

Simple Identification Tools in FishBase

Nicolas Bailly, Rodolfo Reyes Jr., Rachel Atanacio, Rainer Froese

Abstract — Simple identification tools for fish species were included in the FishBase information system from its inception. Early tools made use of the relational model and characters like fin ray meristics. Soon pictures and drawings were added as a further help, similar to a field guide. Later came the computerization of existing dichotomous keys, again in combination with pictures and other information, and the ability to restrict possible species by country, area, or taxonomic group. Today, www.FishBase.org offers four different ways to identify species. This paper describes these tools with their advantages and disadvantages, and suggests various options for further development. It explores the possibility of a holistic and integrated computer-aided strategy.

Index Terms — databases, identification strategy, simple identification tools, web integration.



1 INTRODUCTION

FishBase [1] is a Global Species Database (GSD) and a Biodiversity Information System (BIS) on all extant fish species of the world, with about 32,000 valid species currently recognized as valid. It contains a wide range of information and data on biology, ecology, chorology, taxonomy, physiology, human uses, illustrations, etc. and aims at being the web global encyclopaedia on fishes. Four different types of identification tools are made available on the FishBase website (www.fishbase.org), which opens directly the search page:

- ‘Eye-balling’ drawings and key features by decreasing taxonomic level from class downward;
- Display of all pictures available for a given geographic area or a given family with possible restriction on fin ray meristics;
- Classic dichotomous keys; and
- Uses of simple morphometric ratios.

This paper gives a short description of the tools, how they work and can be used, where the users can find them, their advantages, disadvantages and

N. Bailly, R. Reyes Jr. and R. Atanacio are with the Aquatic Biodiversity Informatics Office, The WorldFish Center and the FishBase Information Research Group, Inc., Los Baños, 4031 Laguna, Philippines. E-mail: n.bailly@cgiar.org.

R. Froese is with the Leibniz Institute of Marine Sciences, IFM-GEOMAR, West Shore Campus, Düsternbrooker Weg 20, D-24105 Kiel, Germany. E-mail: rfroese@ifm-geomar.de.

limitations, and their possible improvements in the next future. Tools not yet in FishBase are also discussed. The need to integrate such tools in a common identification strategy is stressed.

2 TOOLS

2.1 QUICK IDENTIFICATION TOOL WITH OUTLINES

Description, how to use it, and where to find it. - The user is given the choice between boxes displaying simple outline drawings of species representing a given group, together with a short account of key characters. The user clicks on the box, and the boxes for the next level are displayed. For fishes, we start with 6 classes from the most likely ray-finned fishes to the less likely lampreys and hagfishes. Each class leads to its orders, and then each order to its families. Clicking on a family box leads to the "Identification through Pictures" pages described in the next section. It is possible to restrict the identification process to a large geographic area or to a country. This tool is in the top menu in the search page.

Advantages, disadvantages and limitations. - The main strength of the tool is that it is visual. The user is not obliged to read textual accounts. It is quite simple and useful to search for the family. It may be tricky to identify, e.g., eel-like shape groups that can be found in several classes and orders; here, reading the accounts may help. The number of typical species depicted in a box is limited in large groups, and sometimes well-known outlines but rare shapes are not included, such as coelacanths for the lobe-finned fishes box.

Further developments. - Below the family level, the outline drawings are generally not that useful, because species often have the same shape in a given family. There is a possible improvement using the subfamilies for the most species-rich families (Cyprinidae, Characidae, ...).

2.2 IDENTIFICATION THROUGH PICTURES

Description, how to use it, and where to find it. - The principle is to display one typical picture per species by area or taxonomic group. Clicking on a picture opens the corresponding species account with more information.

The tool is accessible from three sections in the search page as "Identification by pictures" under "Information by Family", "Information by Country / Island", and "Information by Ecosystem". For the first section, species within a family are displayed by alphabetical order of scientific names, which shows closely related species next to each other. From that page, it is possible to restrict the search by large area, and/or by the number of dorsal and anal spines. The broad distribution and maximum length of the species are listed as additional help. For the two last sections, the typical pictures of all species reported in the given geographic area are displayed using a traditional sorting of the fish classification, from hagfishes to coelacanths.

Advantages, disadvantages and limitations. - The main strength of this

tool is that it is visual. The user is not obliged to read texts. It is quite simple and useful to search within a family up to 100 species. But identification with pictures only may be difficult or, worse, misleading. As usual, users should read the species account to verify their identification, which is possible by clicking the picture. For groups with over 100 species the tool becomes problematic. Beyond 500 species, the answer time of the server may be prohibitive. Also, only half of the species have at least one usable illustration for identification in FishBase.

Further developments. - A major FishBase goal is to get at least one picture for every species. This is not easily achieved, because many species are only known from a few museum specimens, if any. The tool could be extended to subfamily and genus ranks for the most specie-rich families, and to orders and classes for those with a few species only. For the pages with geographic areas, there should be a taxonomic table of contents/menu at the top of the page, allowing the user to jump directly to the class, order or family (maybe genus) when he knows it.

2.3 COMPUTER-AIDED DICHOTOMOUS KEYS

Description, how to use it, and where to find it. - These are dichotomous keys digitized as they were published in FAO catalogues, revisions, field guides, and major ichthyofaunas. We have developed our own simple database and webpage format for these keys, but we also use Lucld Phoenix for an enhanced interface (see [2] for a description). For the FishBase format, couplet numbers, character text, and number of the next couplet, number of the previous couplet, illustration of the character or of the species (+name) stand in 5 columns in a row of an html table (it corresponds more or less to the database table format). Clicking on the number of the next couplet leads to the corresponding row in the table, clicking on the species name leads to the species account. The tool is accessible from the search page, and from the species summary pages under the section "Tools / Identification keys". Coming from the search page, it is possible to restrict the key selection by large geographic area, Order, or Family, or to enter the key id when this is known. It is possible to specify whether the list should show only keys available with the Lucld interface.

Advantages, disadvantages and limitations. - Advantages are the inclusion of pictures in keys that had none, and the ability to easily step forward and backward during the identification process. Also, the species account is only one click away when reaching a possible identification, so that it can be verified with additional information.

Further developments. - The use of the Lucld interface is a good development of our simple format. More species pictures and character illustrations are needed. One internal improvement is to be able to give at the same time the name as used in the publication and the current accepted name, highlighting when the name in the publication is now considered as a synonym.

2.4 IDENTIFICATION BY MORPHOMETRICS TOOL

Description, how to use it, and where to find it. - This tool uses measurements that are easy to obtain from specimens or pictures, and computed standard ratios. The user needs to take measurements of Total Length (TL), Head Length (HL), Eye Diameter (ED) and Body Depth (BD). The tool accepts measurements made in centimeters or inches. In the case of pictures, measurements in pixels are accepted. Ratios of the head length, eye diameter and body depth to the total length are computed and compared with values stored in FishBase. Providing the FAO Area, from where the specimen was collected, the Class, and the Family (optional) significantly reduces the number of possible species. In addition, the Total Length (TL), by eliminating species that do not grow as big or larger than the unknown specimen, further focuses the search to a few species in many cases. The tool returns a list of possible species. For each species, a short description, fin counts, a picture and a link to the species summary page in FishBase is included. The "Identification by Morphometrics" can be found in the "Tools" section of the search page.

Advantages, disadvantages and limitations. - This tool can be used when the user does really not know what he has in hands. As far as it was possible, commercial species were covered first. However, such measurements are available "only" for a third of species, limited by the number of suitable pictures from which the reference measurements are taken. It must be also understood that the standard ratios are usually computed on one picture only, so there is no statistical range: the possible range during the matching is predefined.

Further developments. - An advanced interface which includes other measurements such as preorbital length, predorsal length and preanal length is under development. It is expected that including these measurements will shorten the list of possible species most often to 10 species or less.

3 DISCUSSION

Missing tools: interactive keys, image analysis, Barcode. - Since the late 1960s, computer-aided identification softwares have been developed, using a matrix of taxa / character states as the basis of the tool (Delta [3], XPER [4], LucId, etc.). In FishBase, we store data on morphological description of species, but unfortunately in a format that is primarily incompatible with these softwares. A first promising attempt was made to transform the data in a correct format to be included in XPER, which opens the way to the suitable compliance with the TDWG SDD standard (Structured Descriptive Data [5]). But we face the usual content issue that is recurrent for large groups: how to describe a seahorse, a tuna, a hagfish, and a turbot with the same characters? So the limitations are now on the structure and the standardization of descriptions more than a technological issue. Image analysis is another tool that we did not implement. After some tries with both public and commercial products, the result did not seem efficient enough in order to be incorporated in FishBase: at the present stage of these tools, it could help to restrict possibilities only as long as our reference picture collection is more complete than currently; also,

these reference pictures must be prepared “manually” to remove noise (e.g., focusing on the individual and not the background) if we want to increase the true positive matching. The other issue is to find solutions to use at the same time live underwater pictures with different orientations of the individual and dead specimens with different colours, not to speak about growth allometries, sexual dimorphism, and intermediary individuals when their sex changes during their life cycle like in some families. A new approach is the identification offered by the Barcode of Life (BoLD website [6]), if the user has access to a respective test kit for the genetic analyses. It is not clear whether this identification tool can be more integrated in other webpages like FishBase (the two websites are already cross-linked), or if it is best to use the tool from the BoLD website.

Identification strategy. - At the moment, identification tools in FishBase are independent, except that the Quick Identification Tool links in the end with the Identification Through Pictures Tool beyond the family level. The idea is that all the tools should be integrated into one, and that the user could choose to jump from one to another anytime during an identification session, or even better, that the system could guide the user in that choice, according to the declared skill level of the user. Interactive keys are obviously the start of such developments. XPER proposes already some modules that can build the identification pathway according to some constraints (e.g., sort out species first that are the most abundant, the easiest to identify, or have the most striking forms): someone has to enter the relevant data. But we are in need of both the design and the technology to jump from outline drawings to identification key and to real pictures, from morphometrics/meristics to image analysis, restrict to a size or a geographical area. The final vision is that the system would guide the user and suggest how to start and which pathway to use across all tools.

4 CONCLUSION

FishBase has deliberately favoured the simplicity over more elaborated identification tools that are costly to develop and maintain, including in terms of data. Some of these simple tools are really easy to deploy on the web such as the Quick Identification Tool with simple outline drawings, and the computerization of the printed dichotomous keys under a simple database format and web layout. Colleagues could move forward quickly to these simple solutions for other taxa. However, each of the tools existing or not in FishBase is interesting in a given context. A long-term goal for the computer-aided identification domain could be to gather all tools in a unique strategy and interface for the user. But this still requires research and technological development, such as the work being done under the European project *KeyToNature* [7]. The last important point is that illustrations, including correctly identified pictures, must be made publicly available. Images are used in 3 of our 4 tools, and we could design the morphometrics tool only because we had a significant number of pictures available. *Homo sapiens* is a species that uses the visual sense to a high degree, and visual identification is still and may remain for a long time its preferred method.

ACKNOWLEDGEMENT

The various tools in FishBase were developed in the last 20 years during various projects mainly funded by the European Commission. Eli Agbayani, Josephine Barile, Elijah Laxamana, Christian Elloran and Stacy Militante were the successive programmers who developed them.

REFERENCES

- [1] R. Froese and D. Pauly (eds.), *FishBase 2000: Concepts, design and data sources*. Los Baños, Philippines, ICLARM, xvii+344 pp., 2000.
- [2] Lucid Central, "Lucid Phoenix", <http://www.lucidcentral.org/Software/LucidPhoenix/tabid/152/Default.aspx>, 2010.
- [3] M. J. Dallwitz, "Overview of the DELTA System", <http://delta-intkey.com/www/overview.htm>, 2009.
- [4] J. Lebbe and R. Vignes Lebbe, "Xper²". <http://lis-upmc.snv.jussieu.fr/lis/?q=ressources/logiciels/xper2>. 2010.
- [5] G. Hagedorn, K. Thiele, R. Morris and P. B. Heidorn, The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.0. <http://www.tdwg.org/standards/116/>. [Last retrieved 05-May-2007], 2005.
- [6] BoLD, "Barcode of Life Data Systems" <http://www.boldsystems.org/views/login.php>, 2010.
- [7] KeyToNature, "KeyToNature: a new e-way to discover biodiversity", <http://www.keytonature.eu/wiki/>, 2010.