

Short Communications

Computer aided taxonomy (CAT): Approach for understanding systematics of marine biota

Vishwas Chavan, D P Kavlekar & D Chandramohan
National Institute of Oceanography, Dona Paula, Goa 403 004, India

Computer aided taxonomy (CAT) is gaining importance day by day. In order to familiarise the taxonomy and its continuing usage an attempt has been made to develop a module for knowing the systematics of organisms and to generate automatic numerical code for marine biota. It will help not only the taxonomists and systematists but also the research, academicians and students.

Earth is the planet in the universe which is known to support life. Approximately over 50 million species are projected to exist on the earth. Out of these only 1.5 million species are known to us so far¹. It is just not possible to identify and address each of these 1.5 million species without a definite mechanism. For nature conservation policy, a thorough understanding and identification of species is imperative². This is specially true and essential in case of marine biota which constitutes almost 70% of the known flora and fauna.

In the present political climate of financial cutbacks, the branch of taxonomy is under pressure. Rapidly decreasing number of skilled taxonomists lead to genuine concern in the scientific community regarding how to preserve this knowledge. Since the traditional methods of knowing the systematics of a given species is time-consuming, many are reluctant to grasp this knowledge. This is also due to the fact that there is the problem of getting (fast) access to the documented taxonomy. Thus, there is a strong need to solve the problem of vanishing taxonomic expertise and to find a solution to make widely accessible what is presently known. This could be achieved through appropriate documentation of identification characters, development of 'easy-to-grasp' methodology for identification of organisms of one's interest. Realising this an attempt has been made to develop a computer based module for knowing the systematics of marine organisms and generate automatic numerical code for the marine flora and fauna.

It was only in the early 80's that the idea of computer based taxonomic databanks was initiated by UNESCO. Since then numerous attempts have been made to computerize taxonomy. But none could become popular due to one or the other reasons. It is to be noted here that all these attempts were aimed at identifying the organisms based on their taxonomic keys. Either Expert System³, Relational Database⁴ or the Morphometric Values⁵ were used for it. Expert-cum-multimedia Database system has also been employed for this purpose². Unfortunately no serious attempt has so far been made towards knowing the systematics of a species of interest through computerization. The present exercise is the first step of its kind to achieve this goal which would be of help to researchers and academicians in teaching and to students of taxonomy. A package called "CAT ver. 1.0" has been developed utilising the widely used DBMS (Database Management System), dbase III+ for knowing and understanding systematics of marine biota.

NIOCODE—To facilitate the development of the package, a numerical code (NIOCODE) has been developed to identify and address each species and facilitate information retrieval and easy comparison between species. NIOCODE, which is based on NODC's (National Oceanographic Data Center, USA) taxonomic code with few modifications; manifests a hierarchical system of numerical code of 12 digits representing the organisms to the level of species, subspecies or variety (Table 1). Code links the Linnaeus system

of binomial nomenclature to a numerical scheme that facilitates modern methods of data storage and retrieval⁶.

CAT—Designed utilizing Top-Down design and information hiding approach, this package is made up of 10 programmes and 4 procedure files. This package contains 18 minor programmes and three dbf files. Code and organism's details up to the family level have been stored in die1.dbf whereas the code and species description from genus to variety level and its authority in die2.dbf. Synonyms along with authority and corresponding species code have been stored in synonym.dbf. Table 2 lists detailed structure of these dbf files.

As shown in Fig. 1, the package provides four different facilities. One of these is the NIOCODE search for the species of interest (Fig. 2). For those marine organisms for which NIOCODE has not been assigned, it can be generated (Fig. 3). However, this facility is not incorporated in the version supplied to the users and is available only for the authors. This is to avoid duplication of use of same code for more than two species. Instead in the package supplied, request has been made to the users to inform the developer about non-availability of NIOCODE for the species of interest, so that it could be included.

The package also provides the facility of knowing the systematics of species of interest either by its scientific name or by its corresponding NIOCODE (Fig. 4). As shown in Fig. 5, facility has also been provided to know the sublevel of any particular level in the taxonomical hierarchy. Developed and compiled in dbase III+ package runs on any IBM compatible PC.

It has been proposed to supply the package free of cost to educational and research institutes as

Table 1—Structure of numerical taxonomic code (NIOCODE)

02 Digits	Subkingdom, Phylum, Sub-phylum, Class, Superorder, Order
04 Digits	Superclass, Class, Subclass, Superorder, Order, Suborder, Infraorder, Superfamily
06 Digits	Class, Order, Suborder, Infraorder, Superfamily, Family, Subfamily, Infrafamily
08 Digits	Genus
10 Digits	Species
12 Digits	Subspecies, Variety

Table 2—Structure of the database (dbf) files in CAT Ver. 1.0

Field	Field Name	Type	Width
Dic1.dbf	1	Code	Character
	2	Discription	Character
Dic2.dbf	1	Code	Character
	2	Discription	Character
	3	Authority	Character
Synonym.dbf	1	Code	Character
	2	Syn1	Character
	3	Ath1	Character
	4	Syn2	Character
	5	Ath2	Character

well as major taxonomic reference centers. This would help in building the taxonomic code for presently uncovered species. Presently it provides the taxonomic code for 3000 floral and faunal species mainly from marine and estuarine environment. CAT ver. 1.0 will help in providing finger tip systematics of the species of interest. The

CAT Ver 1.0

NIO DISC

MAIN MENU

SEARCH NIO CODE
GENERATE NIO CODE
SYSTEMATICS BY NIO CODE
SYSTEMATICS OF ORGANISM
TAXONOMIC SUBLEVELS
EXIT

Highlight option By using 'UP' or 'DOWN' arrow key and press 'ENTER'
or press appropriate menu number

Fig. 1—Main menu exhibiting the facilities provided by CAT ver. 1.0

package and its frequent updating will help in taxonomy education. It will list the total marine flora and fauna.

Computerizing marine taxonomy: Potentials

Advances in computer science will only be able to overcome the diminishing state of systematics

leading to Computer Aided Taxonomical Identification System (CATIS). With the recent introduction of multimedia technologies, this option becomes more relevant in the present context than ever before. Various tools such as expert system, relational database etc. have been employed in developing computer assisted

CAT Ver 1.0 NIO DISC

NIO CODE SEARCH FACILITY	
LEVEL	SPECIEA
LEVEL NAME	UDOTEA CONGLUTINATA
NIO CODE	0809030101

Press J To continue

Fig. 2—NIOCODE search facility

CAT Ver 1.0 NIO DISC

NIO CODE GENERATION FACILITY	
DATA ENTRY SCREEN	CLASSIFICATION
CLASSIFICATION LEVEL	CHLOROPHYTA (P)
SPECIES	CHLOROPHYCEAE (C)
SPECIES	CAULERPALES (O)
SPECIES	CADIACEAE (F)
CONGLUTINATA	UDOTEA (G)

NEW CODE 0809030103

SPECIFIED ORGANISM IS NOT CODED
DO YOU WANT TO GENERATE A NEW CODE? (Y/N) Y

Fig. 3—NIOCODE generation facility

SYSTEMATICS BY NIO CODE		
ENTERED CODE IS = 0808010101		
TAXONOMIC LEVEL	CLASSIFICATION	NIO CODE
PHYLUM	CHLOROPHYTA (P)	08
ORDER	CLADOPHORALES (O)	0808
FAMILY	CLADOPHORACEAE (F)	080801
GENUS	RHIZOCLONIUM (G)	08080101
SPECIES	RHIZOCLONIUM GRANDE	0808010101

press 'ENTER' to continue

Fig. 4—Systematics by NIOCODE facility. (Similar facility is available to know systematics by providing species)

TAXONOMIC SUBLEVEL FACILITY	
TAXONOMY LEVEL	:= CLASS
CLASS	:= CHLOROPHYCEAE (C)
TAXONOMY SUBLEVEL	:= ORDER
CODE	DISCRIPTION
0802	VOLVOCALES (O)
0803	CHLOROCOCCALES (O)
0804	PRASIOALES (O)
0804	SCHIZOGONIALES (O)
0805	ULOTRICHIALES (O)
press 'ENTER' to see more sublevels	

Fig. 5—Taxonomic facility to know systematics by providing species to know the sublevels in a particular level of interest.

identification. However, no single approach is capable of overcoming this problem. Only way to solve this problem, rationally, is to proceed with the integrated approach of all the available tools/techniques of computer science. It has been foreseen that the combined approach of "expert-cum-relational database" will be the best solution along with fool-proof documentation of available systematics knowledge. However, success of all these attempts depends on the involvement of skilled taxonomists and systematists in development of such package.

Though the package is slow in its execution speed, it is a unique tool towards understanding and learning systematics of marine biota. More work on the package needs to be carried out to improve its execution speed. Efforts are needed also to correlate the print taxonomic keys along with the NIOCODE to form the Computer Aided Taxonomic Identification System (CATIS) for marine flora and fauna. Further more integrated

approach of "Expert-multimedia-cum-relational database" will be the best solution on this complex issue.

Authors are thankful to the Director, for encouragement. Thanks are due to Mr. V.K. Kulkarni for his help in data collection. They are thankful to department of Biotechnology, Government of India, New Delhi for sponsoring the Biotechnology Information System (BTIS) Network facility at NIO, Goa under which this work was carried out.

References

- 1 Swaminathan M S, *Biol Edu*, 8 (1991) 4.
- 2 Schalk P H, *Systematics goes multimedia biological knowledge*, (ETI, University of Amsterdam, Amsterdam) 1992.
- 3 Froose R, *Fishbyte*, 7(1989) 25.
- 4 Froose R, Actienbach I & Papasissi I, *Fishbyte*, 8 (1990) 25.
- 5 Froose R, *Fishbyte*, 7 (1989) 25.
- 6 Chavan V S, Chandramohan D & Parulekar A H, *Biol. Edu.*, 8 (1993) 302.