RESOLVING TAXONOMIC DISCREPANCIES: ROLE OF ELECTRONIC CATALOGUES OF KNOWN ORGANISMS

VISHWAS CHAVAN, NILESH RANE, APARNA WATVE
Information Division, National Chemical Laboratory, Pune 411008, India.

AND

MICHAEL RUGGIERO

Abstract. — There is a disparity in availability of nomenclature change literature to the taxonomists of the developing world and availability of taxonomic papers published by developing world scientists to their counterparts in developed part of the globe. This has resulted in several discrepancies in the naming of organisms. Development of electronic catalogues of names of known organisms would help in pointing out these issues. We have attempted to highlight a few such discrepancies found while developing IndFauna, an electronic catalogue of known Indian fauna, and comparing it with existing global and regional databases.

Key words.—Biodiversity informatics, electronic catalogue, taxonomic discrepancies, Indian fauna

Identification of organisms is fundamental to biodiversity studies. Owing to this, the discipline of taxonomy, especially scientific nomenclature, has gained immense importance. Taxonomy provides a vocabulary to discuss the world (Knapp et al., 2002). Each name is unique and its representative organism is precisely described. It is estimated that about 1.8 million species of organisms have been formally named from the world (May, 1999) and each is recognized by a unique binomial. More than 2000 new generic names and 15000 new specific names alone are added to the zoological literature every year, and with such a multiplicity of names, problems are bound to occur. International mechanisms such as the International Code of Zoological Nomenclature (ICZN, 1999) and the International Code of Botanical Nomenclature (ICBN) are rulebooks that govern how organisms are named and they provide clear instructions on how to go about the process (Knapp et al., 2002).

International codes of nomenclature require taxonomic actions to be published and the data thus made available (Agosti and Johnson, 2002). However, nomenclatural additions or changes have to be conveyed to the ICZN or ICBN by the authors and is usually done when ratification is needed from the international authority. Similarly, discrepancies in the nomenclature are brought to the notice of the ICZN and ICBN by scientists, which are later reviewed. This process requires a long time and the availability of a large amount of literature to the scientists discussing nomenclature. In several cases, especially for taxonomists in developing countries, recent taxonomic literature including the codes themselves are unavailable. Very few libraries around the world have the financial capacity to carry the full range of literature in which systematic results are published (Agosti and Johnson, 2002). Hence, nomenclature changes are in many cases unavailable or become available much later to the developing country scientists than to their counterparts in developed world. This leads to use of old or outdated nomenclature.

On the other hand taxonomic papers by developing country scientists published in journals with regional scope, which are not scientifically abstracted, remain isolated and unnoticed by the wider scientific audience and taxonomic changes proposed or used in such papers are often neglected. This obviously leads to many discrepancies in the information available, especially about the current or correct taxonomic
hierarchy of organisms. It is thus necessary to create a system, which will lead to rapid identification of taxonomic discrepancies and their resolution. In addition, a permanent mechanism for registering and validating scientific names of organisms needs to be created at national as well as global levels. Web-based electronic catalogues can be effective in creating such a central repository of taxonomic information. In this paper, we demonstrate the use of web-based electronic catalogues (ELECATs) in identifying taxonomic discrepancies in an Indian context.

The NCL Center for Biodiversity Informatics (NCBI) is developing an Electronic Catalogue of Known Indian Fauna (IndFauna) (NCL, 2005). So far it has documented baseline information on more than 91,000 scientific names of the known Indian faunal species. The data incorporated in the database is collected from multiple sources. The main focus is on published literature including research papers, faunas, and monographs as sources of authentic and reviewed information. For those taxa especially invertebrates, on which published literature is not readily available, preserved collections from natural history museums, web-based databases and checklists are also being referenced. Thus, when collecting information, highest importance is given to “faunas” and “monographs” followed by “published research papers”, then “online and offline databases” followed by “region- and taxon-specific web sites” followed by “personal communications with experts”, and finally to “non-taxonomic publications”.

The information is carefully scrutinized for validity and accepted only if it is from reputed taxonomic institutions or experts. For each species, the taxonomic hierarchy used by Indian faunas is crosschecked diligently with that used in global taxonomic inventories such as Integrated Taxonomic Information System (ITIS, 2005), Species2000 (Species2000, 2005), Catalogue of Life: 2005 Annual Checklist (Bisby et al., 2005), Index to Organism Names (ION, 2005), European Register for Marine Species (ERMS, 2005), Systema Naturae 2000 (Brands, S. J., 1989-2005) etc. In case of any problems regarding taxonomic placement of the species concerned taxonomy experts are contacted and as per their suggestions the species are being entered in the database.

TAXONOMIC DISCREPANCIES

During this project we have noticed several taxonomic discrepancies, which need to be resolved by application of nomenclatural rules. These discrepancies can be grouped in 3 categories: hierarchical differences, spelling differences and homonymies.

Difference in taxonomic hierarchies

Several examples were found where the taxonomic hierarchy of organisms followed in India did not match that used by ITIS. This is especially true in case of some fishes, nematodes and insects. This is the result of differences in taxonomic opinions or the provisional nature of certain data in ITIS and a consensus is often difficult. In this case, the information managers can display the placement of the taxon according to alternative schemes. In spite of this option, it is necessary to conform to the international taxonomic opinion, to make the datasets interoperable with those developed in other parts of the world. This is an issue that needs to be discussed and resolved by taxonomists working in India. Although making changes in taxonomic hierarchy is technically possible in case of the electronic datasets, each change needs to be validated by the taxonomic community as some taxa may or may not be conforming to that change. Some of the examples where taxonomic hierarchy is different are given in Table 1.

As per ITIS and other taxonomic resources Sub Class Elasmobranchii is placed under Class Chondrichthyes, while Systema Naturae 2000 (Brands, S. J., 1989-2005) still recognizes it as Class Elasmobranchii.

Differences in spelling

The most common problem faced while digitizing the data was different spellings of organisms’ names. Some examples are given in Table 2.

Order Cheilostomata as per ITIS is named differently as Order Cheilostomida by ERMS (2005). Even the hierarchy under this order is not same for many species given in these two databases.
Table 1. Difference in hierarchies used in various sources.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Taxon or Scientific name</th>
<th>Indian sources</th>
<th>Integrated Taxonomic Information System</th>
</tr>
</thead>
</table>

In some cases these were misspellings, especially typographical errors. However, to follow the taxonomic norms, each misspelling needs to be reported along with the valid scientific name for avoiding future problems. Usually such wrongly spelled scientific names are reported as synonyms with a prefix “sic” as per the International Code of Zoological Nomenclature (ICZN, 1999). In some case the difference in spelling was due to different taxonomic opinions, and choosing which to use requires discussion with taxonomists. For example the Blue Whale Shark *Rhincodon* which is spelled *Rhiniodon* (Talwar and Kacker, 1984; Talwar, P. K., 1991) in Indian literature. ITIS shows *Rhiniodon* as a synonym of *Rhincodon* that was suppressed by a ruling (ITIS, 2005).

A few homonyms were identified in the cataloguing process. For instance, genus *Chaunoproctus* and genus *Microcosmus* were observed to be used in different regions for different taxa. This is a direct contradiction to nomenclature rules. Closer scrutiny of literature and taxonomic opinion revealed interesting facts of uses of these two genera. According to ITIS (ITIS, 2005) and the Taxonomicon of Systema Naturae 2000 (Brands, S. J., 1989-2005), genus *Chaunoproctus* Bonaparte, 1850 (Aves: Passeriformes: Fringillidae) is a bird, *Chaunoproctus ferreorostris* Vigors, 1829, which is now extinct as per the IUCN Redlist of Threatened Species (IUCN, 2005).
Table 2. Misspellings and difference in hierarchies in various sources.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Scientific names as per:</th>
<th>Integrated Taxonomic Information</th>
<th>Indian sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kingdom: Animalia</td>
<td>Corynosoma striemosum</td>
<td>(Bhattacharya, 1998)</td>
</tr>
<tr>
<td></td>
<td>Phylum: Acanthocephala</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class: Palaeacanthocephala</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Order: Polymorphida</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family: Polymorphida</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genus: Corynosoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Species: striemosum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Kingdom: Animalia</td>
<td>Class Mammalia</td>
<td>(Agarwal V.C., 1998)</td>
</tr>
<tr>
<td></td>
<td>Phylum: Chordata</td>
<td>Order: Cetacea</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class: Mammalia</td>
<td>Genus: Pseudorca</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Order: Cetacea</td>
<td>Species: crassidens</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family: Delphinida</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genus: Pseudorca</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Species: crassidens</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Genus: Amblypharyngodon</td>
<td>Phylum: Chordata</td>
<td>(Aditya and Raut, 2001)</td>
</tr>
<tr>
<td></td>
<td>Genus previously present in ITIS but not found as on date.</td>
<td>Class: Actinopterygii</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Family: Cyprinidae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Genus: Amblypharyngodon</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Species: melastigma</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Kingdom: Animalia</td>
<td>Kingdom: Animalia</td>
<td>(Nandi, 1993)</td>
</tr>
<tr>
<td></td>
<td>Phylum: Chordata</td>
<td>Phylum: Chordata</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Class: Actinopterygii</td>
<td>Class: Actinopterygii</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Order: Beloniformes</td>
<td>Order: Cyprinodontiformes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Family: Adrianichthyidae</td>
<td>Family: Cyprinodontidae</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genus: Oryzias</td>
<td>Genus: Oryzias</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Species: melastigma</td>
<td>Species: melanostigma</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Family: Cyprinodontidae is present under Order Cyprinodontiformes]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chaunoproctus Pearce, 1906 (Arachnida: Acari) is a group of orabitid mites as per the Indian Faunas (Sanyal and Bhaduri, 1986; Sanyal et al., 2003). The type species designated for this genus is Chaunoproctus cancellatus Pearce, 1906 collected from Sikkim, India. Interestingly, genus Zetorchella Berlese, 1916 reported from Somaliland, and genus Callopia Balogh, 1958 reported from Angola, Africa were regarded as synonyms of the genus Chaunoproctus by Balogh (1965) (Sanyal et al., 2003). Later in 1972 he again considered them as distinct genera. Recently, J. Balogh and P. Balogh (1992) again concluded that genus Zetorchella and genus Caloppia as synonyms of genus Chaunoproctus. In 2003, Sanyal and team described three new species Chaunoproctus orientalis, C. sisiri, and C. amarpurensis belonging to genus Chaunoproctus from Tripura, India (Sanyal et al., 2003). Genus Chaunoproctus is also reported from other parts of the globe as valid genus of Order Acari (Mahunka, S., 1987). As per the International Code of Zoological Nomenclature (ICZN, 1999), generic name Chaunoproctus Pearce, 1906 (Arachnida: Acari) can be a junior homonym of Chaunoproctus Bonaparte, 1850 (Aves: Passeriformes: Fringillidae).

Similarly the generic name Microcosmus Heller, 1878 (ITIS, 2005) [1877 (ION, 2005)] (Chordata, Asciidiacea) is a homonym of Microcosmus Chaudoir 1879 (Insecta, Coleoptera, Carabidae) (Saha, et al., 1992). Both names are in use and refer, respectively, to Asciidiacea as per the ITIS (2005) and ION (2005), and for a beetle as per Indian literature (Saha, et al., 1992). Opinion was sought from carabidologists and finally Wolfgang Schiller resolved this issue. Microcosmus has been described by Chaudoir (Chaudoir, 1879) under tribe Panageini within family Carabidae (Order Coleoptera: Class Insecta: Phylum Arthropoda) for the species M. flavopilosus. However, Heller already designated the same genus Microcosmus in 1877 under family Pyuridae, (Class Asciidiacea, Phylum Chordata). Hence, Emberik Strand proposed new name Microcosmodes (Strand, 1936). Later in 1940, Andrews replaced it to genus Microschemus (Yves, B., 2002). So until today the carabid species is cited as Microcosmodes flavopilosus. Thus, Indian records (Saha, et al., 1992) which show presence of Microcosmus instead of Microschemus needs to be updated.

Polypodium hydriforme, which is under phylum Cnidaria, is still considered as a valid scientific name under Phylum Pisces along with Phylum Cnidaria by ION (2005). Genus Doto described by Oken, 1815 is considered valid for phylum Mollusca while; U.S. National Museum of Natural History database (2005) displays it under phylum Arthropoda as well as phylum Mollusca. In these cases it is really difficult to place the organism in a specific hierarchy. Another example is of the Genus Cyaniris, which is present under family Lycaenidae, order Lepidoptera, class Insecta (Bingham, 1907). The same genus name Cyaniris is also given to insects belonging to family Chrysomelidae, order Coleoptera, class Insecta as per the database of the holotype collections from India present in the Belgium Museum (Drugmand, 2002). The genus is not included in the ITIS database while the other web sites and literature sources shows presence of genus Cyaniris in order Lepidoptera.

While checking hierarchy for a species, namely Idia pristis, which is placed under class Hydrozoa (Ritchie, 1910), surprisingly we came across genus Idia which is also present under class Insecta, order Lepidoptera, family Noctuidae as per the checklist of Noctuoidea of Ontario present on the website hosted by Canadian Biodiversity Information Facility (Troubridge and Lafontaine, 2003). There is also no ready reference available for this generic designation anywhere.

**DISCUSSION**

A consensus on these issues is a matter of taxonomic discussion. They need to be resolved by using nomenclatural rules, which requires further detailed research. However these examples effectively demonstrate the potential of electronic catalogues (ELECATs) in bringing issues or discrepancies to the attention of the taxonomic community, starting a dialogue between taxonomists across the globe and identifying issues of common concern. In order
to notice such discrepancies and resolve them quickly, it is essential that a wrapper be
developed which traverses through various
electronic catalogues searching for taxonomic
anomalies. This calls for increasing
collaboration among the various ELECATs.

The information available so far on the
Internet is largely limited to names and citations.
It is understood, however, that taxonomic
literature generated out of two centuries’ work
cannot be put on the Internet overnight. It is a
formidable task, yet one that must be
accomplished as promptly as possible. With the
growing use of information and communication
technologies in biodiversity research, it should
be possible to make the taxonomic literature
itself available on the Internet, which can be
used for checking inconsistency in taxonomy,
used worldwide. Although taxonomists from
around the world have been dealing with these
discrepancies, it is time-consuming and tedious
to identify, check, and correct them using the
traditional media such as published literature.
Modern information and communications tools
can be of immense help in identifying taxonomic
discrepancies quickly and resolving them in a
collaborative manner leading to globally
acceptable standardized inventories. With the
use of the Internet, there can be a truly two-way
exchange of information between taxonomists
from developed and developing countries.
Active collaboration and commitment of
taxonomists and information managers are
required to work towards the goal of developing
information systems to bring uniformity and
precision to taxonomic inventories across the
world. Many of the discrepancies arise because
taxonomists are unable or find it difficult to
check up on taxon names especially for taxa
outside their field of expertise.

Hence, to build up easy communication
pathways and reduce the time input, it would be
extremely helpful to have a web-based central
registry system for taxonomic names. Checking
of names being used in publication with the
central registry would definitely eliminate many
of the commonly encountered discrepancies
described above. Thorne (2003) also proposed
the need for registration of new taxa names in a
central registry of names. The journal Nature
has already taken a step towards registering of
names (Anonymous, 2002) by requiring the
authors of papers featuring new taxonomy to file
the information with a recognized institute such
as Linnaean Society of London. Central registry
will be a repository of scientific name
information or an index for scientific names in
use, along with their history. It will be a
dynamic register for proposed scientific names
(which will be provisionally accepted, noted),
which can later be added to the repository after
annotation. These will serve as references for
scientists describing new taxa to check if the
name has been used before, and in which
context. This will eliminate generation of
homonyms. It can also provide a point of
“normalization” for data.

ELECATs offer an effective method of
creating unique electronic registers. Owing to
the rules of acceptance of scientific names,
names cannot be registered as valid before the
publication of taxon description in a journal. To
solve this, a precedent can be set that in case of
each new description, together with the type
specimen deposition number, a provisional
registration number in the global, regional or
national web based ELECAT should be quoted.
There will be two-way information exchanges
with other ELECATs. The registry will compare
between ELECATs information, find out any
points of mismatches or conflicting data, and
also pick up new information automatically from
the ELECATs. Using this, a single number
reference system for each scientific name can be
developed. The central registry can provide a
minimum standard and starting point for use in
other databases.

The Global Biodiversity Information
Facility (GBIF), together with the Taxonomic
Database Working Group (TDWG), is currently
seeking requirements for Globally Unique
Identifiers (GUIDs) for biodiversity informatics
and to establish infrastructure to support their
use. GUIDs once developed can overcome most
of the current problems, such as (a)
identification of same data records served from
multiple locations, (b) referring to data from
outside network, irrespective of frequent change
of URLs, and (c) referring to taxon concepts in
reliable and consistent way. Page (2005)
suggests a system of Life Science Identifiers (LSIDs) as unique numerical identifiers for scientific names and ITIS currently employs a system of unique Taxonomic Serial Numbers (TSNs). Databases could be mapped to TSNs or some parallel concept.

Many organizations are working towards building up registers of published scientific names of taxa such as for beetles (Vratislav, 2005). Plant names can be checked using International Plant Names Index (IPNI, 2005). Index to Organism Names (ION, 2005) database can be used to check zoological names. The most holistic efforts are those of Species2000 and ITIS Catalogue of Life (Leslie, 2005) and GBIF, which aim to create an index of at least 95% of the known species by 2011 (GBIF, 2005). This is a major step towards developing a central register of names, and increasing collaborations between similar efforts worldwide should shorten the time required. To complement these initiatives, International Commission of Zoological Nomenclature announced its intention of setting up of web-based, open access mandatory registration system called “ZooBank” to register descriptions of all new taxa and nomenclatural acts in animal taxonomy. (Polaszek et al., 2005). While similar mandatory registration mechanism exists for descriptions of new bacteria, it needs to be extended for other kingdoms viz., Plantae, Archaea, Chromista, Fungi, Protozoa, and Viruses. Similar to “ZooBank” these mandatory registration mechanisms would facilitate retrospective registration of existing names, and of all nomenclatural acts in respective kingdom. This can be achieved through active linkages and collaborations with existing projects, rather than replacing them.

In addition, links to other databases like image, DNA sequences, protein sequences, lipid sequences, and collection locality maps, etc. will be a major step forward. Further, linking valid scientific names to collection accession numbers of type specimens will help scientists track all collections quickly and know where they are deposited. Applications could be built such as those on the ITIS Canada website or uBio1 that display multiple classifications. In addition, a taxonomist’s time would be saved by having a tool that can readily compare data with that in the central file (such as ITIS’ Taxcompare tool). The ability to check quickly for homonyms will also save time. In addition, development and use of national ELECATs should be encouraged to collate information at the national levels and make it available to the global users. This is especially important, as these national registers will be able to easily access locally available primary taxonomic information.

It is also necessary to track changes in species concepts over time. ITIS has developed a capability for change tracking, but has not yet implemented it. Availability of specimens, images, protologues, and classifying characters in use in different countries, comparing between specimens of a species with wide distribution-transcending political boundaries and building biogeographic distribution maps, language barrier - translating Latin diagnoses, and picture data are some of the capabilities required to ensure accurate results in biodiversity research projects.

These advances suggest that, in the future, the taxonomic discipline will make broad use of the web-based information and benefit greatly from it. Therefore, it is crucial at the moment to build up or improve the collaborative activities among domain experts, information managers and users of taxonomic information. This would ultimately help in strengthening the biodiversity research necessary for conservation and management of global natural resources.

CONCLUSIONS

The discrepancies found while developing IndFauna, an electronic catalogue of known Indian fauna and comparing it with existing databases can help to solve several issues like taxonomic ambiguities, inadequate documentation and incorrect placements of species. Development of electronic catalogues of names of known organisms (ELECATs) will help in pointing out these issues. International organizations like GBIF are trying to make all biodiversity data accessible to the largest possible section of the human population. Recently GBIF, Species 2000, ITIS and uBio (GBIF, 2005) have decided to cooperate on

1 http://www.ubio.org/
compiling and utilizing taxonomic information resources. National and regional resources such as IndFauna, after solving the types of discrepancies described here, can make valuable contributions to preparing of global taxonomic databases and standards.

REFERENCES

Drugmand D. 2002. Entomological and Arachnological Collections. Royal Belgian Institute of Natural Sciences, Brussels.³
European Register for Marine Species (ERMS). 2005. European Register for Marine Species (ERMS2), Flanders Marine Institute, Oostende, Belgium.³

⁵ [http://www.iczn.org/iczn.htm](http://www.iczn.org/iczn.htm)
⁶ [http://www.biosis.org.uk/ion/search.htm](http://www.biosis.org.uk/ion/search.htm)


National Chemical Laboratory (NCL). 2005. Electronic Catalogue of Known Indian Fauna. National Chemical Laboratory, Pune, India.10


Species2000. 2005. Species2000, School of Plant Sciences, The University of Reading, Reading, United Kingdom.11


