



The science in FishBase

Daniel Pauly, Senior Advisor*

* ACP-EU FishBase Project, International Center for Living Aquatic Resources Management (ICLARM), PO Box 2631, 0718 Makati City, Philippines

keywords: fishbase; metataxonomy; research; iucn 'red list'; biodiversity; acp-eu fisheries research initiative



Daniel Pauly
Doctorate and habilitation in fisheries/marine sciences;
University Professor and Principal Science Advisor,
ICLARM

Introduction

FishBase is an electronic encyclopaedia on fish whose development and dissemination, funded mainly by the European Community, were discussed in several previous issues of this Bulletin (Box 1).

FishBase was conceived in the mid 1980s as a database that would contain biological information, including growth and mortality parameters of fish, to facilitate management-oriented research on major commercial species. As such, its conception went beyond that of the computerised fisheries databases then existing, which were limited to bibliographies, or barely annotated list of names.

Guiding principles in conceptualisation

Contrary to the 'flat file' structure of these earlier databases, and thank to the foresight of Rainer Froese, its major architect, FishBase started, in 1989, as a relational database, which allowed combining the contents of its various tables, and thereby generate new products and insights. Thus, for example, by combining the fish in FishBase that are marked as included in the IUCN "List of Threatened Animals" with a FishBase list of the fishes occurring in a given country, one gets, without further ado, a "Red List" for the country in question. This sounds simple and trivial, and at first, only obvious relations were identified, based on the first attributes that were entered for large number of species.

However, FishBase has grown enormously since it was originally designed. This is due both to data inputted by the FishBase team, equivalent to about 50 person-years of data encoding, and, increasingly, also to complementary databases supplied by collaborators

(e.g., on the scientific nomenclature of all fishes known so far, by W.E. Eschmeyer, or on all fish introductions, by R. Welcomme), jointly corresponding roughly to another 50 person-years worth of encoding.

In the process, FishBase became the largest database in the world for several of its component tables (e.g., nomenclatural information, genetics, growth parameters, recruitment time series), thus creating the potential for more relations being established among the entries of its tables. Two recent trends are now enabling us to realise this potential:

- the work of R. Froese on methods to monitor and facilitate the work of taxonomists, leading to a new subdiscipline which may be called "metataxonomy"; and
- the development, driven mainly by this author, of interactive graphs giving new quality to the data in FishBase, and which have turned this database into a new, powerful tool for all disciplines related to ichthyology and to biodiversity and fisheries research.

The Emergence of 'Metataxonomy'

One of the tasks of taxonomists is to identify/describe new species and fit them (i.e., 'classify' them) into extant, or revised taxonomies. This task involves dealing both with museum specimens, representing species or subsets thereof, and with the published literature, including old or otherwise inaccessible documents, in which these specimens and (sub)species are described. Computerised tools to facilitate this work - now increasingly needed in view of major threats to the world's biodiversity - have tended to be limited to lists of specimens in various museums, and to bibliographies.

It is therefore much to the credit of Rainer Froese to have developed, in the last years and implemented through FishBase a series of new 'meta-taxonomic' tools, consisting, among others, of:

- a routine for checking list of scientific names, using a combination of taxonomic rules (as included in the International Code of Zoological Nomenclature), the fish classification of W.E. Eschmeyer, logic, and some

smart 'tricks' to identify valid names of fishes, given synonyms and/or misspelled names;

- a routine for dynamically linking the original names in publications with current valid names, such that the updates that are generated from this are themselves automatically updated. This can be used to generate, e.g., current 'erratum sheets' for older taxonomic works.
- a routine to compute, based on (ii), the fraction of names in a given publication that are still valid. This allows, via comparisons with the mean replacement rate of valid names, a quantitative evaluation of the 'quality' of a given taxonomic work, given its publication date and the fraction of its names that are still valid.



Box 1. Details on previous contributions on FishBase in the "Bulletin"

Pauly, D. and R. Froese, 1991. The FishBase project - or how scattered information on fish can be assembled and made useful for research and development. *EC Fish. Coop.Bull.*, 4(4):1-6 [Describes structure of early version of FishBase, goals and mode of cooperation with collaborators]

Pauly, D. and R. Froese, 1995. EC Project reviewed in *Nature*. *EC Fish.Coop.Bull.*, 8(3):4-5. [Comments on and partly reproduces very positive review of FishBase vers. 1.2 by R.A. Mc Call and R.M. May in *Nature* (Vol. 374, p. 735); mentions that few points of criticism would be accommodated in subsequent releases, as indeed happened]

Vakily, J.M., R. Froese, M.L.D. Palomares & D. Pauly, 1997. Fisheries and biodiversity management: How can ACP countries meet the challenge? *EC Fish.Coop.Bull.*, 10(1):4-6 [presents a strategy and a project for disseminating FishBase in ACP countries through training courses and regional 'nodes', and for further development of FishBase, notably by involving European museums in the repatriation of biodiversity data]



The last example, dealing with the quality of taxonomic work - something rather elusive, and which was not quantifiable in earlier times - brings us to the second series of developments concerning the Science in FishBase.

Graphs and data quality

The word 'quality' has two major meanings; of these, one refers to the relative value of an item, as due, e.g., to the work that went into its creation. The other meaning generally refers to the 'properties' of items.

Thus, claiming, as we do, that FishBase adds new 'quality' to data is making the double claim that data, through their incorporation in FishBase, become value-added and thereby acquire new properties. This occurred as a gradual process during the evolution of FishBase.

The first obvious step is that a graph representing numerical data in a database allows to identify outliers, due either to:

- an encoding error; or
- an error in the document from which data were extracted;
- an attribute not previously considered that causes a legitimate deviation from the trend identifying the outlier; or
- the outlier is 'correct', and it is the trend and its underlying data which are erroneous.

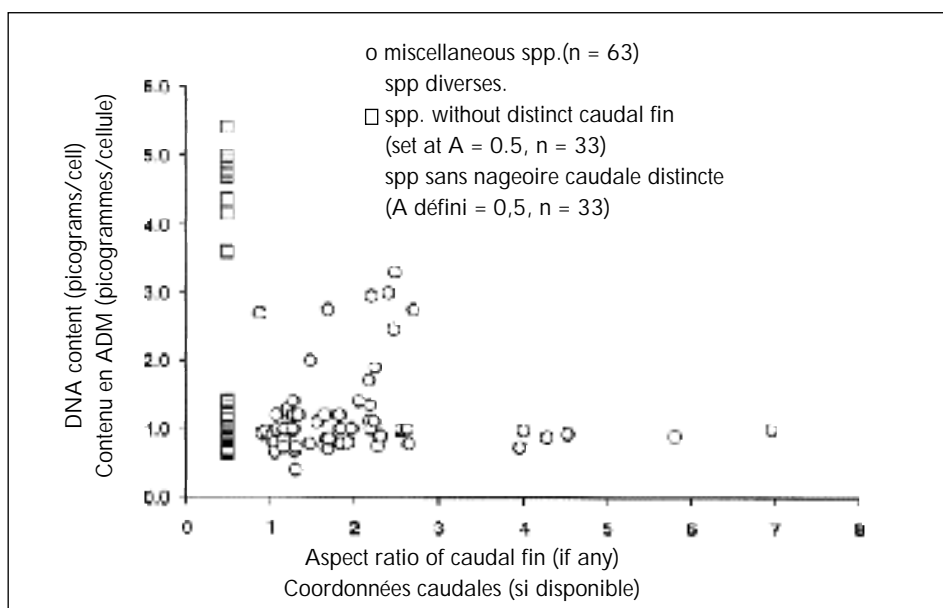


Fig. 1 - Plot of DNA contents per cell (a measure of cell size, see Box 2) vs. the aspect ratio of the caudal fin of fish (a measure of their metabolic level). This graph, generated by FishBase '97, tests an important hypothesis concerning the evolution of fishes (see Box 2). - Graphique des contenus en ADN par cellule (une mesure de la taille des cellules, voir Encadré 2) comparé aux coordonnées caudales des poissons (une mesure de leur niveau de métabolisme). Ce graphe, produit par FishBase 97 vérifie une hypothèse importante concernant l'évolution des poissons (voir Encadré 2).

Resolving which of (a) - (c) applies is usually straightforward. On the other hand, should (d) apply (it has not occurred to date, but still

may), this would imply an important discovery of some sort. Either way, 'quality' is added to the data in a way that would never have



Box 2. Testing a hypothesis on DNA cell contents and metabolism of fish

The DNA content of plant and animal cells is extremely variable. Only one broad generalisation has so far emerged which can be used to predict the amount of DNA in the cells of a given group of organisms: that the DNA content of cells tends to vary with cell size, implying a rough proportionality between the amount of DNA per cell, and the amount of living cellular material involved in various syntheses controlled by that DNA (Cavalier-Smith, 1991).

This implies that DNA content per cell, as recorded in the GENETICS table of FishBase, is a measure of cell size.

Organisms with small cells tend to have high metabolic rates, and conversely. However, as pointed out by Cavalier-Smith (1991), there is a lower limit to the size of cells: the fact that capillaries (which are formed by single cells) cannot have a diameter much smaller than that of red blood cells.

Thus, one can hypothesise that a plot of DNA content vs. the caudal aspect ratio of fish (an index of metabolic intensity, see the SWIMMING table of FishBase) should have on the

left side of the plot a wide range of DNA contents associated with low aspect ratios (including aspect ratio set at 0.5, to represent fish which do not use the caudal fin as their main organ of propulsion, and which tend to have low metabolic rates), and, on the right side of the plot, a narrow range of (low) DNA content associated with high aspect ratios (Pauly, Casal & Palomares, 1997). Fig. 1 displays this triangular feature, thus corroborating the hypotheses linking DNA content - via cell size - to metabolic rate.

References

- Cavalier-Smith, T. 1991. Coevolution of vertebrate genome, cell and nuclear sizes, p. 51-86. In G. Ghiara et al. (eds.) Symposium on the evolution of terrestrial vertebrate. Selected Symposia and Monographs. U.Z. I. 4, Modena.
- Hinegardner, R. and D.E. Rosen. 1972. Cellular DNA content and the evolution of teleostean fishes. *Am.Nat.*, 106(951):621-644.
- Pauly, D., C. Casal & M.L.D. Palomares, 1997. DNA, cell size and fish swimming. Box 24, p. 182 IN Froese, R. & D. Pauly (eds.). Fishbase 97: concepts, design and data sources. Manila, ICLARM

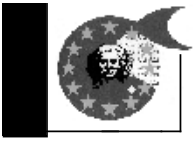


Box 3. Global trends in the mean trophic levels of fisheries catches

Trophic levels, ranging in fish from 2 (in herbivores and detritivores) to about 4.5 (in species feeding on carnivorous fishes) express the trophic role of organisms within ecosystems. Fisheries greatly impact the structure of ecosystems, and this can be shown by changes in the mean trophic levels of the fishes and invertebrates landed in the last decades.

FishBase 97 is now able to perform such analyses, and to output the relevant plots, as it includes estimates of trophic levels from a large number of Ecopath models (see this issue of the *Bulletin*), covering all groups included in the global FAO fisheries statistics that are also included in FishBase.

Fig 2 shows the last two decades of the trends which emerge from such analysis, for all marine and all freshwater fisheries of the world. More detailed analyses of these alarming trends can be performed with FishBase 97, and have recently been submitted to a major international journal.



become possible, were it not for their entry in the database, and their display through a graph.

The second step is that, by combining the entries of various tables, graphs can be generated which establish relationships that test simple or sophisticated hypotheses previously not tested, or not amenable to direct testing. An example is given in Fig.1 and in Box 2, which discusses the test of an hypothesis so far not tested in quantitative terms.

The third step is that, again by combining entries in various FishBase tables, graphs can be constructed documenting processes previously not known to occur, e.g., the steady decline of mean trophic level in the catch of global fisheries (see Fig. 2 and Box 3).

The two aspects of the new quality discussed here are obvious, as the graph in Fig. 2 present

a newly discovered process, yet is based on data not usually perceived as being of sufficient accuracy to enable sophisticated analyses.

Another aspect of quality as generated through FishBase is that it leads to (e)quality among its contributors. Thus, it does not really matter if a data point included in one of the FishBase graphs is derived from the work of a world famous scientist, or that of one of her students. Nor does it matter whether that data point was originally published in a prestigious international journal, or in a technical report of the Department of Fisheries of a small, isolated island country: either the point fits into a graph, or it doesn't. In the latter case, it must be treated as an outlier (see above).

Yet another aspect of this new (e)quality is the repatriation of data which FishBase is fostering. This will make available, in their countries of

origin (through its forthcoming dissemination phase; see Vakily *et al.* 1997 in Box 1), the many occurrence records and other information on fish presently locked in inaccessible publications, or the unpublished files of European and other museums.

We envisage that this data repatriation process will contribute to (e)quality by putting our colleagues in ACP countries, and other developing-country scientists on a more equal footing - at least with regards to access to such records - with those of developed countries. Also, we hope that this process will contribute to the quality of the management advice based - if only in part - on repatriated and other data in FishBase, thus making the Science in FishBase useful to all. ■