conservation priorities for mammals in megadiverse Mexico: the efficiency of reserve networks. *Ecol. Appl.* **17**, 569–578 (2007).
39. Linke, S., Turak, E. & Nel, J. Freshwater conservation planning: the case for systematic approaches. *Freshw. Biol.* **56**, 6–20 <https://doi.org/10.1111/j.1365-2427.2010.02456.x> (2011).
40. Escalante, T. *et al.* Evaluation of five taxa as surrogates for conservation prioritization in the Transmexican Volcanic Belt, Mexico. *Journal for Nature Conservation.* **54**, 125800 <https://doi.org/10.1016/j.jnc.2020.125800> (2020).
41. Mateo, R. G. *et al.* The mossy north: an inverse latitudinal diversity gradient in European bryophytes. *Sci Rep.* **6**, 25546 <https://doi.org/10.1038/srep25546> (2016).
42. Lessmann, J. *et al.* Freshwater vertebrate and invertebrate diversity patterns in an Andean-Amazon basin: implications for conservation efforts. *Neotropical Biodiversity.* **2**, 99–114 <https://doi.org/10.1080/23766808.2016.1222189> (2016).
43. Similä, M., Kouki, J., Mönkkönen, M., Sippola, A. L. & Huhta, E. Co-variation and indicators of species diversity: Can richness of forest-dwelling species be predicted in northern boreal forests? *Ecol. Ind.* **6**, 686–700 <https://doi.org/10.1016/j.ecolind.2005.08.028> (2006).
44. Báldi, A. Using higher taxa as surrogates of species richness: a study based on 3700 Coleoptera, Diptera, and Acari species in Central-Hungarian reserves. *Basic Appl. Ecol.* **4**, 589–593 <https://doi.org/10.1078/1439-1791-00193> (2003).
45. Lessmann, J., Fajardo, J., Bonaccorso, E. & Bruner, A. Cost-effective protection of biodiversity in the western Amazon. *Biol. Conservation.* **235**, 250–259 <https://doi.org/10.1016/j.biocon.2019.04.022> (2019).
46. Rodrigues, A. S. L. & Gaston, K. J. How large do reserve networks need to be? *Ecol. Lett.* **4**, 602–609 (2001).
47. Bax, V. & Francesconi, W. Conservation gaps and priorities in the Tropical Andes biodiversity hotspot: Implications for the expansion of protected areas. *Journal of Environmental Management.* **232**, 387–396 <https://doi.org/10.1016/j.jenvman.2018.11.086> (2018).
48. Cuesta, F. *et al.* Priority areas for biodiversity conservation in mainland Ecuador. *Neotropical Biodiversity.* **3**, 93–106 <https://doi.org/10.1080/23766808.2017.1295705> (2017).
49. Klein, A. M. *et al.* Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society B: Biological Sciences* **274**, 303–313, doi:10.1098/rspb.2006.3721 (2006).
50. Bauer, D. M. & Wing, I. S. Economic Consequences of Pollinator Declines: A Synthesis. *Agricultural and Resource Economics Review.* **39**, 368–383 <https://doi.org/10.1017/S1068280500007371> (2010).

51. Kevan, P. G. & Phillips, T. P. The economic impacts of pollinator declines: an approach to assessing the consequences. *Conserv. Ecol.* **5**, 8 <https://doi.org/10.5751/ES-00272-050108> (2001).
52. Hérivaux, C. & Grémont, M. Valuing a diversity of ecosystem services: The way forward to protect strategic groundwater resources for the future? *Ecosystem Services.* **35**, 184–193 <https://doi.org/10.1016/j.ecoser.2018.12.011> (2019).
53. Haefele, M., Loomis, J. & Bilmes, L. J. in Valuing U.S. National Parks and Programs. America's Best Investment (eds Linda J. Bilmes & John B. Loomis) 16–44(Earthscan from Routledge, 2019).
54. Kramer-Schadt, S. *et al.* The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity Distrib.* **19**, 1366–1379 <https://doi.org/10.1111/ddi.12096> (2013).
55. Cayuela, L. *et al.* Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. *Tropical Conservation Science.* **2**, 319–352 (2009).
56. Guisan, A. & Thuiller, W. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* **8**, 993–1009 <https://doi.org/10.1111/j.1461-0248.2005.00792.x> (2005).
57. Thornhill, A. H. *et al.* Spatial phylogenetics of the native California flora. *BMC Biology.* **15**, 96 <https://doi.org/10.1186/s12915-017-0435-x> (2017).
58. Stork, N. E., McBroom, J., Gely, C. & Hamilton, A. J. New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proc Natl Acad Sci U S A.* **112**, 7519–7523 <https://doi.org/10.1073/pnas.1502408112> (2015).
59. SINAC. *IV Informe de País al Convenio sobre la Diversidad Biológica.* Vol. 4 (GEF-PNUD, 2009).
60. Llorente Bousquets, J. & Ocegueda, S. in Conocimiento actual de la biodiversidad Vol. 1 (eds Jorge Llorente Bousquets & Susana Ocegueda) Ch. 11, 283–322(CONABIO, 2008).
61. Sabrosky, C. W. The Yearbook of Agriculture Vol. 2 *Ch.* **1**, 1–37 United States Department of Agriculture (1952).
62. Hanson, P. Los insectos invasores de Costa Rica. *Revista Biocenosis.* **22**, 51–60 (2009).
63. March, I. J. & Martínez, M. (eds) (eds Instituto Mexicano de Tecnología del Agua et al.) 1–73(México, Jiutepec, Morelos, 2007).
64. van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J. & Raes, N. Minimum required number of specimen records to develop accurate species distribution models. *Ecography.* **39**, 542–552 <https://doi.org/10.1111/ecog.01509> (2016).
65. Rosser, N., Phillimore, A. B., Huertas, B., Willmott, K. R. & Mallet, J. Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biological Journal of the Linnean Society.* **105**, 479–497 (2012).
66. Camero, E. R. & Lobo, J. M. The distribution of the species of *Eurysternus* Dalman, 1824 (Coleoptera: Scarabaeidae) in America: potential distributions and the locations of areas to be surveyed. *Tropical Conservation Science.* **5**, 225–244 (2012).
67. Soberón, J. & Peterson, A. T. Biodiversity informatics: managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London Series B Biological Sciences.* **359**,

- 689–698 (2004).
68. Zizka, A. *et al.* Standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution*. **10**, 744–751 <https://doi.org/10.1111/2041-210X.13152> (2019). CoordinateCleaner
69. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2019).
70. Thuiller, W., Georges, D. & Engler, R. *biomod2: Ensemble platform for species distribution modeling. Version 3.3-7*, (2016).
71. Guisan, A., Thuiller, W. & Zimmermann, N. E. *Habitat suitability and distribution models. With applications in R*(Cambridge University Press, 2017).
72. Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*. **29**, 1189–1232 (2001).
73. Breiman, L. Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
74. Phillips, S. J., Anderson, R. P. & Schapire, R. E. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* **190**, 231–259 <https://doi.org/10.1016/j.ecolmodel.2005.03.026> (2006).
75. Araújo, M. B. & New, M. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**, 42–47 <https://doi.org/10.1016/j.tree.2006.09.010> (2007).
76. Fick, S. E., Hijmans, R. J. & Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315, doi:10.1002/joc.5086 (2017).
77. Simard, M., Pinto, N., Fisher, J. B. & Baccini, A. Mapping forest canopy height globally with spaceborne lidar. *J. Geophys. Res.* **116**, G04021 <https://doi.org/10.1029/2011JG001708> (2011).
78. Lin, D., Foster, D. P. & Ungar, L. H. VIF Regression: A Fast Regression Algorithm for Large Data. *Journal of the American Statistical Association*. **106**, 232–247 <https://doi.org/10.1198/jasa.2011.tm10113> (2011).
79. Phillips, S. J. *et al.* Sample selection bias and presence-only Species Distribution Models: Implications for background and pseudo-absence data. *Ecol. Appl.* **19**, 181–197 (2009).
80. Collevatti, R. G. *et al.* A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Mol. Ecol. Notes*. **21**, 5845–5863 <https://doi.org/10.1111/mec.12071> (2012).
81. Rodrigues, A. S. L. *et al.* Effectiveness of the global protected area network in the representing species diversity. *Nature*. **428**, 640–643 <https://doi.org/10.1038/nature02422> (2004).
82. IUCN. *The IUCN Red List of Threatened Species. Version 2019.2*, (2019).
83. Ardron, J. A., Possingham, H. P. & Klein, C. J. *Marxan Good Practices Handbook*. 155(Pacific Marine Analysis and Research Association, 2008).
84. Prioritizr Systematic Conservation Prioritization in R. Version 4.1.1. Available at <https://github.com/prioritizr/prioritizr> (2019).
85. gurobi: Gurobi Optimizer 8.0 interface. R package version 80 – 1 v. 8.1 (2018).

# Figures

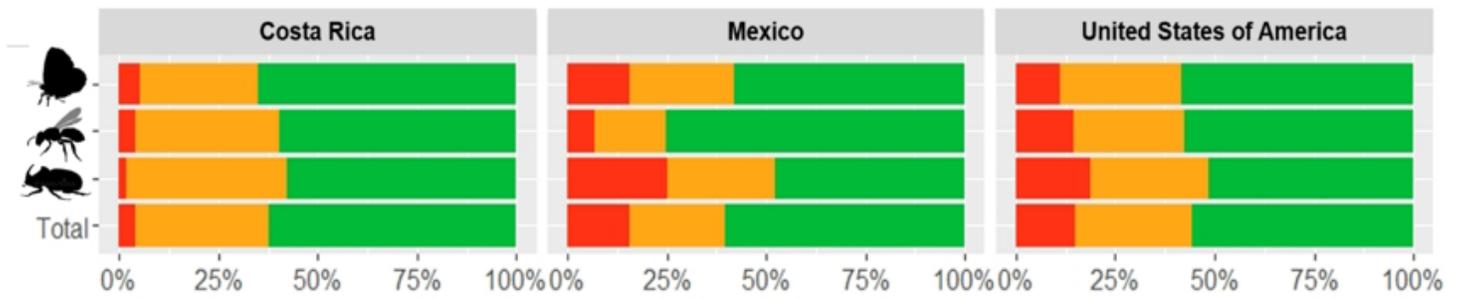
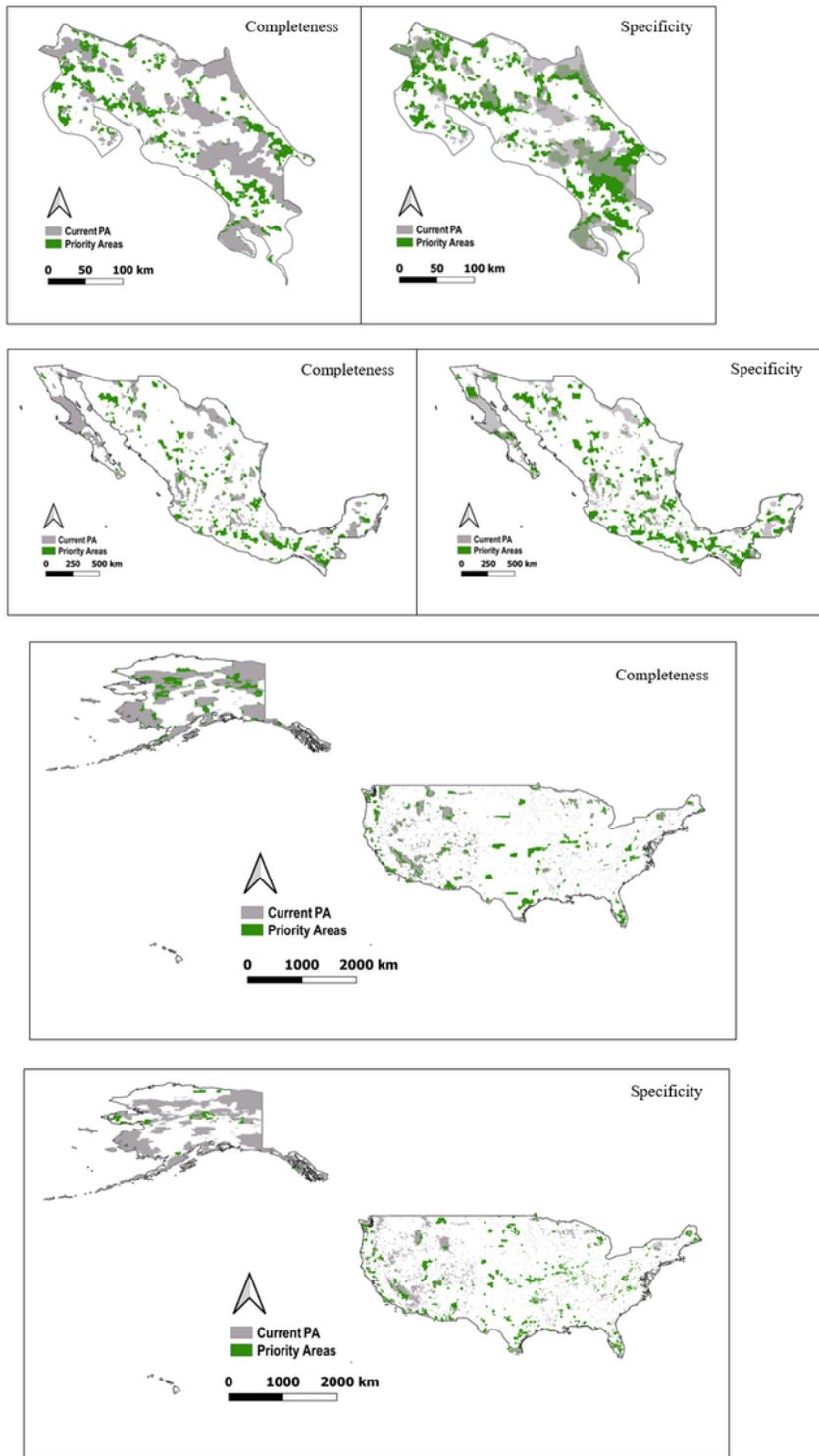


Figure 1

Conservation targets achieved by order and country, and countries' total. Green – extrinsic representativeness (100% conservation target achieved); orange – underrepresented species (>50% - <100% conservation target achieved); red – neglected species ( $\leq 50\%$  conservation target achieved).



**Figure 2**

Priority areas for unseen diversity (green) and existing PA network (gray). In the completeness maps, the newly proposed PA complement the existing PA to achieve conservation targets. In the specificity maps, the existing PA overlay the newly generated PA network, which optimizes the conservation of the elements of unseen biodiversity. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square

concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Poolofspeciesformodelling.pdf](#)