Indian Journal of Marine Sciences
Vol. 34(1), March 2005, pp. 128-135

MASDEA—Marine species database for Eastern Africa

Edward Vanden Berghe
Flanders Marine Data and Information Centre, Flanders Marine Institute
Vismijn, pakhuizen 45-52, 8400 Oostende, Belgium
[E-mail: wardvdb@vliz.be]

Received 1 April 2004; revised 24 December 2004

The Marine Species Database for Eastern Africa (MASDEA) was started in 1996, in the framework of a collaborative project between the Regional Cooperation in Scientific Information Exchange – Western Indian Ocean (RECOSCIX-WIO) and the Kenya Marine and Fisheries Research Institute (KMFRI). Since then, it has grown to a substantial database, with over 20,000 taxonomic and 50,000 distribution records. Since 1999, the database has been online, through the web server of the Flanders Marine Institute (VLIZ), and has attracted a reasonable amount of hits, some resulting in comments, and corrections and additions to the content. Some of the decisions taken at the design stage of the database are discussed. A preliminary analysis of the data in the database is presented. This clearly shows a marked 'observer' bias, and demonstrates the need to continue the efforts of collecting information.

[**Keywords**: Biodiversity; information systems; western Indian Ocean]

[**IPC Code:** Int. Cl.^{7} G11C 13/00]

Introduction

Since the 'Rio Conference' sustainable management of natural resources and conservation of biodiversity is high on the agenda of many governments. Unfortunately, the information to make informed decisions on these matters is not always at hand, and compiling it is no easy task. In order to compile information relevant to biodiversity, a species list is an indispensable tool: taxonomy, and the species names, is the vocabulary of biodiversity. In spite of the crucial role of a species list in any biodiversity study, no such list existed for the eastern African region. These were the reasons why, in 1996, the MASDEA database – Marine Species Database for Eastern Africa – was developed at the RECOSCIX-WIO project (Regional Cooperation in Scientific Information Exchange – Western Indian Ocean), based at the Kenya Marine and Fisheries Research Institute in Mombasa, Kenya. MASDEA consists of taxonomic records, and distribution records, mostly at the country level, for these taxa. Since 1996, the database has been permanently maintained and updated; today it lists just under 20,000 taxa on the genus or species level, more than 1400 records for families, and contains 50,000 distribution records, mostly extracted from primary literature, some extracted from other databases. One of the goals of this exercise is to provide biologists with a reference tool; while the database is being developed, and its contents validated, it can serve as a roadmap to the literature available on biogeography of the Western Indian Ocean. In the long run, as soon as the database is nearing completion and validation, it will be an invaluable tool in environmental management and identification of biodiversity hotspots.

For creating records published literature was the main source of information; where alternative sources were available, preference was given to primary literature. However, for two major taxa, fish and macroalgae, exceptions were made: records, both for taxonomy and for distributions, were uploaded from FishBase[1,2] and for the checklist of macroalgae of the Indo-Pacific Ocean[3]. In part, this project was inspired by the fact that some data had been collected during a project funded by UNESCO/ROSTA, for which no follow-up funding was provided. The MASDEA database was, as a project, formally adopted by the region during the Fourth Session of the IOCINCWIO.

Scope of the Database

The main objectives of the database were to provide biodiversity researchers in the region with a standard, well-documented and properly researched list of taxonomic names. Also taxonomic names no longer considered valid are included: names relegated to synonymy are entered in the database, and a pointer leads to their valid/accepted name. By storing information on synonymy in the database, no knowledge is required for entering records in the database, which makes fast progress in data entry possible – the technical process of entering data is decoupled from the specialist task of interpreting the taxonomic nomenclature. With the extensive list of references, which forms integral part of the database, this taxonomic checklist forms a guide to the biodiversity literature of the wider eastern African region. Based on the database, country-wise list of biodiversity can be created. Another possible application is gap analysis, both in terms of areas/taxa inadequately researched, as in terms of predicting species presence from (apparently) discontinuous distributions.

The database covers all countries that were involved in the RECOSCIX-WIO project, which are: Eritrea, Kenya, Tanzania, Mozambique, Seychelles, Madagascar, Mauritius and Réunion (France). Very little information on individual countries around the Red Sea was found, hence a decision was made early on in the logging process to lump all records, including the ones from Eritrea. South Africa was included later, and only up to Cape Town; including the Atlantic fauna of east of Cape Town would detract from the original scope of the database. Also records for Somalia, Djibouti, Chagos Archipelago and the Comoros were included. Most distribution records were entered in the database at a geographical resolution of country. Also deep-sea records were entered in the database – assigned to the EEZ of the relevant country were possible, or to the 'Western Indian Ocean' where this was not the case. Unfortunately, most data that has been collected is from shallow waters, and open ocean information remains a major hiatus. This, of course, reflects the bias in studies carried out, and hence literature information available.

Database Development

The software for the database was written in Access. A short summary of the structure is presented below. A number of guiding principles were observed while developing the structure and the user interface.

Very often, while scanning the literature, one comes across nomenclature changes, or a statement that an earlier distribution observation was based on a wrong identification. Very often, such wrong species distribution records were propagated through literature, with the result that many species lists are inflated, containing distribution records based on these misidentifications. Also, very often a taxon is included several times: once under the currently valid name, and under one or more invalid synonyms; hence the need arose to store information on synonymy, and to keep the distribution record linked with the name under which it was originally published. All sources where the information is taken from are documented in the database, thus creating an audit trail. Both taxonomy records and distribution records have a field to include a literature reference, so that anyone who wishes can check the contents of the database against the primary sources.

Taxonomic nomenclature is, in itself, a coding system with very elaborate rules and regulations, and well-understood by trained biologists. These rules are internationally accepted, and documented in the 'Code for Zoological nomenclature' and the 'Code for Botanical nomenclature'. There is no need to develop an extra set of codes, which only carries information that is already contained in the taxonomic names. Chances are that these latter codes will be specific to a given databasing exercise, and create unnecessary confusion. There are, obviously, codes that link records in the relational structure of the database, but these are internal to the system and could be completely hidden from the user.

The database structure was kept as simple as possible. Only those fields have been created that were essential to capture the biogeographical information. For all additional information, memo fields are provided. This way, the user is only presented with, and forced to fill, those fields he/she is really interested in. There is also the added advantage of ease of development, and the performance gain of the completed system. The database consists of a synonymised list, with distribution records referring to the taxon name under which the distribution information was originally published. A record in the taxon table corresponds with a single name – whether valid or not. Records for synonyms include a reference to the record with the valid name. This structure makes it possible to store several synonymous names for each valid name, but only one valid name for each synonym.

Each taxonomic name and each distribution record (in principle presence in one of the countries of the region) is referenced to the literature. Also information on synonymy is referenced to the literature. It is, at present, not possible to enter more than one literature reference per taxon. Multiple distribution records for a single combination of taxon and country, however, are possible. It is also possible to keep track of records of which it is known that they are incorrect or doubtful. To this end, two flags are included: one to indicate whether a record is thought to be valid or invalid, another one to show whether knowledge of the validity is known with certainty or not. The combination of the two flags gives a total of four levels: certainly valid, probably valid, probably false and certainly false. These four combinations allowed capturing most of the reservations expressed by authors of biogeographical papers. Two text fields are provided with the taxonomic records (one to enter information on the logging process, one to enter any information on the taxon, such as a description or a general statement of the distribution), one for the distribution records. In all three of these fields, a system of coding can be used to create cross-links between records. These codes are translated in the appropriate hyperlinks by the database engine connecting the database to the web site.

As illustrated in the diagram (Fig. 1), four tables form the core of the structure: 'Countries', 'Records', 'Literature', and 'Taxonomy'. Records table links country and taxonomy, plus a reference to the literature, making it possible to keep track of sources of information on the level of occurrence of taxon in a country. A separate table gives information on the higher taxonomy; bookkeeping of the flow of information into the system is possible through the table with data entry session (log-session) information.

Since 1999, the database is available online, through the VLIZ web server (http://www.vliz.be/vmdcdata/masdea). The CGI-Bin application to connect the database with the web server is developed in Visual Basic; a PHP version is planned in the near future. All information in the database can be searched interactively. Country lists can be produced; the user can determine exactly

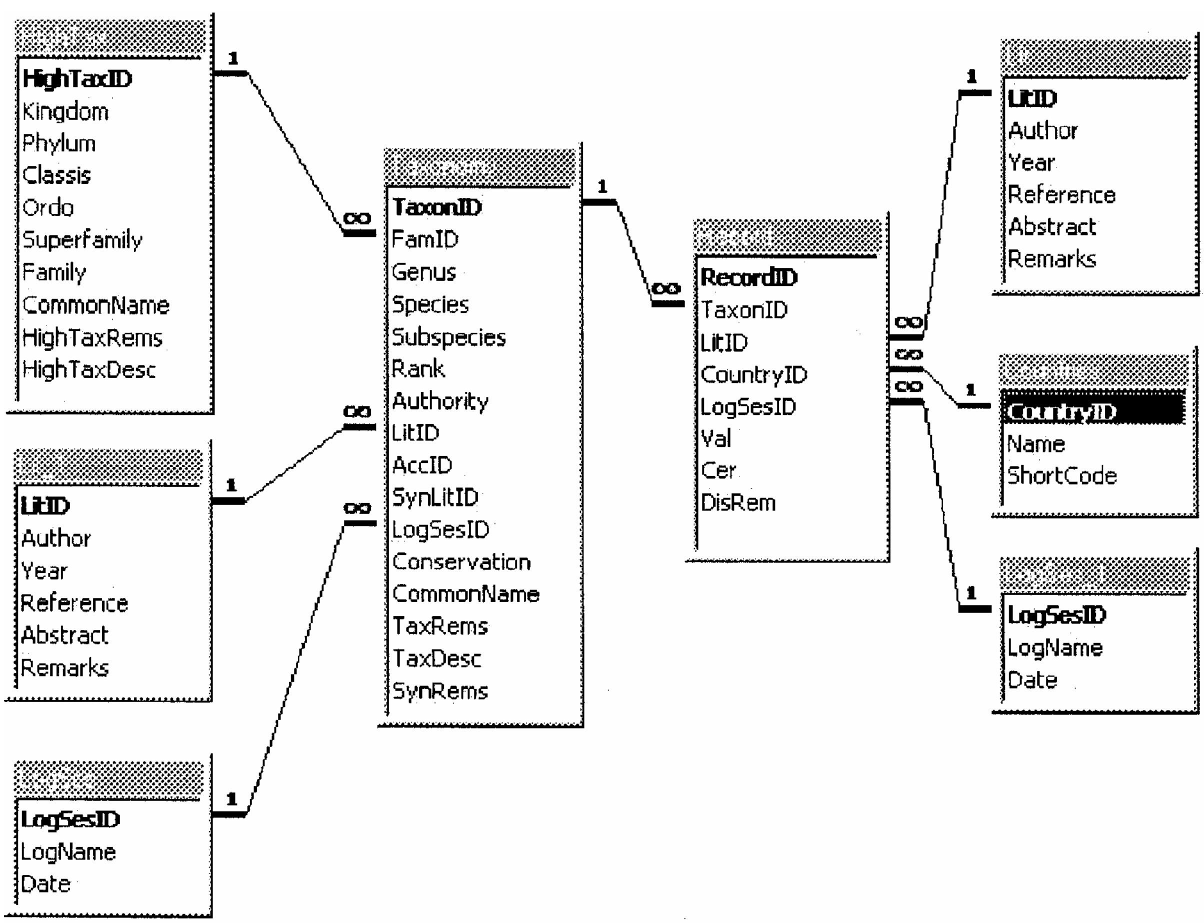

Fig. 1—Structure of the MASDEA database

which names (only valid names or also synonyms) and which records (only certainly valid or also doubtful…) are to be included. The system of coding used in the notes fields in both distribution and taxon records allows creating hyperlinks to another record in the database.

Discussion

About 1,000 published sources have been consulted at the time of writing, resulting in approximately 20,000 taxonomic records and 50,000 distribution records. Information for some taxonomic groups has been extracted from other databases. Fish taxonomy and distribution has been completed using FishBase[1,2]. Seaweeds, apart from being adequately covered from the publications of Isaac[4] and Coppejans and his team[5-9] have been augmented using Silva's catalogue of Indo-Pacific algae[3]. A systematic attempt at tracing all echinoderm literature has resulted in a (what we hope) fairly complete picture for this phylum. Coverage for other phyla remains incomplete or even fragmentary.

Table 1—Number of taxon records (C) per source and per year, for the 30 most important sources

Author	Year	C
FishBase[1]	1996	1988
Silva et al.[3]	1996	1949
Fishbase[2]	2000	1308
Vine[10]	1986	1186
Dautzenberg[11]	1929	878
Drivas & Jay[12]	1988	696
MacNae & Kalk[13]	1958	416
Day[14]	1967	346
Veron[15]	1986	313
Rowe & Gates[16]	1995	289
Vaught[17]	1989	275
Grice & Hulsemann[18]	1967	273
Clark & Rowe[19]	1971	252
Smith & Smith[20]	1963	233
Clark & Courtman-Stock[21]	1976	217
Barnard[22]	1950	215
Sheppard[23]	1987	214
Barnard[24]	1955	204
Day[25]	1967	204
Serene[26]	1984	203
Richmond[27]	1997	177
NODC[28]	1997	176
Spry[29]	1961	173
Guinot[30]	1967	167
Brown, Urban & Newman[31]	1982	167
Spry[32]	1964	163
Randall[33]	1992	154
Griffiths[34]	1974	135
Fischer & Bianchi[35]	1984	135

Table 1 lists the most important sources of taxon records. Apart from the classification of molluscs by Vaught[17], these are also the most important sources for distribution records. Figure 2 shows a scatter plot of the number of taxa recorded in an area, against the number of literature sources from which these distribution records were extracted. The very strong positive correlation (r=0.94) suggests that we did not reach 'saturation' yet in sampling the literature – adding new literature sources will still dramatically increase the number of species recorded from each area. It also shows that, for the time being, it is not really worth doing an analysis of number of species occurring in an area and trying to identify biodiversity hotspots. The preliminary species count per area is given in Table 2.

In Table 3, the number of species in some major groups are compared between MASDEA, the numbers listed by Richmond[36] and the European Register of Marine Species[37]. It is immediately clear that some groups are over-represented in both MASDEA and the publication by Richmond[36], others are under-represented. Being based on a field guide, it comes at no surprise that only the larger, more conspicuous species have been listed here in the list from Richmond[36]. In MASDEA, mainly the

Table 2—Number of species recorded in the database per country. Synonym: records under a name now relegated to synonymy; Valid: records under a name regarded as valid

Name	Synonym	Valid
Aldabra	137	1339
Cargados Carajos	12	175
Chagos	60	1668
Comores	20	995
Djibouti	6	600
East Africa	82	498
Eastern Africa & Madagascar	29	256
Eritrea	9	551
Indian Ocean	3	57
Kenya	100	4232
Madagascar	589	5374
Mascarene Islands	40	950
Mauritius	195	3134
Mozambique	491	6492
Red Sea	349	5666
Reunion	59	1758
Rodriguez	39	317
Seychelles	449	3892
Somalia	27	2101
South Africa	94	4522
Tanzania	187	3612
Western Indian Ocean		187
Western Islands	9	161

overrepresentation of fish, macroalgae and echinoderms is apparent. Fish and macroalgae can be explained by the existence of the very solid foundation provided by FishBase[1, 2] and the work of Silva[3]. Echinoderms were a special research interest of the author, and received more attention than other groups.

One of the great strengths of a database like MASDEA is that it provides a mechanism to integrate results from different field studies. The usefulness of such an activity is clearly demonstrated by the number of holothurians found in each of the four major surveys that were done for this group in Kenya (Table 4).

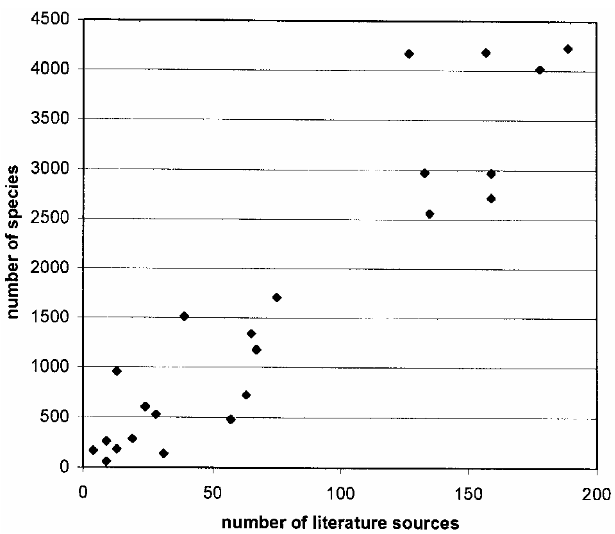

Fig. 2—Number of species recorded vs number of literature sources consulted, for each of the countries included in the database

Since the echinoderms are reasonably complete, certainly for the eastern African region, a number of analyses can be carried out specifically on this group. One form of gap analysis is to predict species presence on the basis of apparently discontinuous distribution patterns; if a species is recorded both to the south and to the north of an area, then, given the presence of suitable habitat, the probability is high that it occurs also in the area under investigation. Figure 3 shows the number of species present in/shared by Tanzania, Kenya and Somalia; three species have been reported from both Somalia and Tanzania, but not from Kenya: *Holothuria parva, Holothuria strigosa* and *Ohshimella ehrenbergii.* Armed with this information, a directed and more efficient search could be made for these 'missing species'.

Figure 4 depicts the number of taxonomic records, as they were entered in the course of time. It is clear that MASDEA as an active project is running a serious danger of stalling. We are now actively looking for partners in this project, who would want to take the project further. We are also looking for taxonomists with knowledge of the flora and fauna of

Table 3—Number of species listed in MASDEA (valid species only), Richmond[31] and the European Register of Marine Species[37]

Kingdom	Phylum	MASDEA	Richmond	ERMS
Animalia	Annelida	829	300	2074
Animalia	Arthropoda	3476	1145	7539
Animalia	Bryozoa	120	500	724
Animalia	Chaetognatha	4	50	42
Animalia	Chordata: Tunicata	146	100	393
Animalia	Chordata: Vertebrata: Pisces	3707	2000	1349
Animalia	Chordata: Vertebrata: Tetrapoda	402		129
Animalia	Cnidaria	1196	695	1329
Animalia	Ctenophora	1		38
Animalia	Echinodermata	1368	419	648
Animalia	Echiura	5	22	19
Animalia	Foraminifera	15		1167
Animalia	Hemichordata	7	20	17
Animalia	Mollusca	3564	3646	3353
Animalia	Nematoda	32		1837
Animalia	Nemertea	5	59	478
Animalia	Plathelminthes	42	100	592
Animalia	Porifera	99	200	1640
Animalia	Sipuncula	24	50	44
Plantae		22	22	5
Protoctista	Macroalgae	2063	1011	1702

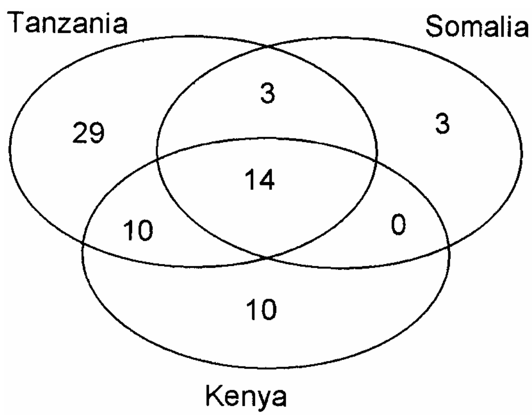

Fig. 3—Species of holothurians unique to, and shared between Tanzania, Somalia and Kenya

Table 4—Number of holothurians recorded in each of the four major surveys for this group in Kenya

Source	Number of holothurians
Levin[38]	7 species
Humphreys[39]	22 species
Massin[40]	11 species
Samyn & Vanden Berghe[41]	26 species
MASDEA	34 species

the region, to check the quality of data entry and to revise the synonymy.

Both the structure and the scope of MASDEA could be revised. The database is, as explained above, rather simple in its structure. It is clear that it would be an advantage to store more than one reference per taxon; or to provide links with full text and illustrations; or to store real georeferenced records… Also, the taxonomic structure itself, now basically a flat table with each field corresponding to a rank in the taxonomic hierarchy could be upgraded to the hierarchical structure as it is used in e.g. Aphia, VLIZ's species register for the North Sea (http://www.vliz.be/vmdcdata/aphia), or (partially) in the Integrated Taxonomic Information System (ITIS, http://www.itis.usda.gov).

The original geographic scope was defined rather haphazardly, based on partnership in the RECOSCIX-WIO project rather than on any biological (or even geographical) reality. To solve this problem, two solutions could be found to rationalise the geographical scope. The first one would be to split the database in a south-western Indian Ocean part, and a Red Sea part. The second solution would be to include the Persian Gulf and Arabian Sea in the

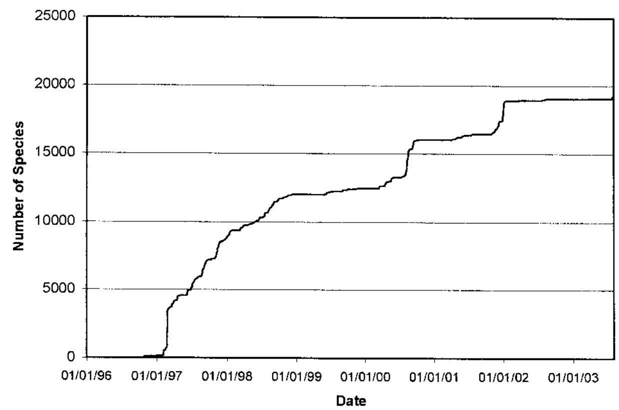

Fig. 4—Number of species in MASDEA, as a function of time

present database, and thus come to a homogenous area of interest. Obviously, both overhauling the engine and revising the scope take time and resources. And the burning question is: who is willing, and has the mandate to do this? Are the nations in the region prepared to use the present database as a tool to consolidate national efforts, and come to a regional project? It is clear that such an activity would be beneficial, and even very much needed. Species distribution information is typically collected in the framework of individual projects, at small scales, and short time spans. But organisms do not stop at national borders; biodiversity conservation and natural resource utilisation have to be planned on a regional scale, not on the level of individual countries. Only by combining and integrating information from individual projects into massive collaborative databases, we will be able to provide the answers on the time scales and geographical scales we need.

Another possible role for a regional biodiversity database is to act as a gateway to global initiatives. MASDEA or a similar database could be used as a tool to catalyse interaction between national/local activities and global initiatives such as the International Taxonomic Information System[42] (ITIS), UNESCO Register of Marine Organisms[43] (URMO) or others. Biologists could take an example of what is happening in physical oceanography: data collected in the framework of individual projects are made available to large data holding facilities, such as the World Data Centres for Oceanography. The combined datasets can then be used in large-scale data and information products, such as the World Ocean Database[44]. Thanks to technological developments, and the creation of collaborative networks such as the Census of Marine Life[45] (CoML) and the Ocean

Biogeographic Information System[46] (OBIS), biologists, not only in the western Indian Ocean but everywhere, are in a position to do the same.

References

1 FishBase, FishBase 96 CD-ROM, (Iclarm, Manila) 1996.

2 Fishbase, FishBase 2000 CD-ROM, (Iclarm, Manila) 2000.

3 Silva C., Basson P & Moe R, Catalogue of the benthic marine algae of the Indian Ocean (U.S.A.) 1996.

4 Isaac W E., Marine botany of the Kenya coast 1. A first list of Kenya marine algae. *J East Afr Nat Hist Soc*, 36 (1967) 75-84.

5 Coppejans E & Beeckman T, *Caulerpa* section Sedoideae (Chlorophyta, Caulerpales) from the Kenyan coast, *Nova Hedwigia*, 49 (1989) 381-391.

6 Coppejans E & Beeckman T, *Caulerpa* (Chlorophyta, Caulerpales) from the Kenyan coast, *Nova Hedwigia* 50 (1990) 111-125.

7 Coppejans E & Gallin E, Macroalgae associated with the mangrove vegetation of Gazi Bay (Kenya) *Belg. J Bot,* 122 (1988) 47-60.

8 Coppejans E, Beeckman T & De Wit M, The seagrass and associated vegetation of Gazi Bay (Kenya) *Hydrobiologia*, 247 (1992) 59-75.

9 Coppejans E., Leliaert F & De Clerck O, Annotated list of new records of marine macroalgae of Kenya and Tanzania, since Isaac's and Jaasund's publications, *Biol. Jb. Dodonaea,* 67 (2000) 31-93.

10 Vine P, *Red Sea invertebrates*, (Immel Publishing, London) 1986.

11 Dautzenberg P, Mollusques testacés marins de Madagascar, Faune des Colonies Francaises, Tome III. 1929.

12 Drivas J & Jay M, Coquillages de La Réunion et de l'île Maurice, (Delachaux & Niestle, Paris) 1988.

13 MacNae W & Kalk M, *A natural history of Inhaca Island, Mozambique*, (Witwatersrand Univ. Press, Johannesburg) 1958.

14 Day J H, Polychaeta of Southern Africa, Part 1. Errantia, [British Museum (Natural History), London] 1967.

15 Veron J E N, Corals of Australia and the Indo-Pacific, (Angus & Robertson Publishers, London) 1986.

16 Rowe F W E & Gates J, Zoological Catalogue of Australia, 33. Echinodermata, (CSIRO, Melbourne, Australia) 1995.

17 Vaught K C, A classification of the living Mollusca, (American Malacologists, Inc., Melbourne, Florida) 1989.

18 Grice G D & Hulsemann K, Bathypelagic calanoid copepods of the Western Indian Ocean, in: *Proceedings of the United States National Museum*, (Smithsonian Institution, Washington, DC) Vol. 122, No. 3583 (1967) 1-67.

19 Clark A M & Rowe F W E, *Monograph of shallow-water Indo-West Pacific echinoderms*, British Museum (Natural History), London. 1971.

20 Smith J L B & Smith M M, *The fishes of Seychelles*, (Department of Ichthyology, Rhodes University, Grahamstown, South Africa) 1963.

21 Clark A M & Courtman-Stock J, *The echinoderms of Southern Africa*, [Publ. No. 766. British Museum (Natural History), London] 1976.

22 Barnard K H, Descriptive catalogue of South African decapod Crustacea, *Ann S Afr Mus,* 38 (1950) 1-837.

23 Sheppard C R C, Coral species of the Indian Ocean and adjacent seas: a synonymised compilation and some regional distribution patterns, *Atoll Res Bull*, 307 (1987).

24 Barnard K H, Additions to the fauna list of South African Crustacea and Pycnogonida. *Ann S Afr Mus,* 43 (1955), 1-107.

25 Day J H, *Polychaeta of Southern Africa. Part 2: Sedentaria,* [British Museum (Natural History) London], 1967.

26 Serene R Crustaces Decapodes Brachyoures de l'Ocean Indien et de la Mer Rouge. Xanthoidea: Xanthidae et Trapeziidae. Orstom, Collection Faune Tropicale No. 24 (1984).

27 Richmond M D, A guide to the seashores of eastern Africa and the western Indian Ocean Islands, (SIDA, Sweden) 1997.

28 NODC, NODC Taxonomic codes, CD-ROM (National Oceanographic Data Center, Maryland, USA) 1997.

29 Spry J F, *The sea shells of Dar es Salaam: Pelecypoda (Bivalves),* Tanganyika Notes and Records 63 (1964).

30 Guinot D, La Faune carcinologique (Crustacea Brachyura) de l'Ocean Indien Occidental et de la Mer Rouge. Catalogue, Remarques bibliographiques et biobliographie, *Bull Inst Fondamental d'Afrique Noire (IFAN),* (1967) 237-252.

31 Brown L H, Urban E K & Newman K, *The birds of Africa,* Volume I. (Academic Press, London) 1982, pp. 521.

32 Spry J F, *The sea shells of Dar es Salaam: Gastropods,* Tanganyika Notes and Records 56 (1961).

33 Randall J E, *Red Sea reef fishes*, (Immel Publishing, London) 1992.

34 Griffiths C L, The Amphipoda of Southern Africa Part 3: The Gammaridea and Caprellidea of Natal, *Annals of the South African Museum,* 62 (1974) 209-264.

35 Fischer W & Bianchi G, *FAO species identification sheets for fisheries purposes: Western Indian Ocean*, (FAO, Rome) 1984.

36 Richmond M D, The marine biodiversity of the Western Indian Ocean and its biogeography: how much do we know? In:*Marine science development in Tanzania and eastern Africa*, edited by M.D. Richmond & J. Francis, (IMS/WIOMSA, Zanzibar, Tanzania) 1999, pp. 241-261.

37 European Register of Marine Species, Available on line at http://www.marbef.org/data/erms.

38 Levin V S, Aspidochirote holothurians of the upper sublittoral zone of Indo-West Pacific: species composition and distribution, *Biol. Morya,* (1979) 17-23.

39 Humphreys W F, The echinoderms of Kenya's marine parks and adjacent regions Zoologische Documentatie 19, Koninklijk Museum voor Midden-Africa (Tervuren, Belgium) 1981, pp. 39.

40 Massin C, Reef-dwelling Holothuroidea (Echinodermata) of the Spermonde Archipelago (South-West Sulawesi, Indonesia), Zoologische Verhandelingen 329, Leiden, 1999, pp. 144.

41 Samyn Y & Vanden Berghe E, Annotated checklist of the Echinoderms from the Kiunga Marine National Reserve, Kenya, Part I: Echinoidea and Holothuroidea, *J East Afr. Nat. Hist. Soc,* 89 (2000) 1-34.

42 Integrated taxonomic Information System, Available online at http://www.itis.usda.gov/

43 UNESCO-IOC Register of Marine Organisms, Available on line at http://www2.eti.uva.nl/database/urmo/default.html.
44 Levitus S, *World ocean database & atlas* 1998, Second Edition, 2000.
45 Census of Marine Life. Available online at http://www.coml.org/coml.htm.
46 Ocean Biogeographic Information System, Available online at http://www.iobis.org.