

Completeness of the global census of marine fish diversity

Camilo Mora, Ransom A. Myers, Derek Tittensor

Biology Department, Dalhousie University, Nova Scotia B3H 4J1, Canada

Spatial variations in the numbers of species are among the most studied patterns in ecology (1) and are of particular interest in pinpointing priority areas for conservation (2-3). In the marine environment, diversity patterns are increasingly being described for a variety of taxa at regional, interoceanic and global scales (3-9). Interest in describing and understanding these patterns has gained momentum because of the increasing effect of human-related threats and the need to prioritise limited resources for conservation (2-3). At the most basic level, these spatial patterns of biodiversity are built upon data collected by censusing the species occurring at given locations. Census data is used to calculate the number of species occurring at different sites or as source of information to extrapolate the extent of occurrence and area of occupancy of species (1). A limitation that has been largely overlooked in this regard is the quality (or completeness) of taxonomic inventories (10-12). Here we quantify the completeness of available biogeographical data for constructing global-scale patterns of the diversity of marine fishes. We found that the global census of marine fishes is relatively complete with ~1500 species remaining to be described. However, at smaller resolutions taxonomic inventories were notably incomplete, particularly among supposedly diverse tropical areas and in developing countries. All inventories nearing completion were in few developed countries. At a fine spatial resolution, less than 3.5% of the surface of the world's oceans has inventories that are over 80% complete. The precarious nature of data currently available raises significant caution over diversity patterns

that have been previously described and over the effectiveness of strategies that are in place for the conservation of global marine biodiversity.

For this analysis, we used the records deposited in the Ocean Biogeographic Information System (OBIS)(13). OBIS is a global concatenation of taxonomic databases and therefore constitutes the largest source of biogeographical data available for marine fishes. We used all of the ~2.1 million records available for marine fish species that contained geographical positions and dates. We calculated the completeness of inventories within square cells across a range of spatial scales from 30 to 360 covering all the world's oceans and within all Exclusive Economic Zones (EEZs). Records were assigned to a given cell or EEZ based on their geographical position, and the dates of these records used to construct the accumulation curve of species over time within each cell or EEZ. To reduce biases introduced by the variety of methods used to collect the data, as well as discontinuities in collection effort, accumulation curves in each cell or EEZ were “smoothed” with a rarefaction approach in which an “average curve” was calculated from 50 curves generated through a random ordering of the years sampled. This smoothed curve represents the statistical expectation for the corresponding curve (10-11). We fitted 11 non-linear models to the resulting smoothed curves (Table 1) and calculated the percent completeness of the inventory in each cell or EEZ by dividing the total number of species currently reported within each cell or EEZ by the expected number of species predicted by the best-fitting model (as predicted by the coefficient of determination of the model).

At the global scale there are currently 12,943 marine fish species described (Fig. 1, Table 1). The model that best fitted the temporal accumulation of these species indicated that the completeness of this inventory is ~88%; i.e., around 1,553 marine fish species remain to be described worldwide (Fig. 1, Table 1). The extent of species that remain to be described is larger in the open and deep ocean (i.e. bathypelagic,

Comment [DT1]: Might it be useful to include some description of the range of variability in this prediction – i.e. the standard deviation of the model or something like that?

bathydemersal and benthopelagic habitats) and smaller in shallower and coastal areas (i.e. reefs and demersal habitats) (Table 1). These variations in the completeness of inventories among habitats likely reflect variations in the accessibility and facility to sample those habitats. The fauna of pelagic fish species had also been well inventoried (Table 1). This has been previously attributed to the large body size of most species in this environment and their ease of capture (14).

We found that the spatial distributions of species already described are remarkably unknown, particularly at finer spatial resolutions (Fig. 2c-f). At a relatively broad 360 resolution, 30% of the world's oceans area has inventories below 50% completeness (Fig. 2c). In contrast, at a finer 30 resolution, over 90% of the world's oceans has inventories below that level of completeness (Fig. 2f). Inventories over 80% complete occur in only 10% and 3.5% of the world's oceans at spatial resolutions of 360 and 30, respectively (Fig. 2 c,f). The incompleteness of taxonomic inventories at finer resolutions highlights a basic flaw in current biogeographical and conservation research. First, large-scale diversity patterns are built upon data of species recorded at smaller resolutions (5-14). With only 3.5% of the world's oceans correctly sampled at a relative fine resolution (i.e. with a conservative 80% of their fish fauna sampled), an emerging question is how reliable are the patterns that have been already described? Secondly, these patterns have been used as the dependent variable in studies seeking to assess the causality of ecological mechanisms (4-9). If existing diversity patterns are flawed by failing to accurately depict true diversity, then support for inference of causal mechanisms from these data should be treated with extreme caution. Finally, these results also cast doubt on the effectiveness of conservation strategies that are aimed at protecting biodiversity (3). For conservation research the issue is particularly relevant because threats to biodiversity often occur at small scales and it is at those scales that decisions are made and where data is particularly precarious (Fig. 2f).

Incomplete taxonomic inventories were not distributed uniformly in space. Globally, most records for marine species have been collected in the near vicinity of continental coasts and are concentrated within the EEZs of few countries (Fig. 2a). Consequently, existing records yield taxonomic inventories over 80% complete for the coasts of only eight countries (i.e. Canada, New Zealand, UK, Greenland, USA, Australia, Bermuda and Spain). Sixty-four countries have inventories between 50% and 79% complete and 154 countries have inventories below 49% completion (Fig. 2b). All the high-seas have inventories below 50% completion. Remarkably, tropical areas, which are supposedly “well-known” for their diversity, have among the worst completeness of all taxonomic inventories (Fig. 2c-f). This pattern of incompleteness was also reflected among all marine habitats (Fig. 3a-e).

Compared to other marine organisms, marine fishes are very likely among the best-studied taxa. Significant efforts have been made to improve and synthesize their taxonomy (e.g. 16) and to index information on their biology, distribution and diversity (e.g. 10,17). Unfortunately, these efforts fall short of what is necessary to describe the worldwide diversity of marine fishes with reasonable accuracy, even at relatively coarse spatial scales. The problem intensifies with finer spatial resolution and in tropical regions where much of the diversity is concentrated. The fact that fishes are one of the best-studied marine taxonomic groups suggests that the situation may well be worse among other taxa. With the current increase of human induced pressures on biodiversity (e.g. overexploitation, climate change, habitat loss and pollution, among others), coping with the uncertainty arising from incomplete data is a problem needing to be quickly overcome. This is going to need considerable renewing interest in taxonomy by both researchers and funding agencies and a more consistent effort among researchers and publishing journals to encourage the storage of raw data in available databases (e.g. OBIS).

1. Gaston, K.J., & T.M. Blackburn. *Patterns and processes in macroecology*. Blackwell Science, Oxford UK. (2000)
2. Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853-858 (2000)
3. Roberts, C.M. et al. Marine Biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280-1284 (2002)
4. Bellwood, D.R., & T. Hughes. Regional-assembly rules and biodiversity of coral reefs. *Science* **292**, 1532-1535 (2001)
5. Mora, C., Chittaro, P.M., Sale, P.F., Kritzer, J.P., & Ludsin, S.A. Patterns and processes in reef fish diversity. *Nature* **421**, 933-936 (2003)
6. Mora, C. & Robertson, D.R. Causes of latitudinal gradients in species richness: a test with the endemic shorefishes of the Tropical Eastern Pacific. *Ecology* **86**, 1771-1782 (2005)
7. Worm, B., Sandow, M., Oschlies, A., Lotze, H.K., Myers, R. Global patterns of predators diversity in the open oceans. *Science* **309**, 1365-1369 (2005)
8. Roy, K., Jablonski, D., Valentine, J.W. & Rosenberg, G. Marine latitudinal gradients: test of causal hypotheses. *Proc. Nat. Acad. Sci.* **95**, 3699-3702 (1998)
9. Macpherson, E. Large-scale species richness gradients in the Atlantic Ocean. *Proc. Royal Soc. London B.* **269**, 1715-1720 (2002)
10. Colwell, R. K. & Coddington, J.A. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. Lond. B.* **345**, 101-118 (1994)
11. Gotelli, N.J. & Colwell, R.K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379-391 (2004)

12. Soberon, J. & Llorente, J. The use of accumulation functions for prediction of species richness. *Cons. Biol.* **7**, 480-488 (1993)
13. www.iobis.org
14. Zapata, F. & Robertson, D.R. How many species of shore fishes are there in the Tropical Eastern Pacific?. *Journal of Biogeography* (in press).
15. Levin, S.A. The problem of pattern and scales in ecology. *Ecology* **76**, 1943-1967 (1992)
16. Eschmeyer, W.N. *The Catalog of Fishes*. California Academy of Sciences Press. USA. (1998)
17. Froese, R. & Pauly, D. Editors. *FishBase*. World Wide Web electronic publication. www.fishbase.org, version (2006).

Acknowledgements We thank the Ocean Biogeographic Information System and all its contributing databases for making their data freely available. Funding was provided by the Sloan Foundation Census of Marine Life through the Future of Marine Animal Populations and the Ocean Biogeographic Information System. We thank Peter Sale, Frederick Grassle and Walter Jetz for helpful comments.

Correspondence and requests for materials should be addressed to C.M. (cmora@dal.ca)

<

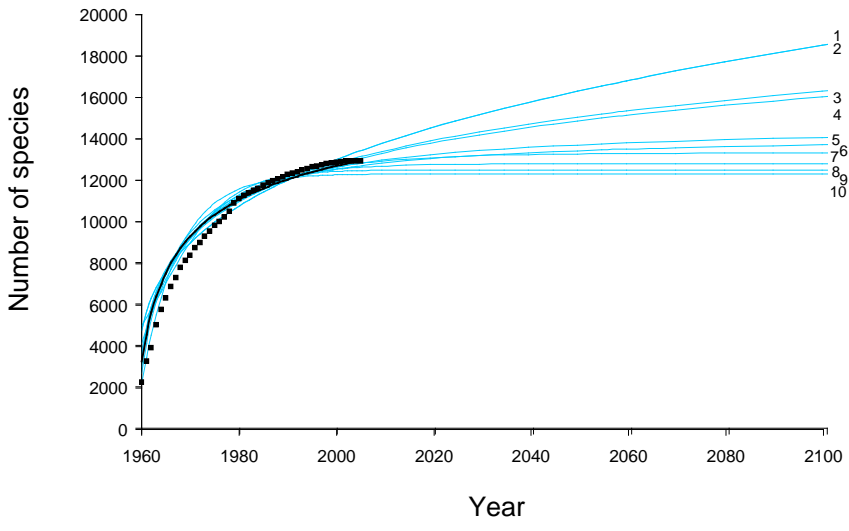
List of figures

Figure 1. Temporal accumulation of the description of marine fish species worldwide (dots). The empirical curve was “smoothed” (solid line) by calculating an average curve through the random ordering of years (50 repetitions). Numbers of the right-hand side of the graph indicate the predictions of the following non-linear models: (1) Hill, (2) Power, (3) Logarithmic B, (4) Exponential, (5) Rational, (6) Chapman-Richards, (7) Clench, (8) Asymptotic, (9) Logistic, (10) Negative exponential and Linear dependence.

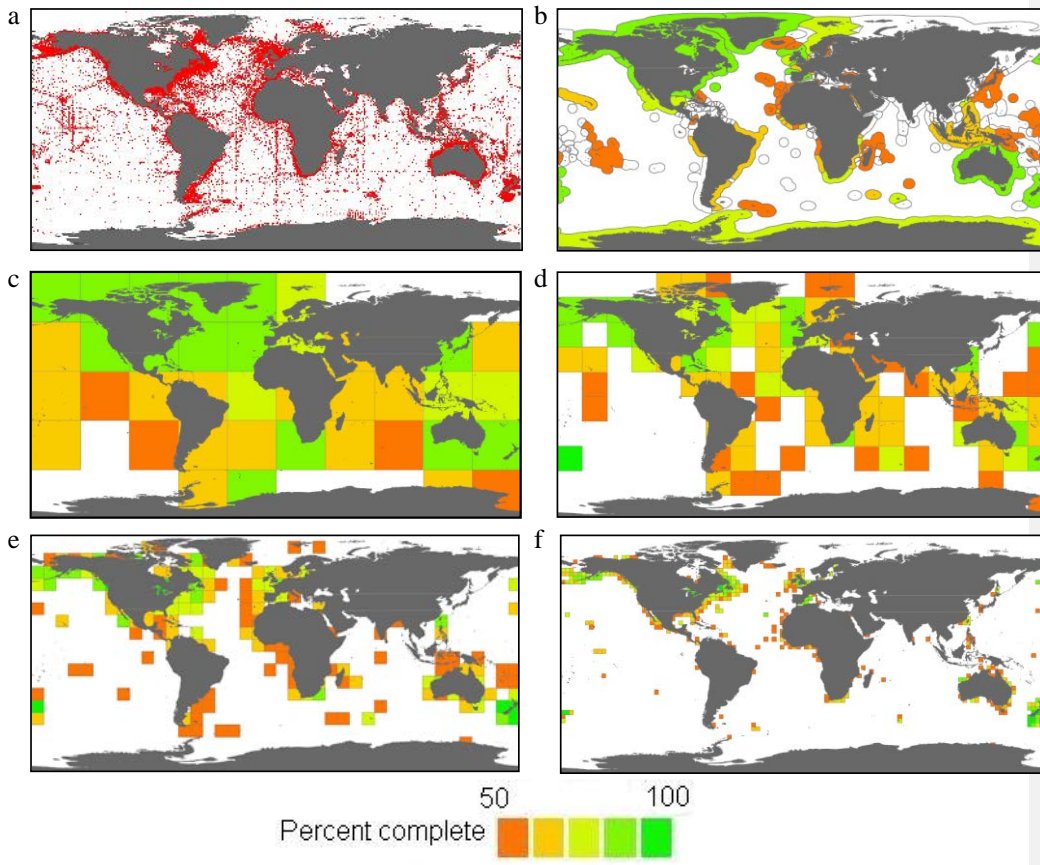
Figure 2. Taxonomic sampling of the marine fishes of the world. Plot (a) indicates the positions of the ~2.1 million marine fish records used in this study. Plots (b) to (f) indicate the completeness of the taxonomic inventories within the Exclusive Economic Zones of the world (b) and at resolutions of 36° by 36° (c), 18° by 18° (d), 9° by 9° (e) and 3° by 3° (f). White areas in the maps indicate locations for which inventories are below 50% complete or for which there are insufficient data. We show the results for cells whose inventories are over 50% complete because in most other cases below that percentage the slope of the curves does not reduce and therefore models tend to predict large values and incompleteness to be exaggerated.

Figure 3. Completeness of taxonomic inventories by marine habitats. Completeness of taxonomic inventories were determined at a 3° resolution for species known to be: pelagic (a), demersal (b), reef-associated (c), benthopelagic (d), bathypelagic (e) and bathydemersal (f). Species-specific habitat associations were obtained from (17). For a spatial reference of where

the different habitats may occur (i.e. all delimited cells) we selected all the cells where at least one species with a particular habitat association has been described



Mora et al Figure 1



Mora et al Figure 2

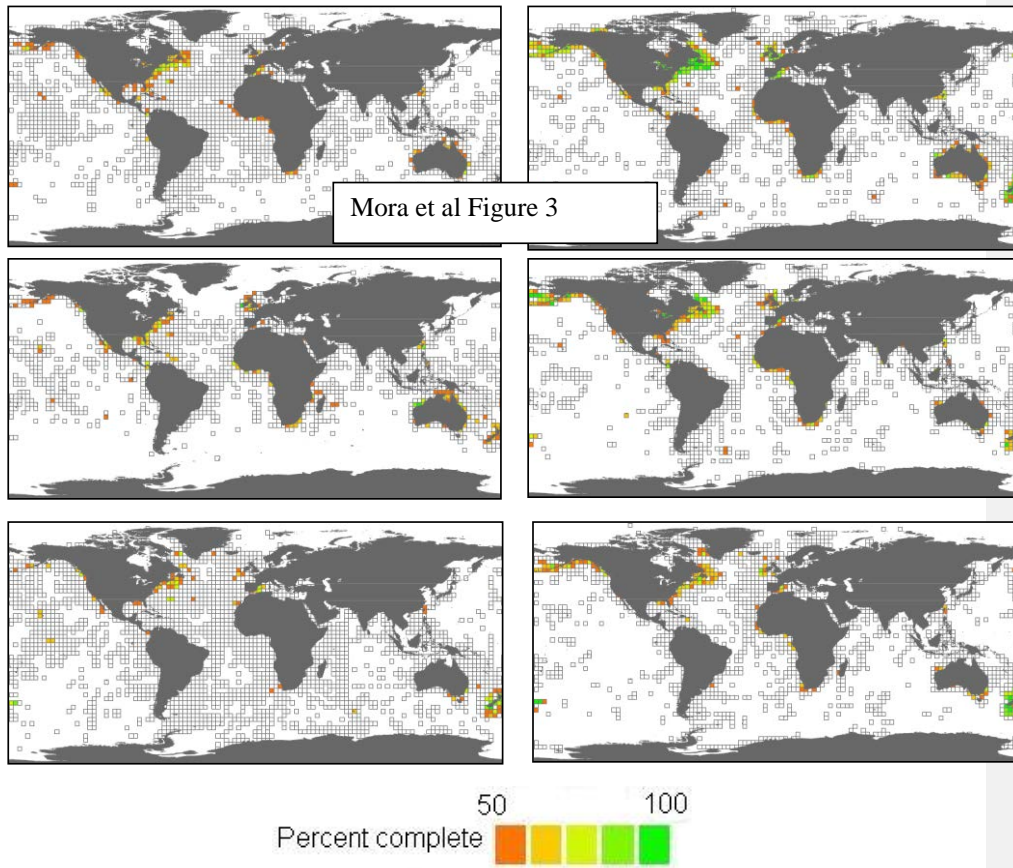


Table 1. Current completeness of the taxonomic inventories of marine fishes as a whole and according to the habitat association of the species. Fit of the models was judge based on the r^2 . We estimated asymptotes for each regression equation at the year 2500.

Species described	Completeness (%)												
	Model best fitting the accumulation curve	Clench $Y=(a*X)/(1+(b*X))$	Exponential negative $Y=a*(1-\exp((-b)*X))$	Linear dependence $Y=(ab)*(1-\exp((-b)*X))$	Exponential $Y=a+(b*\log(X))$	Power $Y=a*X^b$	Logarithmic B $Y=\ln(1+(a*b*X))/b$	Asymptotic $Y=b1-(b*(c^X))$	Chapman Richards $Y=a*(1-\exp(-b*X))^c$	Logistic $Y=a/(1+\exp(-b*c*X))$	Rational $Y=(a+(b*X))/(1+(c*X))$	Hill $Y=(a*b*(X)^c)/((1+b^c))$	
Bathydemersal	1626	61	87	100 [‡]	100 [‡]	66	41	62	98	90	100 [‡]	82	41
Bathypelagic	1104	69	95	100 [†]	100 [†]	69	53 [‡]	68	100	98	100 [‡]	92	53 [‡]
Benthopelagic	727	88	92	100 [†]	100 [†]	68	49 [‡]	66	100	97	100 [‡]	89	49 [‡]
Demersal	4678	95	89	100 [‡]	100 [‡]	66	44 [‡]	63	100	95	100 [‡]	86	44 [‡]
Pelagic	884	88	91	100 [‡]	100 [‡]	66	47 [‡]	64	100	98	100 [‡]	89	47 [‡]
Reef-associated	3924	93	95	100 [†]	100 [†]	68	56 [†]	68 [‡]	100	100	100 [‡]	93	56 [†]
All species	12943	88	92	100 [†]	100 [†]	67	48 [‡]	65	100	97	100 [‡]	88	48 [‡]

When no indicated r^2 were larger that 0.99

[‡] r^2 between .95 and .99

[†] r^2 between .90 and .95