

## SPECIAL FEATURE

Approaches for general rules of biodiversity patterns in space and time

# Global distribution of coral diversity: Biodiversity knowledge gradients related to spatial resolution

Buntarou Kusumoto<sup>1,2,3</sup> | Mark J. Costello<sup>4</sup> | Yasuhiro Kubota<sup>2,5</sup> |  
Takayuki Shiono<sup>2</sup> | Chi-Lin Wei<sup>6</sup> | Moriaki Yasuhara<sup>7</sup> | Anne Chao<sup>8</sup>

<sup>1</sup>Royal Botanic Gardens, Kew,  
Richmond, UK

<sup>2</sup>Faculty of Science, University of the  
Ryukyus, Okinawa, Japan

<sup>3</sup>Okinawa Prefecture Environment  
Science Center, Okinawa, Japan

<sup>4</sup>School of Environment, The University  
of Auckland, Auckland, New Zealand

<sup>5</sup>Marine and Terrestrial Field Ecology,  
Tropical Biosphere Research Center,  
University of the Ryukyus, Okinawa,  
Japan

<sup>6</sup>Institute of Oceanography, National  
Taiwan University, Taipei, Taiwan

<sup>7</sup>School of Biological Sciences and Swire  
Institute of Marine Science, The  
University of Hong Kong, Hong Kong,  
China

<sup>8</sup>Institute of Statistics, National Tsing Hua  
University, Hsin-Chu, Taiwan

## Correspondence

Buntarou Kusumoto, Faculty of Science,  
University of the Ryukyus, Okinawa,  
Japan.  
Email: kusumoto.buntarou@gmail.com

## Funding information

JSPS KAKENHI, Grant/Award Number:  
16H02755; University of the Ryukyus;  
Program for Advancing Strategic  
International Networks to Accelerate the  
Circulation of Talented Researchers

## Abstract

Biodiversity knowledge shortfalls, especially incomplete information on species distributions, can lead to false conclusions about global biodiversity patterns. Diversity estimation theory statistically uses species occurrence records and sampling completeness (coverage) to predict diversity in terms of species richness, dominance and evenness. We estimated Scleractinia coral species diversity at different spatial resolutions, based on 109,296 occurrences and range data for 697 species, using an incidence-based Hill's numbers approach through a rarefaction and extrapolation technique. We found that spatial patterns of diversity estimates were dependent on a geographic scale. The latitudinal and longitudinal diversity gradients, particularly at finer spatial scales, differed from species range-based coral biodiversity hotspots of previous studies. The western Indian Ocean was predicted to have the most coral species, with greater diversities than in the Indo-Pacific Coral Triangle. We concluded that the identification of marine biodiversity hotspots is sensitive to species commission errors (from range maps) and biased sampling coverage. Moreover, estimates of the geographic distribution of species richness informed us of a set of priority areas (the northeastern coast of Australia, central Coral Triangle and coast of Madagascar) for future sampling of unknown coral species occurrence. Our findings of biogeographical survey priorities contribute to filling biodiversity shortfalls for tropical coral reefs through sampling completeness, and consequently for development of conservation planning.

## KEYWORDS

coral reefs, Hill's number, species incidence, sampling bias, marine biodiversity hotspots

## 1 | INTRODUCTION

Fine-scale biodiversity mapping is fundamental to macroecological studies and spatial conservation prioritization (Gaston, 2000; Jetz, McPherson, & Guralnick, 2012;

Moilanen, Wilson, & Possingham, 2009; Tittensor et al., 2010). In recent decades, this approach has led to the creation of databases referencing the occurrence of millions of species (or specimen records), which are potentially significant for biodiversity assessment in space and time

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Ecological Research* published by John Wiley & Sons Australia, Ltd on behalf of The Ecological Society of Japan

(Costello, Vanhoorne, & Appeltans, 2015; Graham, Ferrier, Huettman, Moritz, & Peterson, 2004; Ponder, Carter, Flemons, & Chapman, 2001). However, the incompleteness of occurrence data is a ubiquitous problem (Boakes et al., 2010; Maldonado et al., 2015) that is inherently associated with the spatial and temporal biases of sampling effort and species detection. Indeed, deficiency in species geographical information—the so-called Wallacean shortfall (Hortal et al., 2015; Jetz et al., 2019)—has been reported for various taxonomic groups (e.g., García-Roselló et al., 2015; Hermoso, Kennard, & Linke, 2015; Troia & McManamay, 2017). The Linnean (Brito, 2010) and Darwinian (Diniz-Filho, Loyola, Raia, Mooers, & Bini, 2013) shortfalls are also implicated in terms of lack of taxonomic and phylogenetic information. Jointly, these shortfalls hinder ecological interpretation and/or conservation planning (Girardello, Martellos, Pardo, & Bertolino, 2018; Hortal, Jiménez-Valverde, Gómez, Lobo, & Baselga, 2008; Reddy & Dávalos, 2003). Therefore, improving biodiversity shortfalls is critical (Meyer, Kreft, Guralnick, & Jetz, 2015; Sánchez-Fernández, Lobo, Abellán, & Millán, 2011).

One promising approach for filling gaps in knowledge of biodiversity geographical patterns is the theory of diversity estimation based on species accumulation curves (Yang, Ma, & Kreft, 2013) and non-parametric estimation (Chao, 1987). In this context, Hill's numbers (Hill, 1973) have been recognized as the statistically best measure of diversity (Jost, 2006). Recently, Chao et al. proposed a unified theory for interpolation (rarefaction) and extrapolation of Hill's numbers on the basis of sampling completeness rather than sample size (Chao et al., 2014; Chao & Jost, 2012). They then developed a tool (iNEXT) to estimate species diversity (Hsieh, Ma, & Chao, 2016) using empirical species occurrence data sampled in particular areas or across a large-scale gradient. Based on their approach, it is possible to predict large-scale species diversity patterns. Their estimation of the effective number of species at given sampling completeness and their estimation error are informative in assessing our current knowledge of species biogeography. These estimates can help identify priority areas for allocating future sampling effort to improve our knowledge of biodiversity patterns.

From a viewpoint of marine biodiversity conservation, tropical stony corals (Scleractinia) are a foundational taxon that build coral reef ecosystems in shallow marine waters and support about one-quarter to one-third of all marine species (Costello, 2015) and associated ecosystem services (Moberg & Folke, 1999). Indeed, loss of coral biodiversity due to human impacts and climate change emphasizes the failure to implement effective conservation strategies (e.g., Bellwood, Hughes, Folke, &

Nyström, 2004; Huang, 2012; UNEP, 2010). To prioritize areas for systematic conservation planning (Asaad, Lundquist, Erdmann, & Costello, 2018), expert range maps of coral species have been commonly stacked and then investigated to better understand large-scale regional diversity patterns (Bellwood, Hughes, Connolly, & Tanner, 2005; Sanciangco, Carpenter, Etnoyer, & Moretzsohn, 2013; Tittensor et al., 2010). A particular concern with this approach is that coral species distributions are mostly delineated empirically (Veron, 2000) or modeled by a habitat suitability model (Couce, Ridgwell, & Hendy, 2012). Specifically, coral diversity maps based on stacking species distribution ranges can include false-positives of species presence (commission error) that may hinder a precise understanding of coral diversity patterns and the underlying mechanisms. From this point of view, diversity estimation theory using species occurrence data is helpful: these field observations suffer more from omission than commission error. A comparison of range maps with species occurrence diversity estimation may reveal artifacts related to data incompleteness when identifying coral diversity hotspots at the global scale.

In this study, we compiled 109,296 occurrence records for 697 species of Scleractinia stony corals inhabiting tropical and temperate shallow water across the world. We estimated species diversity using Hill's numbers standardized using a rarefaction and extrapolation technique based on sampling completeness (Chao et al., 2014). Sampling completeness, species diversity values and their estimation errors were evaluated per grid cell at different spatial resolutions. We examined scale-dependent geographical patterns of the species diversity values, including latitudinal and longitudinal diversity gradients, and then compared them with a range-map-based diversity pattern (Veron, 2000). In addition, we investigated genus-level diversity estimations and explored their geographical patterns to determine the impact of taxonomic uncertainty—another serious problem in species occurrence records for coral species (Veron, 2013). Finally, we recommend field survey priorities to fill diversity knowledge shortfalls for coral reefs, and thus provide more accurate diversity estimation.

## 2 | METHODS

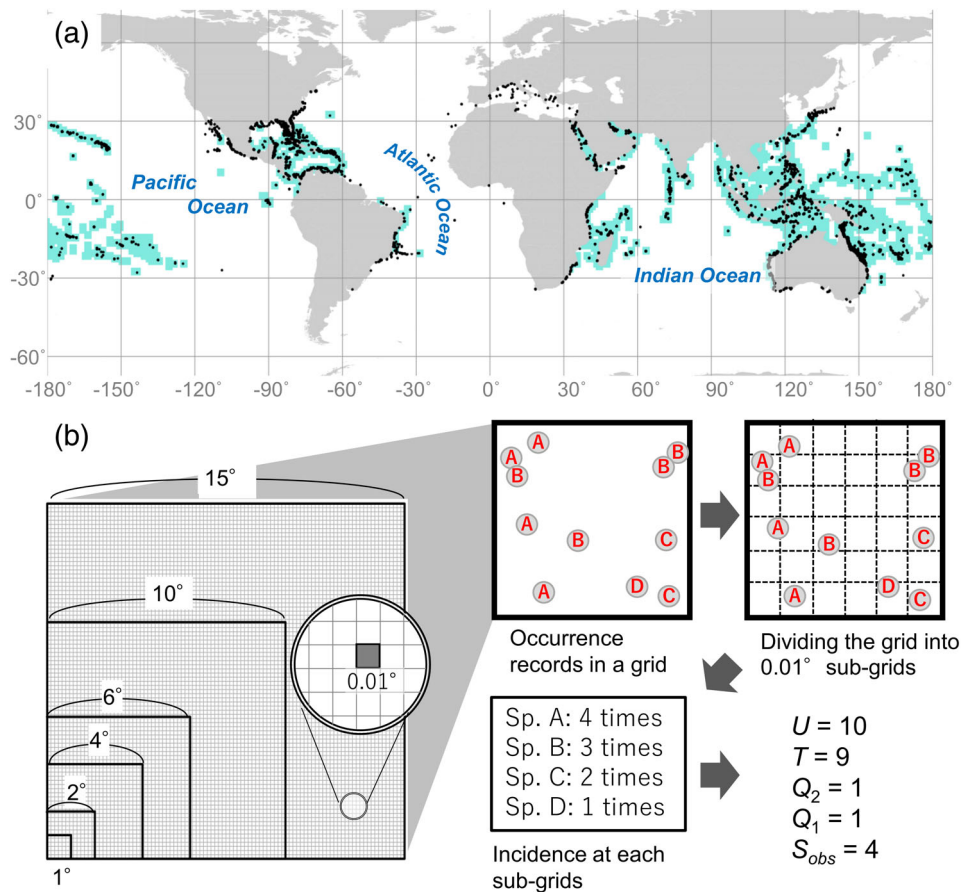
### 2.1 | Occurrence data

We downloaded global-scale occurrence (incidence) data of Scleractinia species from the Global Biodiversity Information Facility (GBIF; <https://www.gbif.org/>) and the Ocean Biogeographic Information System (OBIS;

www.iobis.org/). To compile a data set of species occurrence, we excluded occurrence records not identified to species level and/or without a geographical coordinate. We also removed duplicated information (i.e., identical species names, longitudes and latitudes, where the averaged positioning resolution was approximately  $0.2^\circ$ ) within and between the two data sources. Species names were standardized following the World Register of Marine Species (WoRMS) (Horton et al., 2019); subspecies and variety were merged into their respective species. To focus on tropical coral reefs, we filtered out particular species (e.g., deep-sea or cold-water corals) not previously recorded from coral reef areas (Figure 1; ReefBase; <http://www.reefbase.org/main.aspx>). We finally obtained 109,296 geo-referenced occurrence points for 697 species (Figure 1a).

## 2.2 | Defining species incidence

We defined the frequency of species incidence at six spatial resolutions: 1, 2, 4, 6, 10 and  $15^\circ$  grid cells (Figure 1b). We divided each grid cell into  $0.01^\circ$  of sub-gridded cells, counted the number of sub-gridded cells that contained occurrence records for individual species, and then created the dataset of species incidence at the grid cell level. To obtain reliable estimates of species diversity, we excluded grid cells with few occurrence records from analysis (Figure S1), that is, if the observed number of species was less than six, the number of sub-gridded cells with at least one incidence was less than six, or the total number of species incidences was equal to the number of unique species (species that are each detected in only one sub-grid cell).



**FIGURE 1** Global distribution of occurrence records of tropical Scleractinia stony corals and the evaluation procedure of species incidence to estimate species diversity. (a) Occurrence records of coral species derived from the Global Biodiversity Information Facility and Ocean Biogeographic Information System (109,367 points). Light-blue areas represent the coral reef distribution provided by ReefBase (<http://www.reefbase.org/main.aspx>). (b) Compilation of species incidence data: dividing the globe into grids of different size of (1, 2, 4, 6, 10 and  $15^\circ$ ); subdividing each grid into  $0.01 \times 0.01$  sub-gridded cells; counting the number of sub-gridded cells containing occurrence records in each grid for each species; creating species incidence distribution in each grid and calculating relevant values ( $U$  = sum of species incidences,  $T$  = number of sub-grids where at least one incidence was found,  $Q_2$  = number of duplicates,  $Q_1$  = number of uniques,  $S_{obs}$  = observed number of species)

## 2.3 | Environmental data

For analysis, we defined the potential habitats for coral species based on environmental conditions that included mean sea surface temperature (World Ocean Atlas: <https://www.nodc.noaa.gov/OC5/woa18/>) and water depth (ETOPO1: <https://www.ngdc.noaa.gov/mgg/global/>) at a 0.01° spatial resolution.

## 2.4 | Diversity estimation and comparison

We estimated the Hill's diversity indices ( ${}^qD$ ,  $q = 0, 1, 2$ ) based on the species incidence data in each grid cell at different spatial resolutions (1, 2, 4, 6, 10 and 15°) as:

$${}^qD = \left( \sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad (1)$$

where  $p_i$  is the relative incidence frequency of the  $i$ th species in a grid:  $i = 1, 2, \dots, S$ . That is,  $p_i$  represents the proportion of sub-grids that the  $i$ th species can be detected (i.e., at least one individual of the  $i$ th species can be detected) in the grid. The parameter  $q$  determines the sensitivity to species incidence frequencies (Chao et al., 2014):  ${}^0D$  is species richness, where frequency of incidence is not accounted for (i.e., each species has equal weight);  ${}^1D$  is the exponential of the Shannon entropy index, where species are weighted by their incidence frequency (i.e., highlighting frequent species);  ${}^2D$  is the inverse of the Simpson concentration index (i.e., highlighting highly frequent species). We evaluated the standard error of these estimates ( ${}^0D$ ,  ${}^1D$  and  ${}^2D$ ) by a bootstrapping method, and computed sampling completeness by sampling coverage (SC) with respect to species incidence (Chao et al., 2020 under review):

$$SC = 1 - \frac{Q_1}{U}(1-B)$$

$$B = \begin{cases} 2Q_2 / [(T-1)Q_1 + 2Q_2], & \text{if } Q_2 > 0 \\ 2 / [(T-1)(Q_1 - 1) + 2], & \text{if } Q_2 = 0, Q_1 > 0 \\ 1, & \text{if } Q_1 = Q_2 = 0 \end{cases} \quad (2)$$

where  $U$  is the total number of species incidences;  $T$  is the number of sub-gridded cells where at least one incidence was recorded; and  $Q_1$  and  $Q_2$  are the number of unique and the number of duplicates (species that are each detected in exactly two sub-grid cells), respectively. We refer to  ${}^qD$  at observed SC as reference diversities

( ${}^qD_{ref}$ ). Asymptote diversity  ${}^qD_{asym}$  was estimated at  $SC = 1$  under the assumption that sampling effort is hypothetically expanded to be large, and sample coverage approaches unity, by means of extrapolation.

Generally, when  $q = 0$ ,  ${}^qD_{asym}$  provides a minimum number of true species richness (or Chao 2 estimator; Chao, 1987), whereas for  $q = 1$  and 2,  ${}^qD_{asym}$  infers the true diversity in a grid. We also assessed the associated standard error for each asymptotic richness estimate to reflect sampling uncertainty. To fairly compare among the grid cells in terms of sampling completeness, we estimated  ${}^qD$  at the fixed SC ( ${}^qD_{SC\#}$ ) using 1 and 5% percentile of SC at doubled sample sizes (15–57%; see Table S1 for specific figures for each spatial resolution). The relatively small thresholds of SC were to minimize the extrapolation of more than doubled sample size: either rarefaction or extrapolation within doubled sample sizes was applied to 99% (or 95%) of grid cells.

## 2.5 | Biogeographical patterns

We mapped SC, the diversity estimates, and the standard errors at different spatial resolutions and examined their geographical patterns. We then investigated the patterns of the diversity estimates across latitudinal and longitudinal gradients using LOWESS smoothing splines. We also compared the smoothing curves of the diversity estimates with those derived from the expert range maps (Figure S2; Veron, 2000) for 690 species common with this study, which were resampled at corresponding spatial resolutions for the analysis.

## 2.6 | Optimal sampling effort allocation

From a statistical viewpoint, grid-cells with low SC can be considered as priority areas for additional sampling. Meanwhile, in practice, our research effort is defined by the number of occurrence records, and its total amount has a limitation in a unit of time. In addition, a biased allocation of research effort may promote the difference between under-sampled and over-sampled sites, which make a fair comparison difficult. Therefore, a sampling strategy which effectively fills spatial knowledge gaps is helpful in conservation planning. Here, we simulated an optimal sampling strategy, based on an optimization analysis, to fill knowledge gaps in the current occurrence dataset of coral species. We identified spatial priority areas for future additional sampling to effectively improve the spatial bias of sampling completeness. We used grid cells at a 1° resolution because a high-resolution map would be required from a practical point

of view. In the simulation, we assumed that the potential number of incidences in a grid cell (i.e., “ $T$ ” in the equation for SC) was represented by the number of sub-gridded cells ( $0.01^\circ$ ) which were available habitat for coral species in terms of sea surface temperature ( $18\text{--}36^\circ\text{C}$ ) and sea depth ( $-40$  to  $0$  m). The simulation was carried out using the following procedures: we (a) calculated the current inequality of SC using the Gini index ( $\frac{2\sum_{i=1}^n iSC_i}{n\sum_{i=1}^n SC_i} - \frac{n+1}{n}$ ), where  $SC_i$  is SC in the  $i$ th cell of  $n$  cells sorted by ascending order; (b) added a species incidence in each grid cell according to the relative species incidence frequencies in the cell, where we assumed that an unrecorded species has the same relative frequency as that of uniques; (c) calculated SC; (d) calculated a potential increase in SC for each grid cell and subsequent change in the Gini-index; (e) allocated a new sample to the cell in which an additional sample has most impact on the Gini index; and (f) repeated these procedures until we reached a given limit of sample-size. We assumed a sample limit of 15,000, which roughly corresponds to the annual average of the number of registered samples of stony corals in GBIF for the 10 years between 2008 and 2018 (129,133 records). We repeated this simulation 30 times.

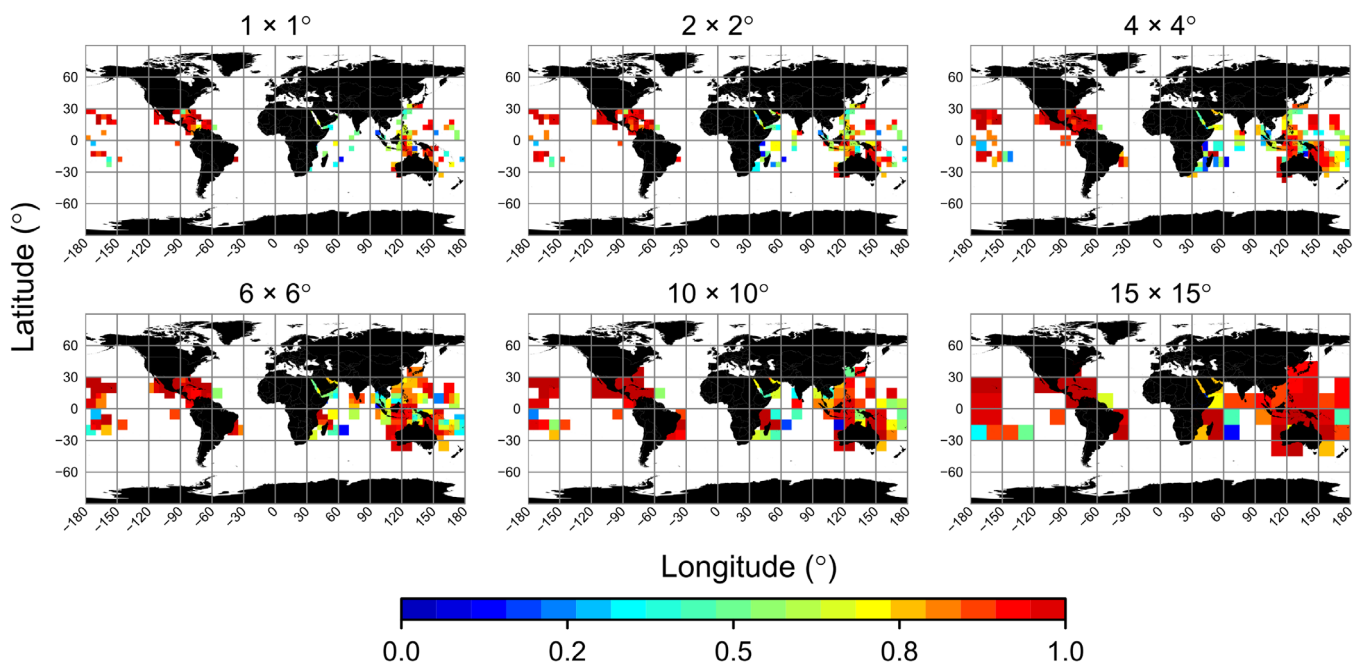
All analyses and graphic works were conducted using R and related packages: “`rgbif`” for retrieving occurrence records from GBIF (Chamberlain et al., 2019), “`iNEXT`” for biodiversity estimation (Hsieh et al., 2016), “`maptools`” (Bivand & Lewin-Koh, 2019), “`raster`”

(Hijmans, 2018), and “`spatialEco`” (Evans, 2018) for editing raster data.

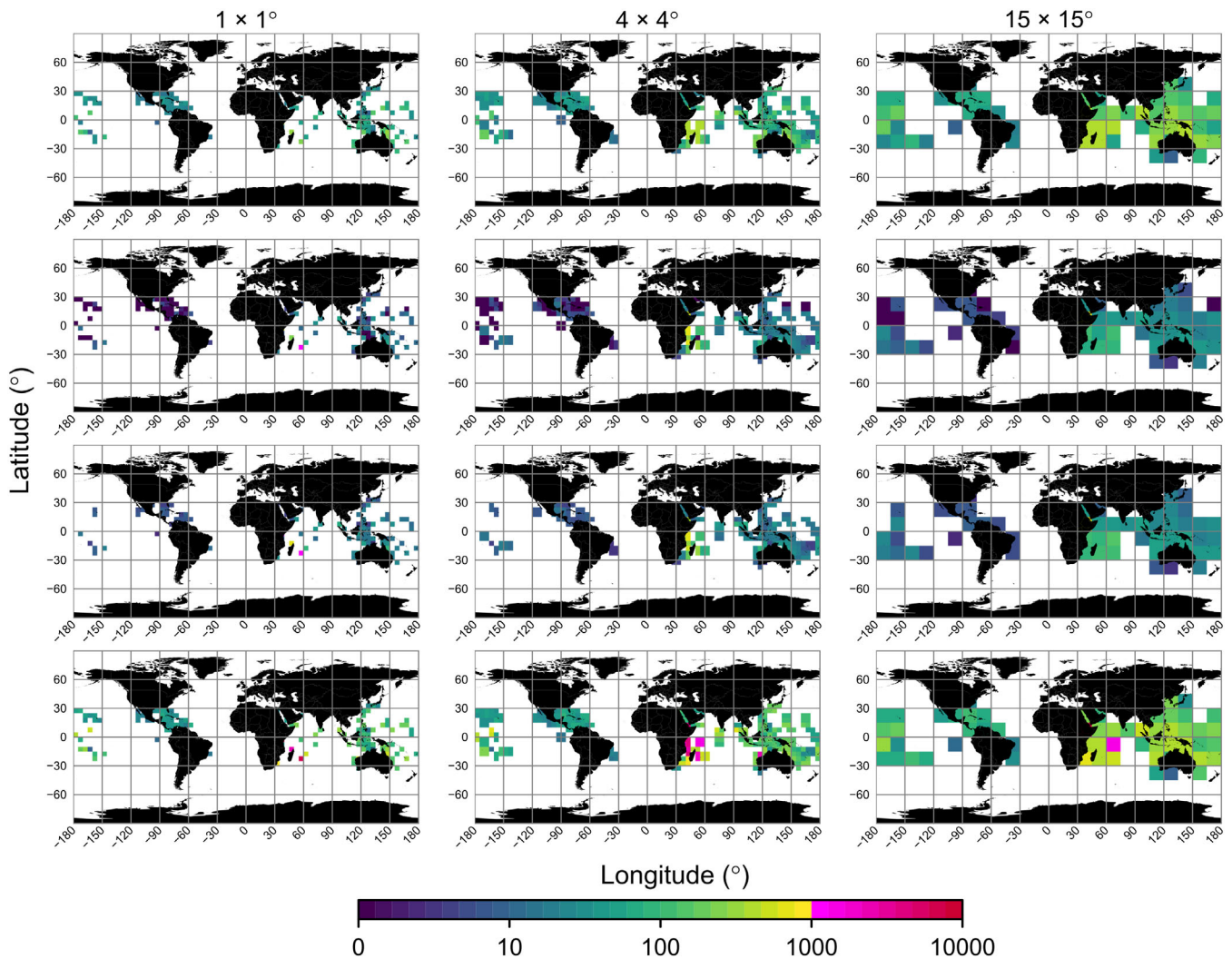
### 3 | RESULTS

Sampling completeness (measured by sampling coverage, SC) of coral species occurrence was relatively consistent across the spatial resolutions ( $>0.8$  on average; Figure S3a). SC was especially high in coastal areas around Australia and Central America, while relatively low on the coast of Madagascar and some Polynesian islands (Figure 2). Standard errors of diversity estimations were not affected by spatial resolutions (Figure S3b–d).

The maps of observed species diversity ( ${}^0D_{ref}$ ) showed areas of high species richness on the north coast of Australia and in the western Indian Ocean (Figures 3 and S4). The estimation error was highest on the coast of Madagascar compared with other regions, and consistently so across different spatial resolutions (Figure S5). The diversities standardized for sampling completeness ( ${}^0D_{SC1}$ ,  ${}^0D_{SC5}$  and  ${}^0D_{asym}$ ) were consistently greater in the western Indian Ocean at all spatial resolutions, and relatively lower on the coast of Central America and in the eastern Pacific. In general, those diversity patterns with the species diversity of the smallest order ( $q = 0$ ) were similar to those with the species diversities of larger orders ( $q = 1, 2$ ) (Figure S6–S15).



**FIGURE 2** Geographic maps of sampling coverage (SC) at different spatial resolutions (1, 2, 4, 6, 10 and  $15^\circ$  cells). Note, all maps are drawn at  $5 \times 5^\circ$  resolution for visibility



**FIGURE 3** Geographical patterns of coral species diversity (Hill's number of order  $q = 0$ ) at different spatial resolutions (1, 4 and 15° cells). Diversity is estimated at four sampling coverages (SC): observed SC ( ${}^0D_{ref}$ ) (top row), 1% percentile of SC values of doubled sample size ( ${}^0D_{SC1}$ ) (second from the top row), 5% percentile of SC values of doubled sample size ( ${}^0D_{SC5}$ ) (third row) and infinite SC (i.e., asymptote diversity,  ${}^0D_{asym}$ ) (bottom row). Exact values of percentiles of SC values are listed in Table S1. Maps for other spatial resolutions (2, 6 and 10° cells) are provided in Figure S4. For visibility, maps with higher spatial resolutions (1 and 4° cells) are drawn at 5 × 5° resolution. Values are log10-transformed. Potential outliers (i.e., extremely high values compared with others) are represented by reddish color scale

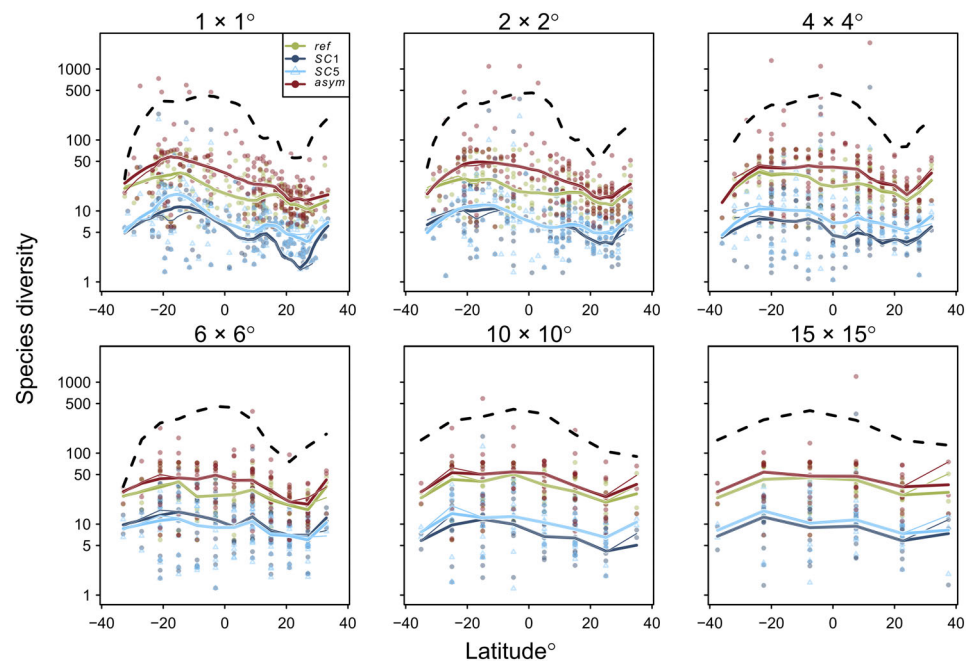
The observed ( ${}^qD_{ref}$ ,  $q = 0, 1, 2$ ) and SC-standardized species diversities ( ${}^qD_{SC1}$ ,  ${}^qD_{SC5}$  and  ${}^qD_{asym}$ ,  $q = 0, 1, 2$ ) showed a latitudinal gradient. This peaked at around 20°S (including northern Australia and Madagascar) and decreased towards the south and north poles, especially at finer spatial resolutions (Figures 4, S16, S17). This latitudinal gradient became non-distinct (nearly flat) at coarser spatial resolutions.

Longitudinal diversity gradients were also found (Figures 5, S18, S19). The observed diversities ( ${}^qD_{ref}$ ,  $q = 0, 1, 2$ ) showed that a unimodal-shaped longitudinal pattern peaked at around 120–150°E (including eastern Australia) at finer spatial resolutions. In contrast, the SC-standardized diversities ( ${}^qD_{SC1}$ ,  ${}^qD_{SC5}$  and  ${}^qD_{asym}$ ,  $q = 0,$

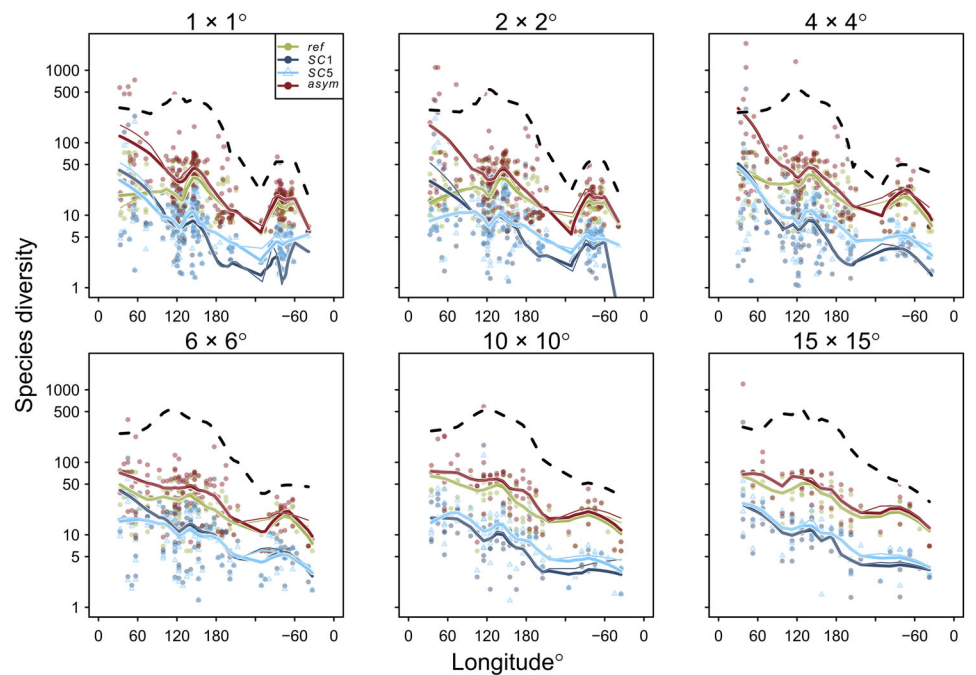
1, 2) showed a peak near 40°E (western Indian Ocean) and decreased linearly eastwards with a spike around 150°E at finer spatial resolutions. This spike disappeared at the coarser spatial resolutions.

Latitudinal and longitudinal patterns of diversity estimates using species occurrence data differed substantially from diversity patterns derived from the expert range maps (dashed lines on Figures 4 and 5). Species range maps showed a near symmetrical unimodal latitudinal pattern peaking at the equator and between 120 and 180°E longitude. These discrepancies in the diversity patterns between the two data types were more distinctive at finer than coarser spatial resolutions.

**FIGURE 4** Latitudinal patterns of coral species diversity (Hill number of order  $q = 0$ ) estimated at different spatial resolutions (1, 2, 4, 6, 10 and  $15^\circ$  cells). Diversity is estimated at four sampling coverages (SC): observed SC (*ref*), 1% percentile of SC values of doubled sample size (*SC1*), 5% percentile of SC values of doubled sample size (*SC5*) and infinite SC (i.e., asymptote diversity; *asym*). LOWESS curves are also shown. Thick and thin lines represent entire cells and cells located on coral reef areas by ReefBase (<http://www.reefbase.org/main.aspx>). The broken line is the number of species calculated by stacking expert range maps (Veron, 2000)



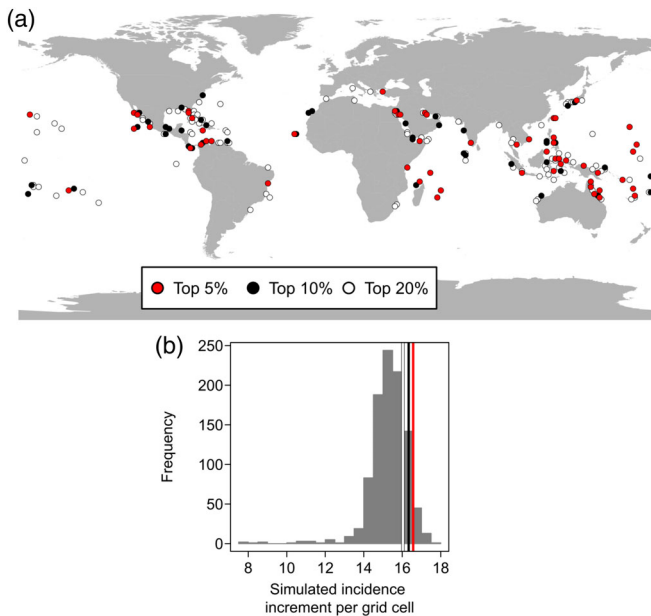
**FIGURE 5** Longitudinal patterns of coral species diversity (Hill number of order  $q = 0$ ) estimated at different spatial resolutions (1, 2, 4, 6, 10 and  $15^\circ$  cells). Diversity is estimated at four sampling coverages (SC): observed SC (*ref*), 1% percentile of SC values of doubled sample size (*SC1*), 5% percentile of SC values of doubled sample size (*SC5*), and infinite SC (i.e., asymptote diversity; *asym*). LOWESS curves are also shown. Thick and thin lines represent entire cells and cells located on coral reef areas as recorded by ReefBase (<http://www.reefbase.org/main.aspx>). The broken line is the number of species calculated by stacking expert range maps (Veron, 2000)



Results of genus-level analyses were generally similar to those of species-level analyses described above. Coral genus diversity, estimated by controlling SC, was greater in the western Indian Ocean and on the northern coast of Australia (Figures S20–S23). The latitudinal and longitudinal patterns of genus diversity were also similar to those of species diversity, while the discrepancy at the diversity peak of  $0^\circ$ E between the data types was smaller than in species-level analyses (Figures S24 and S25).

Simulation analysis revealed priority areas for additional sampling that not only improved SC but

also reduced the spatial bias of SC among sites (Figure 6). The average SC was improved from 0.39 to 0.79, and the Gini index of SC among the grid cells decreased from 0.53 to 0.13. Some cells prioritized for additional sampling occurred in all regions with corals (Figure 6). The biogeographical survey priorities (top 5, 10, and 20% fractions) were mainly found on the northern coast of Australia, center of the Indo-Pacific Coral Triangle, western Indian Ocean, central areas of the Red Sea and Persian Gulf, and coast of Central America.



**FIGURE 6** Spatial priority areas for future sampling to fill biodiversity knowledge shortfalls of tropical coral species at  $1^\circ$  gridded cells: (a) map; (b) histogram of simulated optimal sampling in each grid cell. The top 5, 10 and 20% fractions of priority grid cells are shown. In the simulation, the number of additional samples is set as 15,000, roughly corresponding with the average number of Scleractinia stony coral specimens registered to GBIF in the 10 years between 2008 and 2018 ( $N = 129,133$ ). The sampling simulation was repeated 30 times, with results per grid cell averaged

## 4 | DISCUSSION

The diversity estimation using the global-scale occurrence data of Scleractinia stony corals controlled the inequality of sampling completeness and revealed fairly comparable species diversity at different spatial scales. We found scale-dependent patterns of coral diversity despite relatively coarse spatial resolutions ( $\geq 1^\circ$  grid cell: about  $111 \text{ km}^2$  at the equator). Latitudinal diversity gradients were more distinct at finer spatial resolutions (Figure 4) while longitudinal gradients were more or less consistent across the spatial resolutions. Such scale-dependent patterns in species diversity suggest shift of ecological and/or evolutionary processes (Willis & Whittaker, 2002) and idiosyncratic relationships between local assemblages and regional species pools (Karlson & Cornell, 1998). Our findings may support a hypothesis of coral diversity hotspots, which states that local species diversities are regulated by idiosyncratic assembly processes, while an extensive species pool was formed by overlapping species distribution ranges at a large scale (Bellwood, Renema, & Rosen, 2012; Gaither & Rocha, 2013). The size of species pools and species sorting from

them to local assemblages may be regulated by region-specific biogeographical settings, including environmental conditions, habitat availability, and geological histories including changes in sea-level (Bellwood, 2001; Costello & Chaudhary, 2017; Leprieur et al., 2016; Renema et al., 2008; Sanciangco et al., 2013; Tittensor et al., 2010).

Interestingly, we found the richest species pool (gamma diversity) and greatest local (alpha) diversity in the western Indian Ocean, comparable or even larger than those of the Coral Triangle. The Indian Ocean includes regions with high species endemism, such as the Red Sea, extended Indo-west Pacific, and South Africa (Costello et al., 2017), and the presence of high diversity areas in the western parts (around Madagascar) has been reported (e.g., Ateweberhan & McClanahan, 2016; Connolly, Bellwood, & Hughes, 2003). At larger spatial resolutions ( $10^\circ$  and  $15^\circ$ ), the species richness even at observed sampling coverage (i.e., no extrapolation nor interpolation) was the highest at the western Indian Ocean. This finding was supported by a regional-scale diversity estimation in the Coral Triangle and near-Madagascar (Chao et al., under review). To confirm such regional diversity bias, a further investigation of alpha/gamma diversity (but also from a view point of species turnover, beta diversity) is needed. In fact, the high diversity areas in the western Indian Ocean were characterized by a low sampling coverage and high estimation error (Figure 2 and S4), suggesting the possibility of overestimation due to extrapolation, especially at finer spatial resolutions.

Large-scale species occurrence datasets (mostly presence-only data) are strongly influenced by sampling bias and thus tend to suffer from false-absence (omission errors) because of insufficient sampling effort (Hermoso et al., 2015). Indeed, we found a significant spatial bias in occurrence records: the sampling completeness was spatially inequivalent. Such spatial knowledge gaps theoretically can be improved by using the diversity estimations standardized by sampling completeness, as we examined for coral diversity ( ${}^qD_{SC1}$ ,  ${}^qD_{SC5}$  and  ${}^qD_{asym}$ ). Therefore, these estimates allow fair comparisons of species diversity between places with different sampling effort (Chao & Jost, 2012). Meanwhile, stacked species range maps contain many commission errors and thereby overestimate the number of co-occurring species (Hurlbert & White, 2005; Jetz et al., 2019). Such a difference in data types in terms of species geography affects the geographic pattern of species diversity (Hurlbert & White, 2005). Indeed, our estimates of coral diversity patterns based on species incidence across a global-scale gradient were inconsistent with those derived from expert range maps (Figure S2; Veron, 2000), especially at finer spatial resolutions. A



contrasting pattern of coral diversity between species occurrence and range map data (e.g., weak latitudinal diversity gradient) suggests that the two datasets showed different aspects of coral diversity. The former described scale-dependent structure of local (alpha) diversities across spatial scales, while the latter consistently described the maximal size of a species pool at a site (i.e., gamma diversity). These should be similar patterns when spatial resolution becomes larger, as shown in Figures 5 and 6. Furthermore, this finding of contrasting diversity estimates from point versus range map data are likely to apply to other taxa.

In macroecological studies of Scleractinian corals, the Linnean/Darwinian shortfalls are another critical issue, especially for estimating species diversity (Fukami, 2015). Nonetheless, our genus-level diversity estimates showed a similar diversity pattern (Figure S20–S25). This suggests that our observed biogeographical patterns were relatively robust to taxonomic errors, in contrast to the Wallacean shortfall (deficiency of geographical occurrence), at least in the current nomenclatural system (i.e., WoRMS; Horton et al., 2019).

We also identified geographical survey priorities by means of an optimization analysis that accounted for the spatial bias of sampling completeness (Figure 6). Ideally, coral sampling effort should be allocated across the world in proportion to the diversity of a region; intuitively, more diverse regions may need more sampling to estimate their species richness. We found many unexplored areas with limited species incidence (Figure S1), low SC (Figure 2) and/or high estimation errors (Figures S5, S10 and S15). Most importantly, the northeastern coast of Australia, center of the Coral Triangle, coast of Madagascar, central areas of the Red Sea and Persian Gulf, and coast of Central America and higher latitudinal regions (e.g., Japanese archipelago) were identified as coral survey priorities (Figure 6) to effectively fill knowledge gaps in coral diversity. These priority areas may be consistent with macro-ecologically important areas with evolutionary potential for coral reef diversity (Spano, Hernández, & Rivadeneira, 2016) or frontlines of poleward shift in coral distribution in response to global warming (Mizerek, Baird, Beaumont, & Madin, 2016).

In our analysis, we only used the two major databases of GBIF and OBIS, but there may be other unknown data sources as yet not digitized and included in any database (Page, MacFadden, Fortes, Soltis, & Riccardi, 2015). For example, Speed et al. (2013) mobilized regional data from published literature and unpublished data to examine coral reef dynamics in Western Australia. In this sense, our sampling completeness may underestimate our potential knowledge. Therefore, the priority areas that

we identify (Figure 6) are regarded as important coral areas not only for additional field sampling but also for effective mobilizing occurrence data.

## 5 | CONCLUSION

The estimated species diversity of Scleractinia stony corals showed a distinct scale-dependent pattern along latitudinal and longitudinal gradients that differed substantially from previous range-map-based diversity patterns, especially at finer spatial scales. The geographical pattern of coral species diversity estimated at various spatial-scales may shed new light on gaps in our knowledge. Greater species diversity was identified in the western Indian Ocean, for example, than in the Coral Triangle in the Indo-Pacific region, which has been argued to represent a coral diversity hotspot. The theory of biodiversity estimation statistically connects the occurrence records (which reflect our sampling effort), sampling completeness, the state of true diversity, and the reliability of diversity estimates. This framework has successfully evaluated a huge spatial bias in coral species occurrences and its sampling completeness. More importantly, it has informed us of a set of priority areas for future sampling of unknown coral species occurrence. Our findings of biogeographical survey priorities contribute to filling biodiversity shortfalls of tropical coral reefs through sampling completeness, and consequently to developing conservation planning.

## ACKNOWLEDGEMENTS

We are grateful to members of the data management team of Kubota-lab, the University of the Ryukyus, for their support with data compilation. We thank the individuals and institutions supporting regional and national nodes of the Ocean Biogeographic Information System and Global Biodiversity Information Facility for making data freely available online without which the present research would not have been possible. This study was financially supported by the JSPS KAKENHI (no. 16H02755) and Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers, by the JSPS KAKENHI.

## REFERENCES

- Asaad, I., Lundquist, C. J., Erdmann, M. V., & Costello, M. J. (2018). Delineating priority areas for marine biodiversity conservation in the coral triangle. *Biological Conservation*, 222, 198–211. <https://doi.org/10.1016/j.biocon.2018.03.037>
- Ateweberhan, M., & McClanahan, T. R. (2016). Partitioning scleractinian coral diversity across reef sites and regions in the

- Western Indian Ocean. *Ecosphere*, 7(5), <https://doi.org/10.1002/ecs2.1243>
- Bellwood, D. R., Hughes, T. P., Connolly, S. R., & Tanner, J. (2005). Environmental and geometric constraints on Indo-Pacific coral reef biodiversity. *Ecology Letters*, 8(6), 643–651. <https://doi.org/10.1111/j.1461-0248.2005.00763.x>
- Bellwood, D. R., Hughes, T. P., Folke, C., & Nyström, M. (2004). Confronting the coral reef crisis. *Nature*, 429(6994), 827–833. <https://doi.org/10.1038/nature02691>
- Bellwood, D. R. (2001). Regional-scale assembly rules and biodiversity of coral reefs. *Science*, 292(5521), 1532–1535. <https://doi.org/10.1126/science.1058635>
- Bellwood, D. R., Renema, W., & Rosen, B. R. (2012). Biodiversity hotspots, evolution and coral reef biogeography. In D. Gower, K. Johnson, J. Richardson, B. Rosen, L. Ruber, & S. Williams (Eds.), *Biotic evolution and environmental change in Southeast Asia* (pp. 216–245). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511735882.011>
- Bivand, R., & Lewin-Koh, N. (2019). Maptools: Tools for handling spatial objects. *R Package Version 0.9-9*. <https://doi.org/https://CRAN.R-project.org/package=maptools>
- Boakes, E. H., McGowan, P. J. K., Fuller, R. A., Chang-qing, D., Clark, N. E., O'Connor, K., & Mace, G. M. (2010). Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. *PLoS Biology*, 8(6), e1000385. <https://doi.org/10.1371/journal.pbio.1000385>
- Brito, D. (2010). Overcoming the Linnean shortfall: Data deficiency and biological survey priorities. *Basic and Applied Ecology*, 11(8), 709–713. <https://doi.org/10.1016/j.baae.2010.09.007>
- Chamberlain, S., Oldoni, D., Barve, V., Desmet, P., Geffert, L., Mcglinn, D., & Ram, K. (2019). Rgbif: Interface to the global “biodiversity” information facility API. R package version 1.4.0. Retrieved from <https://cran.r-project.org/web/packages/rgbif/index.html>
- Chao, A. (1987). Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, 43(4), 783–791.
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45–67. <https://doi.org/10.1890/13-0133.1>
- Chao, A., & Jost, L. (2012). Coverage-based rarefaction and extrapolation: Standardizing samples by completeness rather than size. *Ecology*, 93(12), 2533–2547. <https://doi.org/10.1890/11-1952.1>
- Chao, A., Kubota, Y., Zelený, D., Chiu, C-H., Li, C-F., Kusumoto, B., ... Colwell, R. K. (under review). Quantifying sample completeness and comparing diversities among assemblages. *Ecological Research*.
- Connolly, S. R., Bellwood, D. R., & Hughes, T. P. (2003). INDO-PACIFIC BIODIVERSITY OF CORAL REEFS: DEVIATIONS FROM A MID-DOMAIN MODEL. *Ecology*, 84(8), 2178–2190. <https://doi.org/10.1890/02-0254>
- Costello, M. J., Tsai, P., Wong, P., Cheung, A., Basher, Z., & Chaudhary, C. (2017). Marine biogeographic realms and species endemism. *Nat Commun*, 8(1), 1057. <https://doi.org/10.1038/s41467-017-01121-2>
- Costello, M. J. (2015). Biodiversity: The known, unknown, and rates of extinction. *Current Biology*, 25(9), R368–R371. <https://doi.org/10.1016/j.cub.2015.03.051>
- Costello, M. J., & Chaudhary, C. (2017). Marine biodiversity, biogeography, {Deep-Sea} gradients, and conservation. *Current Biology*, 27(11), R511–R527. <https://doi.org/10.1016/j.cub.2017.04.060>
- Costello, M. J., Vanhoorne, B., & Appeltans, W. (2015). Conservation of biodiversity through taxonomy, data publication, and collaborative infrastructures. *Conservation Biology*, 29(4), 1094–1099. <https://doi.org/10.1111/cobi.12496>
- Couce, E., Ridgwell, A., & Hendy, E. J. (2012). Environmental controls on the global distribution of shallow-water coral reefs. *Journal of Biogeography*, 39(8), 1508–1523. <https://doi.org/10.1111/j.1365-2699.2012.02706.x>
- Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. *Trends in Ecology & Evolution*, 28(12), 689–695. <https://doi.org/10.1016/j.tree.2013.09.003>
- Evans, J. (2018). spatialEco. *R Package Version 0.1.1-1*. Retrieved from <https://cran.r-project.org/package=spatialEco>
- Fukami, H. (2015). Regional specific approach is a next step for setting evolutionary-based conservation priorities in the scleractinian corals. *Animal Conservation*, 18(4), 318–319. <https://doi.org/10.1111/acv.12224>
- Gaither, M. R., & Rocha, L. A. (2013). Origins of species richness in the indo-Malay-Philippine biodiversity hotspot: Evidence for the Centre of overlap hypothesis. *Journal of Biogeography*, 40(9), 1638–1648. <https://doi.org/10.1111/jbi.12126>
- García-Roselló, E., Guisande, C., Manjarrés-Hernández, A., González-Dacosta, J., Heine, J., Pelayo-Villamil, P., ... Lobo, J. M. (2015). Can we derive macroecological patterns from primary global biodiversity information facility data? *Global Ecology and Biogeography*, 24(3), 335–347. <https://doi.org/10.1111/geb.12260>
- Gaston, K. J. (2000). Global patterns in biodiversity. *Nature*, 405(6783), 220–227. <https://doi.org/10.1038/35012228>
- Girardello, M., Martellos, S., Pardo, A., & Bertolino, S. (2018). Gaps in biodiversity occurrence information may hamper the achievement of international biodiversity targets: Insights from a cross-taxon analysis. *Environmental Conservation*, 45(4), 370–377. <https://doi.org/10.1017/S0376892918000115>
- Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution*, 19(9), 497–503. <https://doi.org/10.1016/j.tree.2004.07.006>
- Hermoso, V., Kennard, M. J., & Linke, S. (2015). Assessing the risks and opportunities of presence-only data for conservation planning. *Journal of Biogeography*, 42(2), 218–228. <https://doi.org/10.1111/jbi.12393>
- Hijmans, R. J. (2018). Raster: Geographic data analysis and Modeling. *R Package Version*, 2, 8–4. Retrieved from <https://cran.r-project.org/package=raster>
- Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, 54(2), 427–432. <https://doi.org/10.2307/1934352>
- Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven shortfalls that beset

- large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, 46(1), 523–549. <https://doi.org/10.1146/annurev-ecolsys-112414-054400>
- Hortal, J., Jiménez-Valverde, A., Gómez, J. F., Lobo, J. M., & Baselga, A. (2008). Historical bias in biodiversity inventories affects the observed environmental niche of the species. *Oikos*, 117(6), 847–858. <https://doi.org/10.1111/j.0030-1299.2008.16434.x>
- Horton, T., Kroh, A., Ahyong, S., Bailly, N., Boyko, C. B., Brandao, S. N., ... Zhao, Z. (2019). World Register of Marine Species. <https://doi.org/10.14284/17>
- Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, 7(12), 1451–1456. <https://doi.org/10.1111/2041-210X.12613>
- Huang, D. (2012). Threatened reef corals of the world. *PLoS One*, 7(3), e34459. <https://doi.org/10.1371/journal.pone.0034459>
- Hurlbert, A. H., & White, E. P. (2005). Disparity between range map- and survey-based analyses of species richness: Patterns, processes and implications. *Ecology Letters*, 8(3), 319–327. <https://doi.org/10.1111/j.1461-0248.2005.00726.x>
- Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., ... Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, 3(4), 539–551. <https://doi.org/10.1038/s41559-019-0826-1>
- Jetz, W., McPherson, J. M., & Guralnick, R. P. (2012). Integrating biodiversity distribution knowledge: Toward a global map of life. *Trends in Ecology & Evolution*, 27(3), 151–159. <https://doi.org/10.1016/j.tree.2011.09.007>
- Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2), 363–375. <https://doi.org/10.1111/j.2006.0030-1299.14714.x>
- Karlson, R. H., & Cornell, H. V. (1998). Scale-dependent variation in local vs. regional effects on coral species richness. *Ecological Monographs*, 68(2), 259. <https://doi.org/10.2307/2657203>
- Leprieux, F., Descombes, P., Gaboriau, T., Cowman, P. F., Parravicini, V., Kulbicki, M., ... Pellissier, L. (2016). Plate tectonics drive tropical reef biodiversity dynamics. *Nature Communications*, 7(1), 11461. <https://doi.org/10.1038/ncomms11461>
- Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., ... Antonelli, A. (2015). Estimating species diversity and distribution in the era of big data: To what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973–984. <https://doi.org/10.1111/geb.12326>
- Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. *Nature Communications*, 6(1), 8221. <https://doi.org/10.1038/ncomms9221>
- Mizerek, T. L., Baird, A. H., Beaumont, L. J., & Madin, J. S. (2016). Environmental tolerance governs the presence of reef corals at latitudes beyond reef growth. *Global Ecology and Biogeography*, 25(8), 979–987. <https://doi.org/10.1111/geb.12459>
- Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29(2), 215–233. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9)
- Moilanen, A., Wilson, K. A., & Possingham, H. (2009). *Spatial conservation prioritization: Quantitative methods and computational tools*. Oxford: Oxford University Press.
- Page, L. M., MacFadden, B. J., Fortes, J. A., Soltis, P. S., & Riccardi, G. (2015). Digitization of biodiversity collections reveals biggest data on biodiversity. *Bioscience*, 65(9), 841–842. <https://doi.org/10.1093/biosci/biv104>
- Ponder, W. F., Carter, G. A., Flemons, P., & Chapman, R. R. (2001). Evaluation of museum collection data for use in biodiversity assessment. *Conservation Biology*, 15(3), 648–657. <https://doi.org/10.1046/j.1523-1739.2001.015003648.x>
- Reddy, S., & Dávalos, L. M. (2003). Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography*, 30(11), 1719–1727. <https://doi.org/10.1046/j.1365-2699.2003.00946.x>
- Renema, W., Bellwood, D. R., Braga, J. C., Bromfield, K., Hall, R., Johnson, K. G., ... Pandolfi, J. M. (2008). Hopping hotspots: Global shifts in marine biodiversity. *Science*, 321(5889), 654–657. <https://doi.org/10.1126/science.1155674>
- Sánchez-Fernández, D., Lobo, J. M., Abellán, P., & Millán, A. (2011). How to identify future sampling areas when information is biased and scarce: An example using predictive models for species richness of Iberian water beetles. *Journal for Nature Conservation*, 19(1), 54–59. <https://doi.org/10.1016/j.jnc.2010.05.003>
- Sanciangco, J. C., Carpenter, K. E., Etnoyer, P. J., & Moretzsohn, F. (2013). Habitat availability and heterogeneity and the Indo-Pacific warm Pool as predictors of marine species richness in the tropical Indo-Pacific. *PLoS One*, 8(2), e56245. <https://doi.org/10.1371/journal.pone.0056245>
- Spano, C. A., Hernández, C. E., & Rivadeneira, M. M. (2016). Evolutionary dispersal drives the latitudinal diversity gradient of stony corals. *Ecography*, 39(9), 836–843. <https://doi.org/10.1111/ecog.01855>
- Speed, C. W., Babcock, R. C., Bancroft, K. P., Beckley, L. E., Bellchambers, L. M., Depczynski, M., ... Wilson, S. K. (2013). Dynamic stability of coral reefs on the west Australian coast. *PLoS One*, 8(7), e69863. <https://doi.org/10.1371/journal.pone.0069863>
- Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, 466(7310), 1098–1101. <https://doi.org/10.1038/nature09329>
- Troia, M. J., & McManamay, R. A. (2017). Completeness and coverage of open-access freshwater fish distribution data in the United States. *Diversity and Distributions*, 23(12), 1482–1498. <https://doi.org/10.1111/ddi.12637>
- UNEP. (2010). Strategic plan for biodiversity 2011–2020 and the Aichi targets. Retrieved from 534 [http://scholar.google.com/scholar?q=Strategic plan for biodiversity 2011–2020 and the 535 Aichi targets&btnG=&hl=en&num=20&as\\_sdt=0%2C22VN - readcube.com](http://scholar.google.com/scholar?q=Strategic+plan+for+biodiversity+2011-2020+and+the+535+Aichi+targets&btnG=&hl=en&num=20&as_sdt=0%2C22VN-readcube.com)
- Veron, J. (2000). *Corals of the world* (Vol. 295, pp. 1–3). Townsville: Australian Institute of Marine Sciences.
- Veron, J. (2013). Overview of the taxonomy of zooxanthellate Scleractinia. *Zoological Journal of the Linnean Society*, 169(3), 485–508. <https://doi.org/10.1111/zoj.12076>
- Willis, K. J., & Whittaker, R. J. (2002). Species diversity—scale matters. *Science (New York, N.Y.)*, 295(5558), 1245–1248. <https://doi.org/10.1126/science.1067335>
- Yang, W., Ma, K., & Kreft, H. (2013). Geographical sampling bias in a large distributional database and its effects on species

richness-environment models. *Journal of Biogeography*, 40(8), 1415–1426. <https://doi.org/10.1111/jbi.12108>

### SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Kusumoto B, Costello MJ, Kubota Y, et al. Global distribution of coral diversity: Biodiversity knowledge gradients related to spatial resolution. *Ecological Research*. 2020;35:315–326. <https://doi.org/10.1111/1440-1703.12096>