RAINER FROESE

4
The science in FishBase

INTRODUCTION

FishBase is an online information system with key information on all the known fishes of the world, i.e., over 30 000 species. This key information has been extracted, standardized, and evaluated by a team of specialists from over 40 000 scientific publications. The rationale and development of FishBase are presented in Palomares and Bailly (this volume), who demonstrate that FishBase, the successive editions of the book that document it (notably Froese and Pauly, 2000) and individual chapters therein are heavily cited in both grey and peer-reviewed scientific literature. Indeed, such information has been crucial for numerous high profile studies, including in high profile outlets such as *Science* and *Nature*.

Nevertheless, there have been suggestions that FishBase is a laudable exercise in compiling scientific information, similar to a scientific library, but that its creation and maintenance are not "science," or even "research." Using that logic, one could argue that the work done by all scientists who collect and standardize their data prior to analyzing them is not "science." Also, one could argue that the evaluation of published data prior to their encoding and the tagging of some estimates as "doubtful" (as done by the experienced FishBase encoders) is equivalent to the critical assessment performed by the authors of scientific reviews – who undoubtedly do science.

Rather than developing these arguments, however, I suggest that the scientific status of FishBase can be evaluated by establishing that, based largely on data extracted from FishBase, new insights have been made and published, and that the papers in question have been cited by other scientists. In the following, I present three examples, jointly illustrating the science of FishBase.

Ecosystem Approaches to Fisheries: A Global Perspective, ed. V. Christensen and J. Maclean. Published by Cambridge University Press.

FISHING DOWN MARINE FOOD WEBS

The first and most widely cited study, done with the help of FishBase, and which could not have straightforwardly been done without it, was a team publication in *Science* (Pauly *et al.*, 1998, 937 citations in Web of Science^{SM}, December 2009). For this, we used the Food and Agriculture Organization of the United Nations (FAO) time series of global catch data for over 1000 species and species groups, and assigned each to one of more than 200 trophic level estimates, as incorporated in FishBase from published diet compositions or from Ecopath models. For every year, we calculated the mean trophic level of the catch, weighted by the catch of the respective species or groups. Plotting mean trophic levels of global catches over time showed a continuous decline from 1970 on (and from 1950 when we omitted a single species, the Peruvian anchoveta).

This global trend is now widely known as "fishing down marine food webs," and is verified by many local studies based on more detailed catch data, but mostly using trophic level estimates from FishBase (Stergiou and Christensen, this volume). Figure 4.1, for example, documents "fishing down marine food webs" in the northwestern Atlantic, where it is particularly strong, due to both the complete collapse of a key high-trophic level species – northern cod – and the availability of detailed fisheries statistics (fishing down cannot be

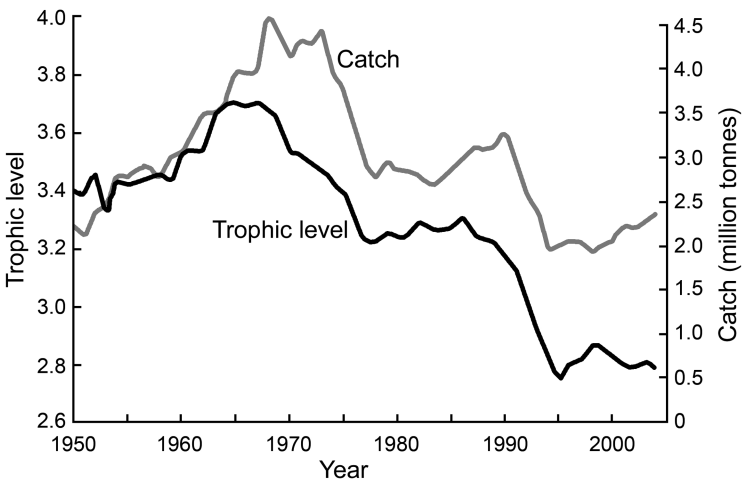

Figure 4.1 Catch and mean trophic level in the northwestern Atlantic.

documented without catch statistics that are well disaggregated). It is a disconcerting finding from the figure that both trophic levels and catches have been declining since the late 1960s. Stergiou and Christensen (this volume) may be consulted for more information on this, the first high-impact paper that used FishBase extensively.

PATTERNS AND PROCESSES IN REEF FISH DIVERSITY

My second example is a study by Mora *et al.* (2003) that examines three hypotheses about the geographic species richness of Indo-Pacific reef fishes. The authors at first used an existing database of Indo-Pacific coastal fishes compiled from published checklists for a previous study (Bellwood and Hughes, 2001). However, in the course of their study, Mora *et al.* (2003) detected that "[m]ore than 300 species were duplicated in the original database as a result of synonymy, misspelled names or misallocations of species to families." After they adopted the FishBase standard for scientific names (i.e., Eschmeyer's *Catalog of Fishes*), and the ensuing corrections to their original list, they then complemented their database with checklists from "the Philippines, Madagascar, Eastern Island, Cook Islands (all from www.fishbase.org)."

Mora *et al.* (2003) then showed that the mid-range points of the species' distributions occur over-proportionally in the Indonesian-Philippine region, thus refuting two hypotheses that proposed that speciation happened outside this area, with the high species richness stemming from the overlapping of the distributions' tails. They conclude that, in contrast to a widely held belief, "the processes of speciation, extinction and dispersal that yield large-scale patterns of species richness also seem to determine which species are present in local assemblages." Their paper was published in *Nature* and had 87 citations in the Web of Science℠ as of August 2010.

The enormous number of species that must be dealt with when performing analyses of this kind is illustrated in Figure 4.2, which shows the first-ever biodiversity transect across the Indo-Pacific at the equator, derived from the several thousand species in FishBase that have so far been mapped for the Indo-Pacific. Several known diversity patterns are nicely reproduced, such as the lower diversity on the eastern coasts of the Indian and Pacific oceans, the peaks of diversity in shallow waters, and the overall peaks in the Celebes/Halmahera region. The transect is, however, preliminary, as it underestimates the diversity at Celebes/Halmahera because many of the less-common

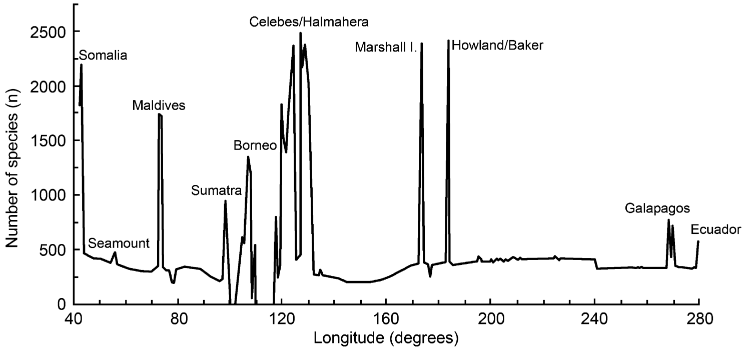

Figure 4.2 Preliminary transect of species richness per half-degree cell along the equator from Somalia to Ecuador, based on several thousand maps of Indo-Pacific fishes, marine mammals, and invertebrates.

species have not been mapped. Also, it overestimates diversity at Marshall and Howland/Baker Islands, because the observed restrictions on species distributions caused by distance from the center have not yet been included properly in the mapping algorithm (Kaschner *et al.*, 2007). Further, the abrupt drop in richness of mostly deep-sea species at 240 degrees latitude is an artifact caused by insufficient sampling of the Southeast Pacific: if no occurrence is reported from an FAO area (here: area 87), then the mapping algorithm prevents the species from spreading there. Despite its preliminary nature, the transect in Figure 4.2 clearly shows the explanatory potential of large datasets, such as those underlying the maps in FishBase.

FISHING ELEVATES VARIABILITY IN THE ABUNDANCE OF EXPLOITED SPECIES

Hsieh *et al.* (2006) explored the temporal variability of exploited versus unexploited fish stocks occurring in the same environment. These authors used larval surveys to estimate the abundance of adults, and FishBase to obtain most of the life-history traits of the adults. They found that exploited species showed higher variability in abundance in the same year and environment than unexploited species. This remained true when differences in life-history traits, such as maximum size, age and size at maturity, fecundity, duration of spawning period, and trophic level were taken into account. They concluded that the increased variability was “probably caused by fishery-induced truncation of age structure, which reduces the capacity of populations to

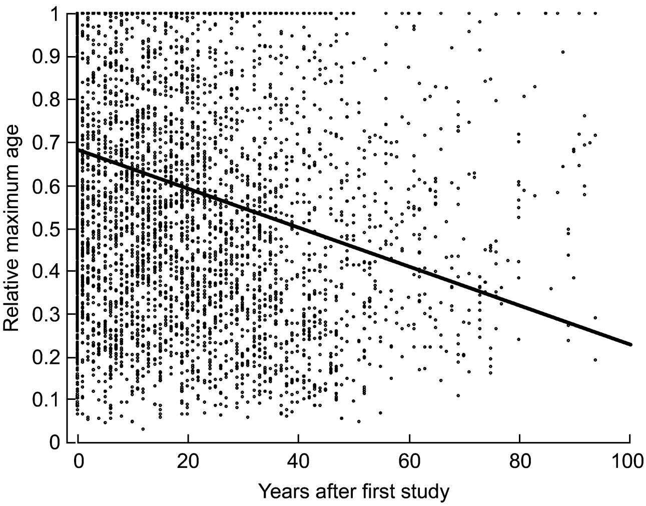

Figure 4.3 Maximum age in exploited fish populations (derived from growth studies as age at 95% of asymptotic length (L_∞), and shown as a fraction of the largest observed value) over years after the first study was done. A robust regression results in relative maximum age = 0.683 – 0.00454 years, with n = 5201 and $r^2 = 0.0928$.

buffer environmental events" and that "to avoid collapse, fisheries must be managed not only to sustain the total viable biomass but also to prevent the significant truncation of age structure." Their paper was published in *Nature* in 2006 and had received 65 citations in the Web of Science^{SM} as of August 2010.

That such truncation of ages occurs in exploited fish stocks can be shown using FishBase. Figure 4.3 shows the maximum ages in various stocks (as fractions of the overall maximum age recorded for the species in question) plotted against the time elapsed since the first study in a given species. While the scatter of these 5201 exploited populations is large, there is a significant decrease in maximum age over time, accounting for 9.3% of the variation in the data.

CONCLUSION

I believe the above examples give an unambiguous answer to the question of whether creating and maintaining FishBase is doing "science." FishBase contains – indeed, consists of – scientific data that were standardized and evaluated by scientists. It is used by numerous scientists to generate new knowledge in the peer-reviewed literature.

The relevance of this new knowledge is shown by the leading status of the journals and the citation record of the papers presenting it.

As an afterthought, it might be the fact that biologists are not yet used to working with large, international datasets that brought up this issue. I have never heard of oceanographers or meteorologists questioning whether their global datasets were part of science.

ACKNOWLEDGMENTS

I thank Daniel Pauly, once the “Science Advisor” of FishBase, for many years of discussing and improving the science in FishBase. Also, I would like to thank the FishBase team for nearly two decades of reliable encoding of data in FishBase. I thank especially Eli Agbayani, Josephine Rius, and Kathleen Reyes for assistance with Figure 4.2, and Crispina Binohlan for assembling the growth/age data shown in Figure 4.3.

REFERENCES

Bellwood, D. R. and Hughes, T. P. (2001) Regional-scale assembly rules and biodiversity of coral reefs. *Science*, 292, 1532–1535.

Froese, R. and Pauly, D. (2000) *FishBase 2000: Concepts, designs and data sources*. Los Baños, Laguna, Philippines: ICLARM.

Hsieh, C., Reiss, C. S., Hunter, J. R., Bedlington, J. R., May, R. M. and Sugihara, G. (2006) Fishing elevates variability in the abundance of exploited species. *Nature*, 443, 859–862.

Kaschner, K., Ready, J. S., Agbayani, E., Rius, J., Kesner-Reyes, K., Eastwood, P. D., South, A. B., Kullander, S. O., Rees, T. and Close, C. H. (2007) AquaMaps: predicted range maps for aquatic species. Available online: www.aquamaps.org.

Mora, C., Chittaro, P. M., Sale, P. F., Kritzer, J. P. and Ludsin, S. A. (2003) Patterns and processes in reef fish diversity. *Nature*, 421, 933–936.

Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. and Torres Jr, F. (1998) Fishing down marine food webs. *Science*, 279, 860.