



Global biogeographical data bases on marine fishes: caveat emptor

D. Ross Robertson

Smithsonian Tropical Research Institute, Balboa,
Republic of Panamá

ABSTRACT

A review of georeferenced collection-site records for Caribbean shore-fishes served by major online distributors of aggregated biodiversity data found large-scale errors in over a third of the species and genera, in nearly two-thirds of the families. To avoid compromising the value of their services to the global science community online providers must actively address the question of data quality.

Keywords

Marine fishes, global data bases, erroneous GIS data, faulty biogeographical analyses.

Correspondence: D. Ross Robertson,
Smithsonian Tropical Research Institute, Balboa,
Republic of Panamá. E-mail: drr@stri.org

An interlinked group of major online services supplies the global science community with various types of data bases on the biodiversity of terrestrial and marine organisms. Those data bases should be extremely useful for biogeographical analyses, provided that the data are accurate. Unfortunately, major errors are rife among georeferenced collection-site data (GIS data) on marine fishes served by the oldest such service, FishBase (<http://www.fishbase.org>), and by newer systems that serve data from FishBase. These include the Ocean Biogeographical Information System (<http://www.iobis.org>), the Global Biodiversity Information Facility (<http://www.gbif.org>) and the Encyclopedia of Life (<http://www.eol.org>).

When comparing OBIS-generated map displays of GIS data against published descriptions of the geographical ranges of shallow-water shore-fishes from the Greater Caribbean region, I frequently encountered obvious large-scale location errors. These included data-points in the wrong ocean, on the wrong side of the Atlantic, in the wrong hemisphere, in the centre of a deep ocean basin (inappropriate for inshore, demersal fishes) and in the centre of a continent. These observations stimulated a formal assessment of the extent, nature, distribution across major taxonomic groups and provenance of such errors. For this I examined the first species (alphabetic order) with OBIS GIS data in each genus of each family of Greater Caribbean shore-fishes, compared an OBIS map of those data to published descriptions of that species geographical range (Robins *et al.*, 1986; McEachran & Fechhelm, 1998, 2002; Carpenter, 2002; Floeter *et al.*, 2008; and individual species pages in FishBase at <http://www.fishbase.org>), and tallied the types of large-scale errors listed above. Furthermore, as FishBase and OBIS do not segregate GIS data on adults and pelagic larvae for any species, I also tallied records of larval data-points far outside the limits of the adult range, i.e. cool temperate and mid-ocean records for inshore demersal tropical species. While pelagic larvae of demersal marine organisms often are carried on ocean currents well beyond

the boundaries of the adult range, data on larvae collected well outside that range are inappropriate for inclusion in analyses of the geographical distributions of adults.

I found that large-scale errors in OBIS GIS data are taxonomically widespread, with 36.8% of the species and genera in 64.7% of the families of sharks, rays and bony fishes having such adult and/or larval data errors (see Appendix S1 in Supporting Information). Most of those errors are from collections of adults (errors in 32.3% of species and genera, 60.8% of families), with obvious large-scale errors due to the admixture of larval data with adult data occurring in only 6.9% of species and genera and 11.8% of families. However, such larval data errors affect some taxa (notably eels) at relatively high rates (see Appendix S1). GIS data provided by FishBase to OBIS are responsible for errors in 96.8% of the species that have erroneous data. Errors in the distributions of adults of most (87.6%) species reflect faulty georeferencing of named sites (e.g. West African coordinates for Puerto Rico; see Appendix S1). Other errors are due to misidentifications, the admixture of data on adults of other species, or the use of outdated (at least 5 years old) taxonomy. These adult and larval data errors originated from 32 primary sources, including both FishBase and OBIS.

For this analysis I tallied only the most obvious large-scale errors, excluded likely errors that I could not fully resolve, and did not include smaller scale, but biogeographically still important errors, such as those within the Greater Caribbean. Hence overall error rates among GIS data on marine fishes served by FishBase and OBIS are higher than the present analysis indicates. This high frequency of errors likely is not unique to Greater Caribbean shore-fishes, as a number of the main primary sources of erroneous data (major museums) that are served by FishBase have worldwide collections of fishes and other marine organisms.

Any biogeographical analysis of fish distributions that uses GIS data on marine fishes provided by FishBase and OBIS 'as is' will be seriously compromised by the high incidence of species

with large-scale geographical errors. A major revision of GIS data for (at least) marine fishes provided by FishBase, OBIS, GBIF and EoL is essential. While the primary sources naturally bear responsibility for data quality, global online providers of aggregated data are also responsible for the content they serve, and cannot side-step the issue by simply including generalized disclaimers about data quality. Those providers need to actively coordinate, organize and effect a revision of GIS data they serve, as revisions by individual users will inevitably lead to confused science (which version did you use?) and a tremendous expenditure of redundant effort. To begin with, it should be relatively easy for providers to segregate all data on pelagic larvae and adults of marine organisms that they serve online. Providers should also include the capacity for users to post readily accessible public comments about the accuracy of individual records and the overall quality of individual data bases. This would stimulate improvements in data quality, and generate 'selection pressures' favouring the usage of better quality data bases, and the revision or elimination of poor-quality data bases. The services provided to the global science community by the interlinked group of online providers of biodiversity data are invaluable and should not be allowed to be discredited by a high incidence of known serious errors in GIS data among marine fishes, and, likely, other marine organisms.

REFERENCES

- Carpenter, K.E. ed. (2002) *The living marine resources of the western central Atlantic*. Food and Agriculture Organization species identification guide for fisheries purposes and American Society of Ichthyologists and Herpetologists Special Publication No. 5, Rome. Vols. 1–3.
- Floeter, S., Rocha, L., Robertson, D.R., Joyeux, J.-C., Smith-Vaniz, W., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcon, M., Bowen, B.W. & Bernardi, G. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22–47.
- McEachran, J.D. & Fechhelm, J.D. (1998) *Fishes of the Gulf of Mexico*. University of Texas Press, Austin, Texas, Vol. 1.
- McEachran, J.D. & Fechhelm, J.D. (2002) *Fishes of the Gulf of Mexico*. University of Texas Press, Austin, Texas, Vol. 2.
- Robins, C.R., Ray, G.C. & Douglass, J. (1986) *A field guide to Atlantic coast fishes of North America*. Houghton Mifflin, Boston, Massachusetts.

Editor: Omar Defeo

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Errors in Greater Caribbean fish distributions.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.