Aligning marine species range data to better serve science and conservation

Casey C. O'Hara1, Jamie C. Afflerbach1, Courtney Scarborough1, Benjamin S. Halpern1,2,3

Short title: "Aligning marine species range data"

1. National Center for Ecological Analysis and Synthesis, University of California, 735 State Street Suite 300, Santa Barbara CA 93101
2. Bren School of Environmental Science and Management, University of California, Santa Barbara CA 93106
3. Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Rd, Ascot, West Berkshire SL5 7PY, United Kingdom

*Corresponding author:* Casey O'Hara, National Center for Ecological Analysis and Synthesis, University of California, 735 State Street, Suite 300, Santa Barbara CA 93101; (805) 892-2500; [ohara@nceas.ucsb.edu](mailto:ohara@nceas.ucsb.edu)

## Abstract

Species distribution data provide the foundation for a wide range of ecological research studies and conservation management decisions. Two major efforts to provide marine species distributions at a global scale are the International Union for Conservation of Nature (IUCN), which provides expert-generated range maps that outline the complete extent of a species' distribution; and AquaMaps, which provides model-generated species distribution maps that predict areas occupied by the species. Neither dataset claims to represent the "truth" of species' spatial distributions, but rather each communicates a distinct understanding of the underlying distribution, with a distinct purpose in mind.

Together these databases represent 24,637 species (92.9% within AquaMaps, 16.3% within IUCN), with only 2,279 shared species. Ideally, the decision to use one dataset one over the other should carefully consider the purpose for which it was designed, but this small overlap between the two suggests that data availability is likely an important factor as well.

Differences in intent and methodology can result in very different predictions of species distributions, which bear important implications for scientists and decision makers who rely upon these datasets to inform conservation policy and management actions. Comparing distributions for the small subset of species with maps in both datasets, we highlight several differences resulting from differences in methodology and intent. In particular, we find that IUCN maps greatly overpredict coral presence into unsuitably deep waters, and we demonstrate that AquaMaps methodology can produce odd discontinuities at the extremes of a species range. We illustrate the scientific and management implications of these tradeoffs by repeating a global analysis of gaps in coverage of marine protected areas, and find significantly different results depending on how the two datasets are used.

It remains essential to recognizes the tradeoffs between these two valuable marine datasets, and understand the implications of these tradeoffs for conservation research and decision-making.

# Introduction

Knowing where species exist and thrive is fundamental to the sciences of ecology, biogeography, and conservation. This body of science has resulted in various compiled databases of species distribution and range maps which provide foundational information for understanding species diversity and extinction risk, predicting species responses to human impacts and climate change, and managing and protecting species effectively. No species range dataset can claim to represent the "truth" of any species' spatial distribution, but rather each offers its own distinct understanding of that latent distribution. These varying predictions of species presence and absence, driven by intent and methodology, can result in very different predictions of species presence; however, conservation managers and policy makers must base their actions on conclusions drawn from these imperfect datasets. It is critical to understand the differences in spatial range datasets and the implications of these differences for conservation research and decision-making.

The two most comprehensive, widely-used global-scale repositories that predict marine species ranges throughout the world's oceans are AquaMaps, which rely primarily on model predictions to communicate the distribution of a species based on habitat suitability [1], and range data from the International Union for Conservation of Nature (IUCN), which rely primarily on expert opinion to communicate range size as a criterion for extinction risk assessments [2].

While the two datasets ostensibly describe the same information, i.e. where can a particular species be found, they communicate fundamentally distinct concepts. The IUCN range map for a given species describes the "extent of occurrence" for that species, a contiguous region that encloses all known occurrences and connecting regions, with the explicit caveat that this "does not mean that [the species] is distributed equally within that polygon or occurs everywhere within that polygon" [2]. The AquaMaps map for a species, on the other hand, excludes unsuitable habitats and thus seems to better be defined as "area of occupancy": the subset of the extent of occurrence in which the species is actually likely to be found [2].

The fundamental differences between the datasets suggest that the choice of one over the other should be carefully matched to the purpose for which it is to be used [3], and yet these datasets have been used in hundreds of studies and applications for a wide range of purposes, including assessing marine species status [4–6], evaluating global biodiversity patterns [7–10], predicting species range shifts [11], and setting conservation priorities [12]. In most of these cases, the implications of choosing one versus the other of these datasets is not evaluated or discussed, yet strong conservation and management conclusions are drawn from the results of the studies.

Here we focus not on how the two datasets should ideally be used; instead, recognizing that each provides distinct value for conservation research, we explore hidden assumptions and sources of error within each dataset, and test the implications of choosing one over the other. You do science with the data you have, not with the data you wish you had; often, the ideal type of spatial data is not available, and non-ideal data is better than no data at all. In such a case, it is crucial to acknowledge, and where possible address, the resulting tradeoffs.

Data availability is a crucial factor in deciding which dataset to use. We compare how each data source represents the global spatial and taxonomic distribution of the 24,637 marine species mapped by one or both datasets. Most notably, AquaMaps includes range maps for many more species (currently 22,889 species; 92.9% of total), such that most global analyses related to marine biodiversity to date have used AquaMaps. IUCN range map data exist for 4,027 unique marine species, with only 2,279 species (9.3% of total) mapped in both datasets. This small overlap means each dataset provides unique value in taxonomic and geographic coverage, but also suggests that data availability likely plays a larger role in selecting a dataset than does the intended purpose of the data.

Although relatively small in number, these overlap species present a unique opportunity to evaluate the two datasets overall. For the species included in both datasets, we examine how well the maps align in both spatial distribution and overall area. Based upon the definitions of extent of occurrence and area of occupancy, we expect that for a given species, the AquaMaps predicted distribution will fall within, and describe a smaller range than, the IUCN predicted distribution. Where these expectations seem to fail, we explore methodological issues that can introduce errors. In particular, we show how explicitly using depth as a constraint on coral ranges in IUCN range data dramatically mitigates a source of commission error, and highlight a side effect of AquaMaps methodology that results in discontinuous species range boundaries.

As a case study, we then reexamine a global analysis of gaps in protection afforded by marine protected areas (MPAs) [12] as a case study to explore the implications of selecting one data set over the other. Repeating the study using IUCN range data rather than AquaMaps data quintuples the apparent proportion of "gap" species, i.e. those species whose ranges remain completely unprotected, from 1.2% to 6.4%. This shift highlights the implications of different data use decisions on our understanding of marine biodiversity status and protection.

Due to the variation in georegional and taxonomic coverage, and fundamental distinctions in method and intent of these two spatial range datasets, there is no way to simply recommend one dataset over the other. Instead, by making explicit the distinctions between the two, and offering methods to identify and reconcile potential errors in each, we hope to reinforce the effective use of these datasets to better inform conservation management and policy decisions.

# Methods and Analysis

## About the datasets

The IUCN publishes species range maps developed by species experts. These experts outline spatial boundaries that define the "limits of distribution" of a given species, based on observation records and informed by expert understanding of species' range and habitat preferences (Fig S1A). IUCN guidelines generally recommend that boundaries be drawn as a "minimum convex polygon": "the smallest polygon in which no internal angle exceeds 180 degrees and which contains all the sites of occurrence" [2]. For the purposes of this analysis, we elected not to use IUCN data for bird species, which are available separately through Bird Life International [13].

In contrast, AquaMaps models species distribution based on envelopes of environmental preference, such as temperature, depth, and salinity; these preference envelopes are deduced from occurrence records, published species databases such as FishBase, and expert knowledge. The AquaMaps model overlays these environmental preferences atop a map of environmental attributes on a global 0.5° grid to determine suitable habitat, resulting in a "probability of occurrence" for each species (Fig S1B). To roughly constrain species ranges to appropriate georegions, the AquaMaps model uses Food and Agriculture Organization of the United Nations (FAO) Major Fishing Area [14] boundaries. Of the resulting maps, 1296 (5.7%) have been further refined through an expert review process [15,16].

**Comparison of taxonomic and regional distribution**: To examine the overall taxonomic distribution across the spatial datasets, we grouped species by taxonomic class and data source, and determined the proportion of each class represented in each dataset. To compare the spatial representation of the two datasets directly, we rasterized the IUCN species polygons to the same 0.5° grid as the AquaMaps species maps; we determined species presence within a grid cell as any non-zero overlap of a species polygon with the cell (Fig S2) For the AquaMaps dataset, we determined per-cell species count by including all species with non-zero probability of occurrence, to best approximate the "extent of occurrence" generally indicated by IUCN maps.

**Comparison of paired maps**: Species range maps communicate a prediction of both the distribution of a species and the size of a species range. To enable direct comparison of these measures between the two datasets, we identified "paired map" species, i.e. the subset of marine species that have range maps in both IUCN and AquaMaps current native distribution, using genus and species binomials as a matching key. We used the taxize package [17] in R [18] to standardize species names and synonyms; for species with separate subpopulation maps in IUCN, we combined all subpopulations to create a single global population. For each of these paired map species, we determined species presence within each spatial cell for each dataset using the same criteria outlined above.

Overlaying paired distribution maps for each species, we defined and calculated *distribution alignment* and *area ratio* :

For each paired map species, and indicate the smaller and larger range representation (regardless of which dataset). represents the amount of overlapping area between the two datasets. Distribution alignment uses overlapping predictions of presence as a measure of concurrence between the two datasets. Area ratio compares range size, used by IUCN as a criterion to help define extinction risk; it also provides a proxy for frequency of commission and/or omission errors.

**Examining issues in paired map alignment**: Corals provide a useful case study to examine differences in species range alignment. Corals are an intensely studied taxonomic group whose importance in supporting biodiversity is undisputed; their dependence on photosynthesis provides habitat-limiting life history traits, and they don't tend to move. Extracting data from the IUCN API [2] we identified depth limitations of each coral species mapped in the IUCN dataset to verify that none are found below the photosynthetic depth limit of 200 m. Using the same methodology as shown in Fig 2, we created a 200 m bathymetry raster from a 200 m bathymetry spatial dataset (public domain; available from Natural Earth, www.naturalearthdata.com) and masked our IUCN coral rasters to identify mapped coral presence below 200 m. The resulting maps were again compared to the AquaMaps rasters to examine distribution alignment and area ratio.

AquaMaps' reliance upon FAO zones to constrain species ranges seems likely to result in abrupt species range boundaries where ecologically suitable habitat intersects with human-defined borders. To explore the impact of these artificial constraints on species range predictions, we chose to focus on boundaries defined by longitude, as they seem less likely to coincide with ecological differences than boundaries defined by latitude. As an example, we identified all AquaMaps species whose eastern or western range limit between between 25° S and 20° N latitudes coincided with the large vertical boundary between FAO regions 71 and 77 at 175° W longitude.

The oceans abound with rare and poorly-understood species which offer sparse information on behavior and habitat preferences to inform models and expert opinion. To identify such data-poor species within the set of paired map species, we used R clients for OBIS [19] and GBIF [20] to identify known occurrences of each species; occurrences were averaged between the two occurrence databases as a proxy for data richness.

## Methods for MPA Gap Analysis case study

To assess the effectiveness of MPAs in protecting biodiversity, Klein et al. [12] compared the coverage of the global MPA network presented by the World Database on Protected Areas (WDPA) [21] to the species ranges described in the 2014 AquaMaps dataset [1]. For the primary analysis, the researchers defined species presence as 50% or greater probability of occurrence.

To reconstruct the primary analysis, we selected the subset of protected areas from the 2014 WDPA dataset classified as IUCN protected area management categories I-IV and spatially overlapping a marine area. The WDPA polygons and marine polygons were rasterized to 0.01° and then aggregated to AquaMaps native 0.5° cells, to calculate proportion of marine protected area within each cell. After verifying our results using the 2014 AquaMaps dataset, we updated the analysis to use the 2015 AquaMaps dataset, at a presence threshold of 50% (to compare to Klein et al. directly) and 0% (to better compare with IUCN spatial data). To analyze MPA coverage against IUCN spatial data, we extracted IUCN polygon weights per 0.5° cell for each species and compared against the protected area raster. Finally, we combined AquaMaps data (at 0% threshold) and IUCN data, using AquaMaps for the 2,279 overlapping species and again compared against the protected area raster.

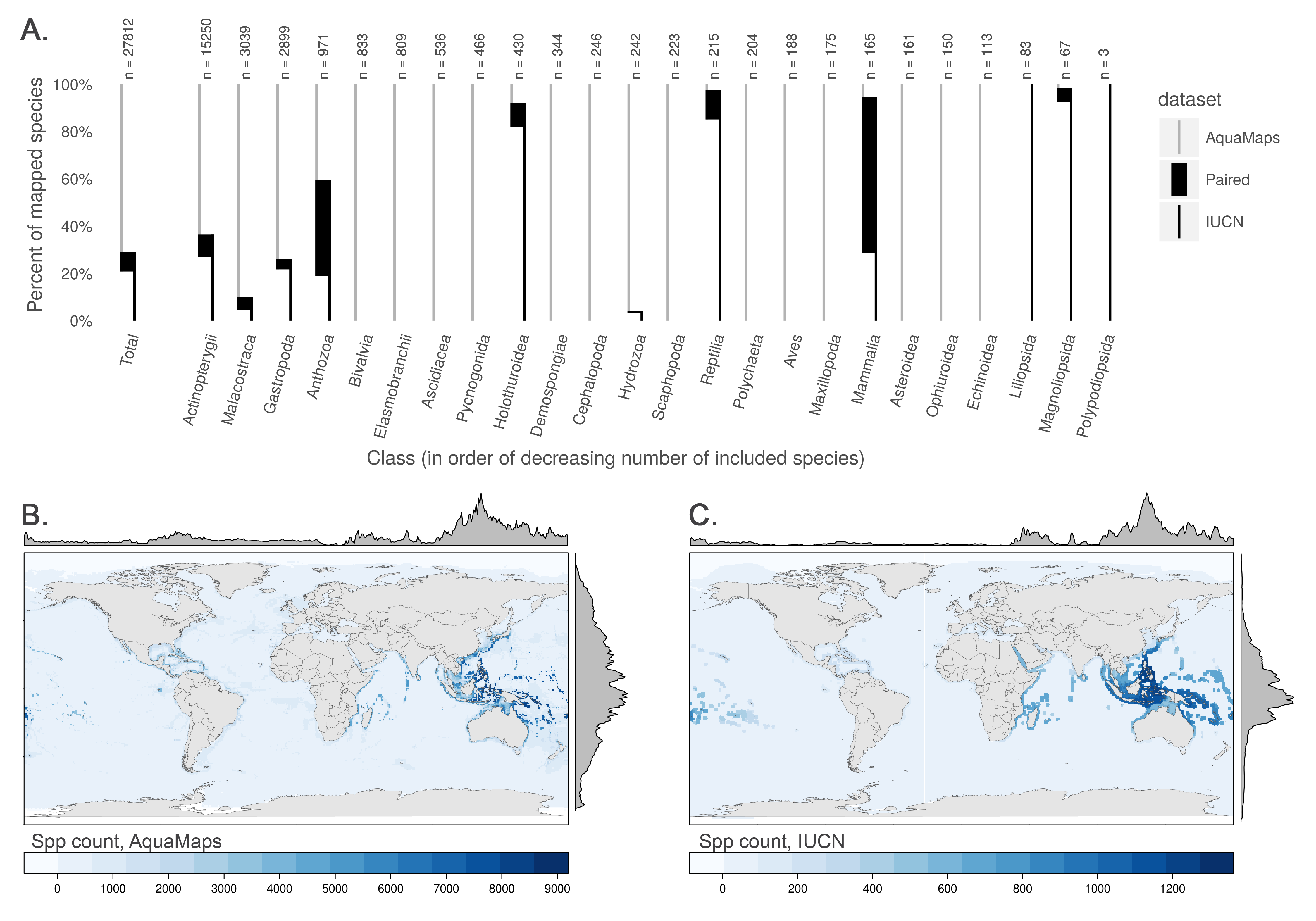
All processing was completed using R statistical software [18], and all code and intermediate data are available on GitHub at <https://github.com/OHI-Science/IUCN-AquaMaps>.

# Results and Discussion

In comparing the IUCN and AquaMaps datasets, it is again important to emphasize that the two differ in both methodology and intent. As area of occupancy is generally defined as a subset of a species' extent of occurrence, AquaMaps ranges (effectively area of occupancy) should generally fall within the bounds of IUCN ranges (effectively extent of occurrence), capturing a smaller total area. However, AquaMaps range maps are created independently from IUCN data and therefore exceptions are certain to arise. Here we are looking for systematic deviations from our expectations that might highlight implications of data use decisions.

## Taxonomic and geographic coverage

The two datasets have notably different taxonomic (Fig 1A) and regional (Figs 1B, 1C) coverage. AquaMaps encompasses a broader range of taxa than IUCN, as IUCN spatial data files are only available for select taxonomic groups that have been comprehensively assessed. Of the 24,637 species mapped within these datasets, only 2,279 (9.3%) are mapped within both, with many taxa completely unrepresented in one dataset or the other. While species numbers in both datasets peak in tropical latitudes near the equator, species counts for IUCN maps drop quickly beyond 30°N and 30°S, while AquaMaps includes distribution of species well into temperate latitudes. The longitude frequency plots show a slight shift in the IUCN dataset away from the Atlantic and eastern Pacific compared to AquaMaps. Together, the limitations of spatial and taxonomic coverage are likely to drive a researcher's choice of dataset far more strongly than the quality, format, or intended purpose of the dataset.

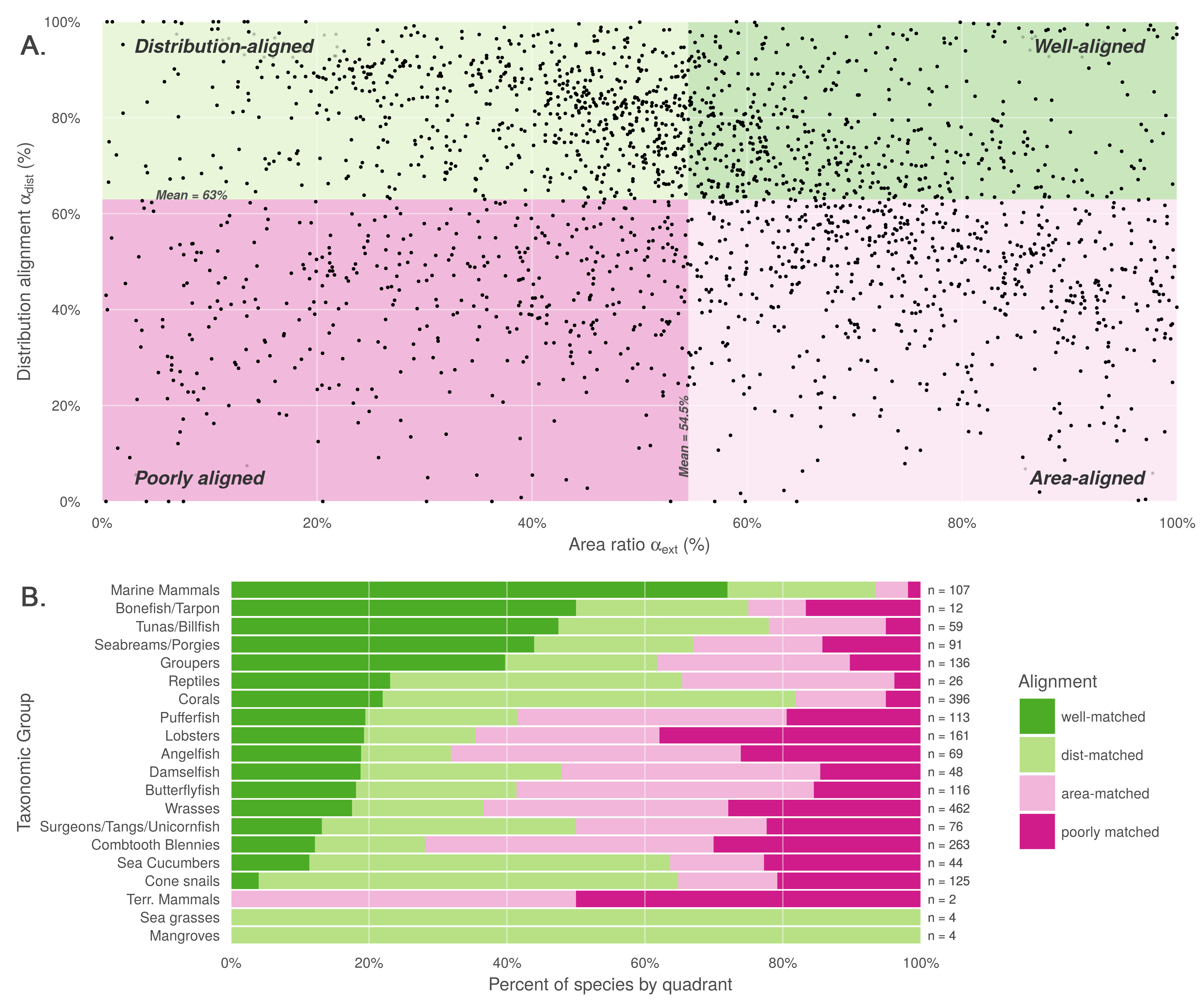


**Fig 1.** (A) Number and proportion of species by taxa included in each dataset. Overlapping species are dominated by bony fishes (983 species, primarily tropical taxa) and corals (396 species). (B, C) Global marine species count per 0.5° cell according to (B) AquaMaps and (C) IUCN. The margin frequency plots show relative species count per cell at each latitude and longitude.

## Distribution and range size alignment

Comparing distribution alignment and area ratio for the 2,279 paired map species (Fig 2A), a weak negative linear pattern appears to emerge, suggesting that increasing similarity in range area correlates very slightly with decreasing distribution alignment (R^2 = .016). The pattern itself is not particularly important, and emerges simply due to the nature of the analysis and the datasets: The AquaMaps model tends to extrapolate species ranges into suitable areas beyond known occurrences more frequently than IUCN maps, such that each additional unit of range predicted by AquaMaps will fall in different locations than an additional unit of range predicted using IUCN methodology. For species with dissimilar range areas, predicted distribution for the smaller range can more easily fall within the generous bounds of the larger range. For species with increasingly similar range areas, differences in methodology become more difficult to "hide," and the distribution alignment generally becomes poorer.

The mean distribution alignment for species included in both datasets was 63%; the mean area alignment was 54.5%. By dividing the paired map species into quadrants based on these means, we highlight categories of relationships that help further explain this general pattern. Representative maps from each category are provided in the supporting materials (Fig S3).



**Fig 2.** (A) Distribution alignment (overlap of smaller range within larger) versus area ratio (the ratio of smaller range area to the larger range area) for 2,279 species included in both IUCN and AquaMaps datasets. The upper right quadrant (quadrant 1) comprises species whose maps largely agree (better than mean value) in both spatial distribution and the extent of described ranges (n = 466; 20.1 %). The upper left quadrant (quadrant 2) comprises species whose maps agree well in distribution, but disagree in area (n = 687; 29.7 %). The lower right quadrant (quadrant 3) includes species for which the paired maps generally agree in range area, but disagree on where those ranges occur (n = 691; 29.9 %). The lower left quadrant (quadrant 4) indicates species for which the map pairs agree poorly in both area and distribution (n = 470; 20.3 %). (B) Alignment quadrant breakdown of species by taxonomic group.

The upper right quadrant includes the species (n = 527) whose described ranges are above average in alignment of both spatial distribution and area. These species tend to be well-studied and include wide-ranging pelagic organisms such as marine mammals, tunas, and billfishes (Fig 2B). This result is not surprising, as species with very large ranges are likely to be more aligned regardless of methodology simply because their extent of occurrence and area of occupancy both span nearly the entire map.

The area-mismatched ranges contained in the upper left quadrant (n = 709) include many species whose spatial distribution is similar, but where one range is notably larger than the other. For 88% of the species in this quadrant, the IUCN range is larger (on average, 2.57 times as large) than the AquaMaps range, as expected when comparing an extent of occurrence to an area of occupancy map. This suggests a high rate of commission and/or omission errors by one or both datasets; further analysis is required to disentangle the source and type of error contributing to poor area alignment. The results of the coral analysis (described below) provide some insight.

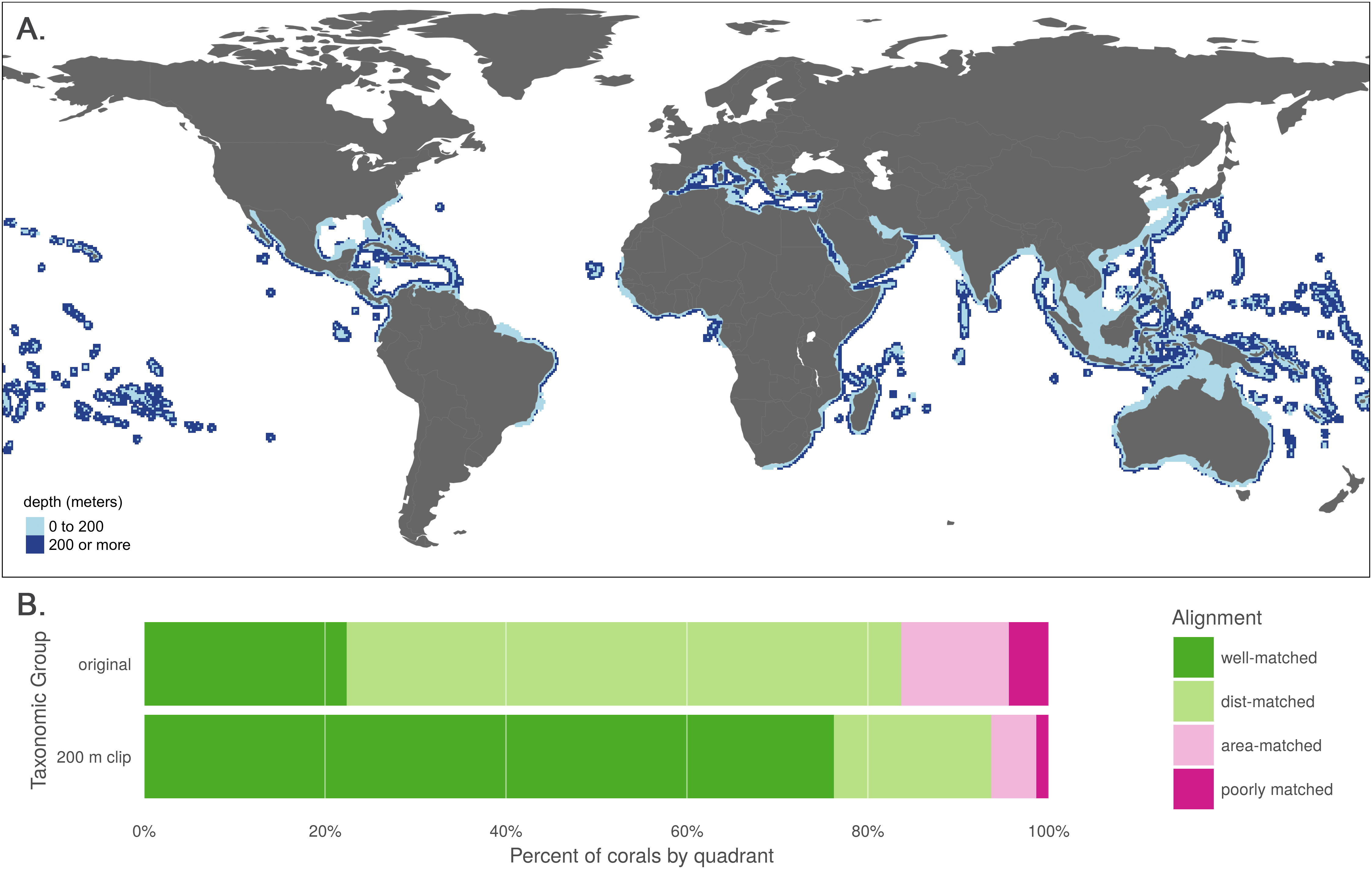
Species found in the lower right quadrant (n = 635) seem to represent cases of "two wrongs make a right." For these species, IUCN and AquaMaps both predict ranges extending far beyond the overlapping region, but the methodological differences result in very different extrapolations. Consequently, area ratios are close to 100%, though the poor distribution alignment indicates that one or both datasets are introducing significant errors. In this quadrant, the IUCN range is the larger for only 56.5% of species, defying our expectations based on definitions of area of occupancy and extent of occurrence. As above, further analysis is required to disentangle the causes of error.

The lower left quadrant includes species (n = 443) where alignment is poor in both dimensions. In this quadrant, the IUCN range is larger for only 24.6% of species; this again defies our expectations for extent of occurrence and area of occupancy range predictions. Data-poor species are more common in this quadrant; indeed, the median number of species occurrence records (averaging occurrences from the Ocean Biogeographic Information System (OBIS) [19] and the Global Biodiversity Information Facility (GBIF) [20]) for this quadrant is 24 records, compared to a median of 97 records for species across the other three quadrants. The AquaMaps dataset offers its own quality metric based on the number of unique cells containing valid occurrences; for this quadrant, the median "occurcells" is 11 compared to a median of 40 across the other three quadrants. Care should be taken when using distribution and range maps based upon fewer observations, as they clearly bear greater uncertainty; AquaMaps explicitly warns against using any of its maps generated with an "occurcells" count fewer than 10 [1].

## Coral depth exploration

Noting from Fig 2B that corals dominate the upper-left "distribution-aligned" quadrant of Fig 2A (n = 237; 33.4% of all species in this quadrant), we chose to examine the effect of explicitly restricting IUCN ranges to depths based on species' life histories. Corals offer an excellent case study, due to their foundational role in biodiverse habitats, as well as their lack of mobility and reliance on photosynthesis. Ocean depth preference is explicitly included as a parameter in AquaMaps models. While depth is recommended by the IUCN as a criterion for providers of range maps ("The limits of distribution can be determined by using known occurrences of the species, along with the knowledge of habitat preferences, remaining suitable habitat, elevation limited, and other expert knowledge of the species and its range." [2]), it is not presented as a requirement, so we cannot take its inclusion for granted.

Fig 3A shows aggregated ranges of the 463 coral species mapped in the IUCN dataset, with their ranges broken into proportional area deeper and shallower than 200 m. According to IUCN descriptions, none of these species is indicated to occur deeper than 200 m, and 94% are confined to waters shallower than 50 m; seven of the mapped species had no reported depth information. Clipping coral ranges to shallower than 200 m eliminated an average of 47.6% of the total predicted area while still allowing for a generous estimate of suitable habitat.



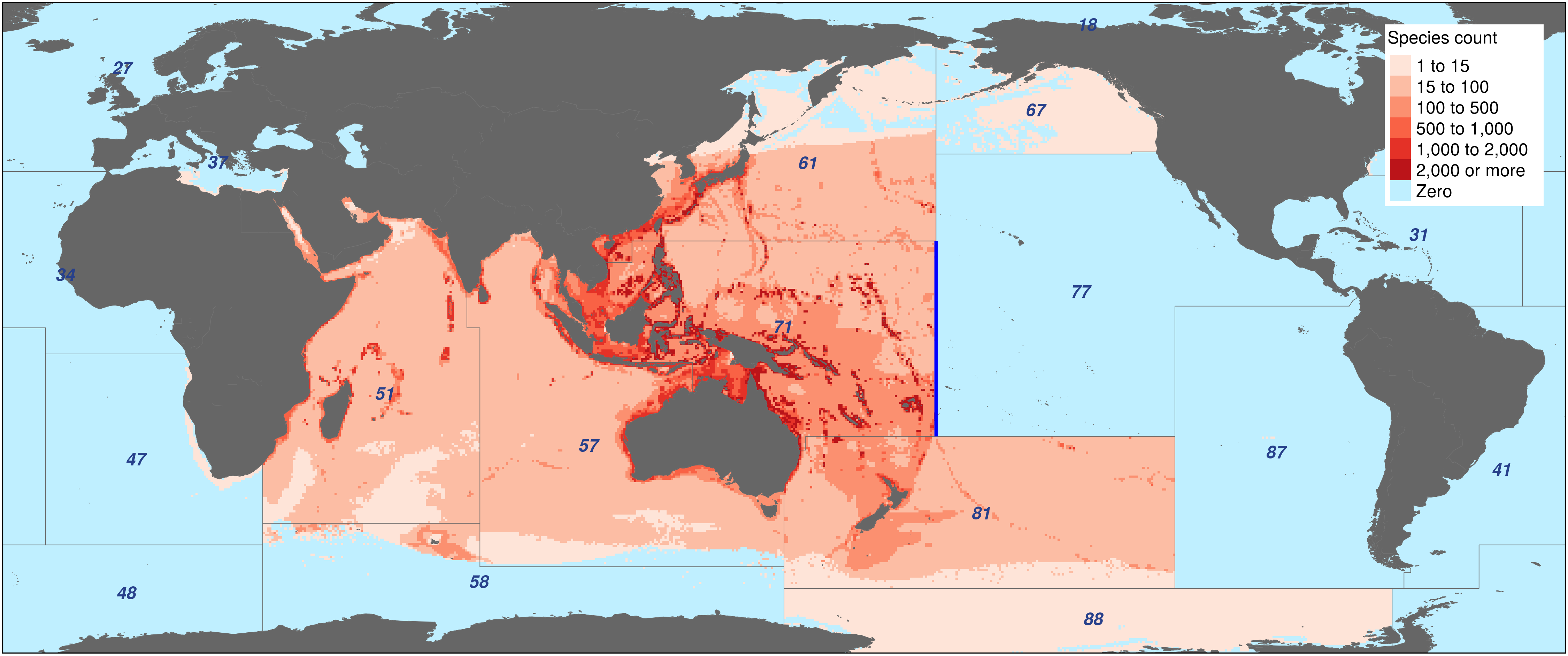
**Fig 3.** (A) Aggregate map combining ranges of the 463 coral species mapped in the IUCN dataset, showing raw ranges and ranges clipped to 200 m depth. (B) Alignment quadrant breakdown of paired map coral species using original data from IUCN and AquaMaps (as in Fig 2B) and the same species with IUCN ranges clipped to 200 m depth.

In constraining coral ranges to appropriately shallow waters, we see a strong increase in the apparent alignment of species maps between IUCN and AquaMaps. Using the original quadrant definitions from Fig 2B, we see in Fig 3B a massive shift in 354 paired-map coral species due to constraining coral depths to 200 m. Membership in the "well-aligned" quadrant jumped from 22.4% to 76.2%, with a corresponding decrease in all other quadrants. By excluding the unsuitable areas from IUCN's extent of occurrence, we eliminate preventable commission errors and more closely approximate the area of occupancy described by AquaMaps. See Fig S4 to examine the shifts of individual species among the quadrants.

We cannot know for certain the true distribution of each of these corals; certainly some commission errors result from satisfying the connectivity requirements of an extent of occurrence, and others may be due to experts taking a precautionary approach. But a sensible shift in method drastically decreases the likelihood of introducing commission errors, with little chance of introducing omission errors, greatly improving our confidence in the remaining reported distributions for most purposes. Note that this change applies just as readily to the IUCN coral maps that are not included in the paired map analysis, and likely to other reef-associated flora and fauna. While species depth preferences are an easy and consistent means of constraining range predictions, other conditions such as salinity and temperature could be cautiously used to refine the results of expert opinions.

## Georegional constraint exploration

From the entirety of the AquaMaps dataset, we identified 3,577 Indo-Pacific species whose equatorial distributions (between 25° S and 20° N) encounter a range limit at 175° W. This longitudinal boundary created a western limit for 369 species, and an eastern limit for 3,208 species. A clear discontinuity in species distributions of the 3,208 west-limited species (Fig 4) matches perfectly with FAO region 77 [14]; other discontinuities are apparent at other FAO boundaries, despite these boundaries not being actively studied in the analysis. Similar discontinuities were clear when mapping the west-limited species.



**Fig 4.** Composite map showing AquaMaps predicted ranges for 3,208 species whose equatorial distribution encounters an eastern discontinuity exactly at 175° W, the boundary between FAO Major Fishing Areas 71 and 77 (shown in blue). Other FAO area boundaries create additional clear discontinuities.

FAO Major Fishing Area boundaries provide a readily available method to roughly contrain AquaMaps predictions to appropriate ocean basins, thus eliminating a large source of potential commission errors and enabling rapid modeling of thousands of species ranges. However, these boundaries are defined for statistical purposes based on economic and political considerations rather than ecological considerations, and can result in odd discontinuities in species range predictions where otherwise suitable habitat is excluded. While such a discontinuous boundary would likely be obvious when inspecting the distribution of an individual species, the distinction is likely to be obscured when aggregating many species ranges as is typical for biodiversity or conservation studies. Ecoregional spatial zones such as Marine Ecoregions of the World [22] may provide an additional useful constraint on species ranges; expert review, though time-consuming, is the most certain route to boosting confidence in these predicted distributions.

## Case Study: MPA Gap Analysis

Klein et al. [12] compare the global distribution of species to the global distribution of marine protected areas to assess how well current MPAs overlap with species ranges and identify which species fall through gaps in protection. The study relied on the 2014 version of the AquaMaps database, using a probability of occurrence threshold of 50% or greater, to determine species presence, and the World Database of Protected Areas to define zones of marine protection. They found that the global MPA network leaves 90.5% of marine species with less than 5% of their overall range represented within MPAs, and 1.4% of species have no protection at all (i.e., "gap" species).

We recalculated the amount of under-protected and gap species using either IUCN or AquaMaps data (using the 2015 AquaMaps data and a 0% threshold to more closely approximate the extent of occurrence represented by IUCN data, Fig 4). We found a five-fold increase in the proportion of gap species (6.4% of species vs. 1.2%) and dramatically larger proportion of species with less than 2% of their range protected (73.2% of species vs. 47.7%). However, this comparison also indicates a larger proportion of well-protected species with greater than 10% of range protected (2.9% of species vs. 1.5%).

**Fig 5.** Percent of species range covered by MPAs based upon methods in Klein et al. (2015). Scenario 1 replicates the original results, measuring protected range of species in AquaMaps 2014 dataset, with a 50% presence threshold, against the 2014 World Database of Protected Areas, filtered for IUCN categories I-IV that overlap marine areas. Scenario 2 updates the results using AquaMaps 2015, showing very small changes despite the inclusion of an additional 5,545 species. Scenario 3, using 2015 AquaMaps data, drops the presence threshold to zero, showing an expected decrease in gap species, but also a decrease in species with 5% or greater protected range. Scenario 4 examines species MPA coverage using only the IUCN dataset.

To achieve more comprehensive global coverage of species ranges it may be desirable to use these two datasets together, as the Ocean Health Index (OHI) does to calculate scores for its Species goal [4,5]. While it may be unrealistic to "fix" one data set to match the other, we may be able to reduce the impact of the tradeoffs inherent in each. Trimming unsuitable habitat from the IUCN's extent of occurrence maps, for example by explicitly clipping them to appropriate depths for corals and reef-associated species, would result in maps that more closely approximate areas of occupancy. Conversely, including all AquaMaps cells with a non-zero "probability of occurrence" (rather than using a probability threshold to determine presence, e.g. greater than 40% for the Ocean Health Index [4,5] or 50% for the MPA gap analysis [12]) would allow for the most generous inclusion of species range, resulting in maps that more closely, though still imperfectly, approximate extents of occurrence.

# Conclusions

No spatial dataset can ever claim to know the "truth" of the whole of marine biodiversity. Although other approaches exist for predicting species distribution, AquaMaps and IUCN range maps are the largest approaches applied globally across a broad range of marine taxa. These datasets show strong agreement for many well-studied species, but for many others, substantial differences arise from differences in methodology and intent of each dataset. While the decision of which dataset to use should ideally be driven by the intended purpose for which it was created, the fact is that taxonomic coverage will likely be a more important factor. Recognizing and acknowledging the advantages and differences of the range maps presented by these datasets will increase their utility for research and conservation actions. Conclusions drawn from each of these datasets could paint dramatically different pictures of global marine biodiversity or the effectiveness of conservation management decisions.

Effective management and protection of marine species depends on a robust understanding of where species exist and where they do not; without this knowledge we risk wasting resources protecting low-value regions while missing opportunities to protect critical ones. By highlighting the distinctions between these two important and fundamentally different marine species range datasets, we improve our ability to inform strategic and effective conservation policy that supports a resilient ocean ecosystem.

# Acknowledgments

We are thankful to Melanie Frazier, Julia Stewart Lowndes, Halley Froelich, Claire Runge, and two anonymous reviewers for their insightful comments on earlier versions of this manuscript, and to the National Center for Ecological Analysis and Synthesis for computation support. We acknowledge financial support from the Gordon and Betty Moore Foundation.

# References

1. Kaschner K, Rius-Barile J, Kesner-Reyes K, Garilao C, Kullander S, Rees T, et al. AquaMaps: Predicted range maps for aquatic species [Internet]. 2015. Available: <www.aquamaps.org>

2. IUCN. The IUCN Red List of Threatened Species [Internet]. International Union for the Conservation of Nature. 2015. Available: <http://www.iucnredlist.org>

3. Gaston KJ, Fuller RA. The sizes of species’ geographic ranges. Journal of Applied Ecology. 2009;46: 1–9. doi:[10.1111/j.1365-2664.2008.01596.x](https://doi.org/10.1111/j.1365-2664.2008.01596.x)

4. Halpern BS, Longo C, Hardy D, McLeod KL, Samhouri JF, Katona SK, et al. An index to assess the health and benefits of the global ocean. Nature. 2012;488: 615–620. doi:[10.1038/nature11397](https://doi.org/10.1038/nature11397)

5. Halpern BS, Longo C, Lowndes JSS, Best BD, Frazier M, Katona SK, et al. Patterns and Emerging Trends in Global Ocean Health. Tsikliras AC, editor. PLOS ONE. 2015;10: e0117863. doi:[10.1371/journal.pone.0117863](https://doi.org/10.1371/journal.pone.0117863)

6. Selig ER, Longo C, Halpern BS, Best BD, Hardy D, Elfes CT, et al. Assessing Global Marine Biodiversity Status within a Coupled Socio-Ecological Perspective. Guichard F, editor. PLoS ONE. 2013;8: e60284. doi:[10.1371/journal.pone.0060284](https://doi.org/10.1371/journal.pone.0060284)

7. Coll M, Piroddi C, Steenbeek J, Kaschner K, Ben Rais Lasram F, Aguzzi J, et al. The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. Bograd SJ, editor. PLoS ONE. 2010;5: e11842. doi:[10.1371/journal.pone.0011842](https://doi.org/10.1371/journal.pone.0011842)

8. Martin C, Fletcher R, Jones M, Kaschner K, Sullivan E, Tittensor DP, et al. Manual of marine and coastal datasets of biodiversity importance. United Nations Environment Programme; 2014 May.

9. Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, et al. The biodiversity of species and their rates of extinction, distribution, and protection. Science. 2014;344: 1246752–1–1246752–10. doi:[10.1126/science.1246752](https://doi.org/10.1126/science.1246752)

10. Kaschner K, Tittensor DP, Ready J, Gerrodette T, Worm B. Current and Future Patterns of Global Marine Mammal Biodiversity. Bograd SJ, editor. PLoS ONE. 2011;6: e19653. doi:[10.1371/journal.pone.0019653](https://doi.org/10.1371/journal.pone.0019653)

11. García Molinos J, Halpern BS, Schoeman DS, Brown CJ, Kiessling W, Moore PJ, et al. Climate velocity and the future global redistribution of marine biodiversity. Nature Climate Change. 2015;6: 83–88. doi:[10.1038/nclimate2769](https://doi.org/10.1038/nclimate2769)

12. Klein CJ, Brown CJ, Halpern BS, Segan DB, McGowan J, Beger M, et al. Shortfalls in the global protected area network at representing marine biodiversity. Scientific Reports. 2015;5: 17539. doi:[10.1038/srep17539](https://doi.org/10.1038/srep17539)

13. BirdLife International and NatureServe. Bird species distribution maps of the world. BirdLife International, Cambridge, UK; NatureServe, Arlington, USA; 2015.

14. FAO Fisheries and Aquaculture Department. CWP Handbook of Fishery Statistical Standards. Section H: FISHING AREAS FOR STATISTICAL PURPOSES. Food; Agriculture Organization of the United Nations; 2015.

15. Kaschner K, Watson R, Trites AW, Pauly D, others. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Marine Ecology Progress Series. 2006;316: 2–3. Available: <http://www.vliz.be/imisdocs/publications/100462.pdf>

16. Ready J, Kaschner K, South AB, Eastwood PD, Rees T, Rius J, et al. Predicting the distributions of marine organisms at the global scale. Ecological Modelling. 2010;221: 467–478. doi:[10.1016/j.ecolmodel.2009.10.025](https://doi.org/10.1016/j.ecolmodel.2009.10.025)

17. Scott Chamberlain, Eduard Szocs. Taxize - taxonomic search and retrieval in R. F1000Research. 2013; Available: <http://f1000research.com/articles/2-191/v2>

18. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available: <https://www.R-project.org>

19. OBIS. Data from the Ocean Biogeographic Information System [Internet]. Intergovernmental Oceanographic Commission of UNESCO. 2016. Available: <http://www.iobis.org>

20. GBIF. Global Biodiversity Information Facility (GBIF) Memorandum of Understanding [Internet]. Global Biodiversity Information Facility. 2010. Available: <http://www.gbif.org/resource/80661>

21. IUCN, UNEP-WCMC. The World Database on Protected Areas (WDPA) [Internet]. Cambridge, UK: UNEP-WCMC. 2014. Available: <www.protectedplanet.net>

22. Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña ZA, Finlayson MAX, et al. Marine ecoregions of the world: A bioregionalization of coastal and shelf areas. BioScience. 2007;57: 573–583. Available: <https://bioscience.oxfordjournals.org/content/57/7/573.full>