



## Note

# An e-key database system for finfish identification: its design, implementation and utility

AJEY KUMAR PATHAK\*, SHRI PRAKASH SINGH AND REETA CHATURVEDI

ICAR-National Bureau of Fish Genetic Resources, Canal Ring Road, Post-Dilkusha, Devikhare, Lucknow - 226 002  
Uttar Pradesh, India

e-mail: pathakajey@gmail.com

## ABSTRACT

This study focused on the development of a computer based e-key system that provides facility for identifying fish species using electronic keys. Electronic keys of fin formula and other 11 distinguishing characters described for Indian finfishes were prepared and the rule based database approach was implemented to develop the system. In addition to identification, the system provides information on taxonomy and distribution along with photographs of the identified finfish species. The system was built by applying Microsoft Access database and Visual Basic technologies for the Windows platform. The system is interactive to the user through windows components. The system facilitates identification and also provide complete information about the fish species and is presently available on compact disk (CD). It was developed assuming the minimum computing (Hardware and Software) resources available to the end users.

Keywords: Database, e-key system, Finfish, Fin formula, Fish identification system, India

Identification of fish species is a primary step towards any fisheries management and conservation programme, however many species of fish are difficult to identify (Lyons *et al.*, 2006). Species identification for fishery purposes has been a major programme for the Food and Agriculture Organisation (FAO) since the 1960s (Guisande *et al.*, 2010). Correct identification of species is the starting point for any type of biological study, particularly on wild populations (Rainboth, 1996). In finfish taxonomy research, information on fins plays significant role in identifying the species and population. Traditionally unknown biological specimens are identified by expert consultation, comparison with specimens of known identity and by the use of keys or similar tools. Out of these, the most common and reliable method for classification of organisms, is use of the taxonomic keys. The most common form of taxonomic/dichotomous key was originated with Jean de Lamarck in 1778 (Woodland, 1991). In this way, identification enables us to retrieve the appropriate facts from the system (classification) to be associated with some specimen at hand and is better described as the recovery side of taxonomy (Blackwelder, 1967).

Morphometric and meristic characters are powerful key characters in identifying the specimens as these are powerful tools for measuring discreteness and relationships among fish species (Guisande *et al.*, 2010). For this reason, analysis of morphometric and meristic characters has been widely used by ichthyologists to

differentiate species and different populations within a species (Rohlf, 1990; Rohlf and Marcus, 1993; Foote, 1997; Cadrin, 2000; Wainwright, 2007). Identification of fish species with accuracy, by comparing the specimen with fin illustrations, written descriptions and range of information is quite tedious and time consuming process for those who are not specialists. In order to reduce burden of routine fish specimen identification task, automated species identification using digital technologies is a better option. New digital technologies provide opportunities to address the challenging task of fish species identification (Lyons *et al.*, 2006). Recent advances in the computer technology allow for the development of exciting new ways to organise and present the information necessary to identify species (Stevenson *et al.*, 2003). Nowadays, a lot of work is being done using computers to achieve accuracy and minimise the processing time.

In the present communication, structure, function and behaviour of an electronic key (e-key) system has been described, which aids in identifying finfish species reported from Indian waters. Information on taxonomy and distribution are also provided with illustrations.

In order to develop e-keys for finfishes reported from the different waters of India, authors used following sources: Talwar and Jhingran (1991); Jayaram (2010) and Vishwanath *et al.* (2007).

Table 1 lists the parameters selected for identification and other general information included in data collection.

A database system approach was introduced to store and organise data of parameters and to analyse the organised data for fish identification.

The identification system was developed using Microsoft Access (MS-ACCESS) 2000 and Visual Basic under Windows Operating System. MS-ACCESS 2000 was used for developing the database and Visual Basic for designing and developing application programming interface (API) using the windows component and to connect the database. The key considerations in the development were use of the commonly used tools and interfaces for data query, entry, summary and report in order to minimise the cost where resources are scarce. The system requires the following minimum configuration to work:

Microsoft Windows 2000/XP/2003; Microsoft ACCESS; at least 128 MB RAM; a 1024x768 high resolution monitor.

An entity relationship model was formed to design the structure of the database. The primary and foreign keys were identified in the tables and relationships between tables was created. Fig. 1 shows the sample screen shot of the tables created in the database. The data on fins and 11 other distinguishing characters for (Table 1) finfish species belonging to fresh, brackish and marine waters were collected and e-keys were prepared and stored in the database through the data entry form module developed in MS-ACCESS 2000 database environment. Table 2 presents the distribution of fish species in the different ecosystems that are included in the database.

Visual Basic 6.0 was used for the design and development of the application interface. Active Window Interface (AWI) components were used for developing the interface. These components were used to provide the interaction and control over the parameter for querying with the database. Open Database Connection (ODBC) was applied to integrate the database for querying and retrieving the information from the database.

The e-key system can be operated by installing the application from the CD under Windows operating environment. The installation loads the application and also copies the database by creating a folder and assigning the application name in the Program folder under Windows. Fig. 2 depicts screen shot for identifying the finfish specimen that appears after the application starts. This screen shot is in two parts. The left part is the query

Table 1. Parameters selected for identification and other general information`

Identification parameters
Fins
Dorsal fin
Anal fin
Pectoral fin
Ventral fin
Adipose fin
Caudal fin
Scales
Predorsal scale
Lateral line scale
Scale below the lateral line
Scale above the lateral line
Lateral line scale transverse
No. of scales in head region
Gill raker
No. of barbels
No. of scutes
Caudal peduncle
Other parameters
Classification
Class
Subclass
Division
Subdivision
Superorder
Order
Suborder
Family
Subfamily
Genus
Distribution
Fin formula
Fish image

window and the right part displays the result of a query. In the left part, action command buttons are provided below to execute the operation of interest.

The identification can be performed through a query function module embedded in the API. Here, the user has to key in the values for the parameters used in identification. After entering values, the user has to click on 'OK' command button to see the result of query in the right part of the interface serving as an output. For example, if the user supplies values as indicated in Fig. 2 and press the OK command button, a message

Table 2. Habitat-wise distribution of fishes in the database

No. of fish species in the database	Cold water	Cold warm water	Warm water	Warm brackish marine	Brackish water	Warm brackish	Brackish and marine	Marine
835	91	26	303	61	14	53	39	248

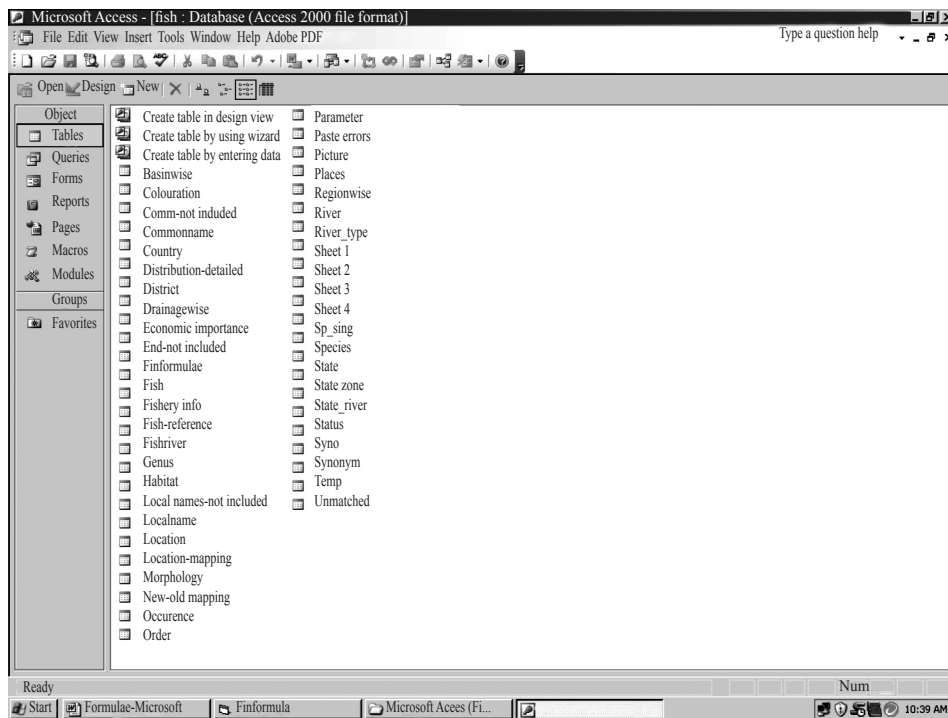


Fig. 1. Sample screen shot of tables in the database

window appears on the screen that displays the message ‘fish found’ (Fig. 3) and the identified fish along with picture and other information is displayed in the right part of the query window (Fig. 4).

Computers are effective means for identification when used interactively. Many computer based identification systems exists for identifying fish species using either of the approaches of a data matrix, rule based expert systems



Fig. 2. Sample screen shot of the interface for identification of finfish specimen

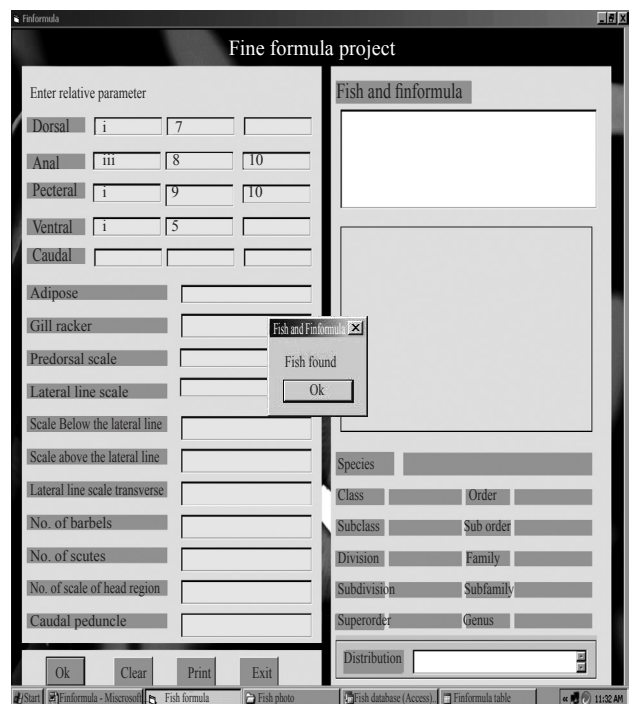


Fig. 3. Sample screen shot of the message window after keying in values of fins

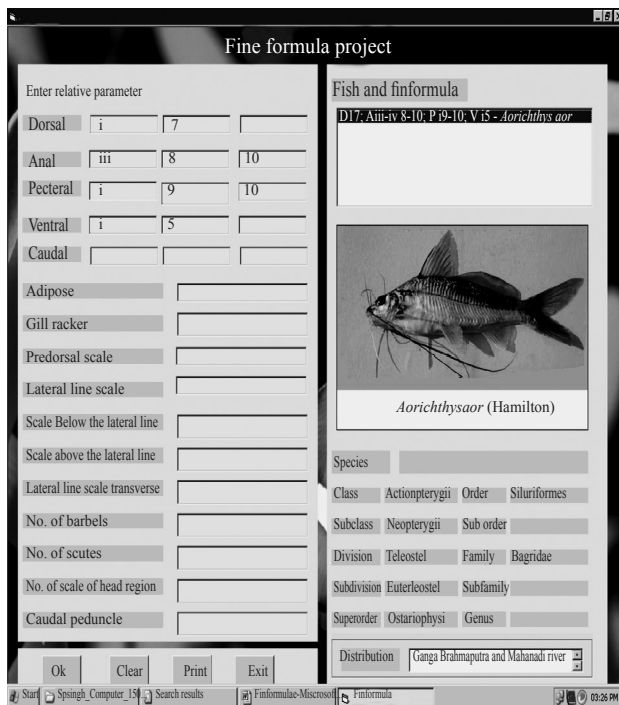


Fig. 4. Sample screen shot of the result with fish photograph after querying with the database

and frame based expert systems. Each approach has limitations and advantage over others. These approaches further employ the different classification techniques and methods to classify the biological organism. Our system akin to rule based, using the database approach is somewhat slow as the most taxonomic information are published in a form more akin to data matrices than rules. The rule based systems are difficult to construct and the information they contain tends to be sparse (Dallwitz, 1992). Earlier to this, in India computers were applied to understand the systematics of marine biota of India and computer aided taxonomy was developed based on the numerical codes (Chavan *et al.*, 2004). This system was quite slow in execution and could not gain much importance. Later, an effort was made to develop an electronic catalogue describing the systematics of Indian biota and further role of electronic catalogues in resolving the discrepancies of known organisms was discussed (Chavan *et al.*, 2005). Recent advancements in the computer and database technology have provided opportunities to address the challenging task of identification and resolving taxonomic ambiguities at the molecular level also. DNA sequencing of the cytochrome c oxidase subunit I gene (COI) from fish tissue samples for obtaining unique, species-specific “barcodes” has become the promising molecular taxonomy tool for accurate fish species identification. In this way, barcode databases developed using computers and database technology (Nagpure *et al.*, 2012) play a significant role in identifying fish species.

But, this method can only complement the existing approaches of conventional taxonomy, because DNA barcodes based on a few specific genes may fail to distinguish closely related species because of the persistence of ancestral polymorphism (Mallet and Willmott, 2003). Thus, it is imperative to record that there is no replacement of conventional taxonomy and computational approaches using the conventional taxonomy to identify fish species and resolving taxonomic ambiguities are considered better. However, this rule based system is more interactive to users through windows component and it can be very well utilised in the fields for identification of finfish specimens. IPEZ, an expert system for taxonomic identification of fishes based on the machine learning technique has already been developed using the morphometric characters in more than 8900 individuals belonging to 6 classes, 43 orders, 192 families, 510 genera and 847 marine and freshwater species, which aids in taxonomic identification of juveniles and adult fishes using 32 measurements (Guisandae *et al.*, 2010). Since our system applies rule over the e-keys stored in the database, it resembles an example of expert system that has been built for the users including students, especially in the developing countries having only minimum computing resources.

The e-key system based on database system approach is a more versatile and scalable tool. At present, the system covers only finfish species and hence only finfishes can be identified. The system has no geographic limit and can be used to identify the specimens of other regions subject to availability of identification records of finfishes of that region. The system is very fast in identification and quickly retrieves the information. The system has a few thousand records of identification and needs upgrading in the near future, if number of finfish species increases voluminously. Based on the experience and working of the system, it is suggested that the present system can be utilised for identifying the fishes of other countries, if the information on fins and other distinguishing morpho-meristic characters are supplemented in the database.

### Acknowledgements

The authors are grateful to the Director, ICAR-National Bureau of Fish Genetic Resources, Lucknow for providing facilities and encouragement. They are also thankful to Late (Dr.) D. Kapoor and Dr. Rajesh Dayal for the help rendered during the work.

### References

- Blackwelder, R. E. 1967. *Taxonomy, a text and reference book*. John Wiley & Sons, Inc., New York, USA.
- Cadrin, S. X. 2000. Advances in morphometric identification of fishery stocks. *Rev. Fish Biol. Fish.*, 10: 91-112.

- Chavan Vishwas, Rrane Nilesh and Watve Aparna 2005. Resolving taxonomic discrepancies: Role of electronic catalogues of known organisms. *Biodivers. Inform.*, 2: 70-78.
- Chavan, V., Watve, A. V., Londhe, M. S., Rane, N. S., Pandit, A. T. and Krishnan, S. 2004. Cataloguing Indian biota: the electronic catalogue of known Indian fauna. *Curr. Sci.*, 87(6): 749-763.
- Dallwitz, M. J. 1992. A comparison of matrix-based taxonomic identification systems with rule-based systems. In: Xiong, F. L. (Ed.), *Proceedings of IFAC Workshop on Expert Systems in Agriculture*, International Academic Publishers, Beijing, p. 215-218.
- Foote, M. 1997. The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.*, 28: 129-152.
- Guisande, C., Manjarres-Hernandez, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuna, A., Prieto-Piraquive, E., Janeiro, E., Matias, J. M., Patti, C. and Patti, B. 2010. IPEz: an expert system for the taxonomic identification of fishes based on machine learning techniques. *Fish. Res.*, 102(3): 240-247.
- Jayaram, K. C. 2010. *The freshwater fishes of the Indian region*, Narendra Publishing House, New Delhi, India, 616 pp.
- Lyons, J., Hanson, P. and White, E. 2006. A photo-based computer system for identifying Wisconsin fishes. *Fisheries*, 31(6): 269-275.
- Mallet, J. and Willmott, K. 2003. Taxonomy: renaissance or Tower of Babel? *Trends Ecol. Evol.*, 18: 57-59.
- Nagpure, N. S., Rashid, I., Pathak, A. K., Singh, M., Singh, S. P. and Sarkar, U. K. 2012. FBIS: A regional DNA barcode archival and analysis system for Indian fishes. *Bioinformatics*, 8(10): 483-488.
- Rainboth, W. J. 1996. *Fishes of the Cambodian Mekong. FAO species identification field guide for fishery purposes*. FAO, Rome, 265 pp.
- Rohlf, F. J. 1990. Morphometrics. *Annu. Rev. Ecol. Syst.*, 21: 299-316.
- Rohlf, F. J. and Marcus, L. F. 1993. A revolution in morphometrics. *Trends Ecol. Evol.*, 8: 129-132.
- Stevenson, R. D., Haber, W. A. and Morris, R. A. 2003. Electronic field guides and use communities in the eco-informatics revolution. *Conservation Ecol.*, 7(1): 3
- Talwar, P. K. and Jhingran, A. G. 1991. *Inland fishes of India and adjacent countries, vol. I & II*, Oxford & IBH Publication Co. Pvt. Ltd., New Delhi, 1062 pp.
- Vishwanath, W., Lakra, W. S. and Sarkar, U. K. 2007. *Fishes of North-east India*, National Bureau of Fish Genetic Resources, Lucknow, India, 264 pp.
- Wainwright, P. C. 2007. Functional vs morphological diversity in macroevolution. *Annu. Rev. Ecol. Syst.*, 38: 381-401.
- Woodland, D. W. 1991. *Contemporary plant systematics*. Englewood Cliffs, Prentice Hall, NJ, USA.