AN ALGORITHM FOR IDENTIFYING MISSPELLINGS AND SYNONYMS IN LISTS OF SCIENTIFIC NAMES OF FISHES

by

R. FROESE (1)

ABSTRACT. - A computerized algorithm is presented for the examination of lists of scientific names, assigning current names to known synonyms and misspellings, and suggesting the most probable correct name for new combinations and new misspellings. The algorithm was implemented in a system that made use of W.N. Eschmeyer's fish genera database of 1990 and of species names and synonyms compiled in FishBase, a biological database on finfish containing, at the time of this study 15,000 of the estimated 25,000 recent species of fish. The algorithm was tested with six independent data sets with altogether 10,711 scientific names. About 60% of these names either directly matched valid names, synonyms or known misspellings, or could be assigned automatically to a valid name. For the remaining 4,385 names, the algorithm made suggestions which led to the identification of 1,218 synonyms, 798 misspellings, and only 72 cases that needed more research and could not be decided based on the information provided by the algorithm. No match was found for 2,146 names referring mainly to species not yet contained in FishBase.

RÉSUMÉ. - Un algorithme pour identifier les synonymes et les erreurs d'orthographe dans les listes de noms scientifiques de Poissons.

Cet article présente un algorithme informatisé pour la vérification de la nomenclature scientifique. Cet algorithme consiste à faire correspondre un nom scientifique à des synonymes ou noms mal orthographiés déjà répertoriés dans la base de données et, à défaut, à suggérer le nom correct le plus probable en procédant à de nouvelles combinaisons avec les synonymes et noms mal orthographiés les plus proches. Cet algorithme est réalisé au sein d'un système qui exploite la base de données sur les genres de poissons de W.N. Eschmeyer datant de 1990 ainsi que les noms et synonymes répertoriés dans FishBase, une base de données biologique sur les poissons contenant 15 000 des 25 000 espèces récentes de poissons estimées. L'algorithme a été testé sur six séries de données indépendantes composées au total de 10 711 noms scientifiques. Environ 60% de ces noms ont pu être associés à des noms valables, à des synonymes ou à des noms mal orthographiés connus, ou ont pu être automatiquement rapportés à un nom valable. Pour les 4 385 noms restants, l'algorithme a suggéré des combinaisons qui ont abouti à l'identification de 1 218 synonymes, 798 mots mal orthographiés et seulement 72 cas pour lesquels aucune décision n'était possible à partir des informations fournies par l'algorithme. Aucun recoupement n'a pu être réalisé pour 2 146 noms d'espèces non encore répertoriés dans FishBase.

Key-words. - Taxonomic databases, Misspellings, Synonyms, Nomenclature.

One of the more tedious tasks of taxonomists is to check and update long lists of species names that biologists, ecologists, authors, resource managers and, more recently, those with interest in biodiversity have sent to them for comment. Typically these lists include many synonyms and misspellings. Also, museums have begun to computerize

(1) International Center for Living Aquatic Resources Management (ICLARM), MCPO Box 2631, Makati City 0718, PHILIPPINES.

ICLARM Contribution No. 1267.

their collections and are faced with the problem of updating names that might have been assigned more than 200 years ago. Manual verification of scientific names can be a frustrating task, as is demonstrated by Bailly and Hureau (1995) who found 14 different spellings for a single species name (*Bregmaceros mcclellandi*) in two bibliographic databases, as well as 14 different citations of the original publication.

This paper presents an algorithm which largely automates and simplifies the process of verifying scientific names. A prototype of this algorithm has been implemented in FishBase, a biological database on finfish developed by the International Center for Living Aquatic Resources Management (ICLARM) in collaboration with the Food and Agriculture Organization of the United Nations (FAO) and numerous individual experts and institutions, with support from the European Commission (EC) (Pauly and Froese, 1991; McCall and May, 1995; Froese and Pauly, 1995, 1996).

The goal of the algorithm is twofold: first, it aims to assign a valid name to as many as possible of the submitted names; second, it aims to provide all necessary information to assist the user in selecting a valid name for cases where a valid name can not be assigned automatically.

MATERIALS AND METHODS

The algorithm is based on the rules of the *International Code of Zoological Nomenclature* (1985; ICZN) and was optimized from the experience gained with the prototype described below. It is divided into three parts which are documented in Appendix 1, 2 and 3, respectively. The chosen format should make it easy to implement the algorithm in any database programming language.

The algorithm uses the following sources of information:

1. The FishBase FAMILIES table which contains the 469 family names of fishes as recognized by Eschmeyer (1990), plus additional information from other sources, notably Nelson (1994). The algorithm used only the field containing the valid family name.
2. The GENERA table developed by W.N. Eschmeyer and corresponding to his monograph *Catalog of Genera of Recent Fishes* (Eschmeyer, 1990), which contains the more than 10,000 generic and subgeneric names used for fishes until the end of 1989. W.N. Eschmeyer made GENERA available for use and distribution through FishBase. For the purpose of this study, only four fields of GENERA were used: GenName, Status (as assigned by Eschmeyer, 1990), CurrentGenus, and Family;
3. The FishBase SYNONYMS table, which contains about 15,000 valid names of the estimated 25,000 species of recent finfish plus about 24,000 synonyms classified in one of the following categories: junior synonym, original combination, new combination, misspelling, misidentification, questionable, other; only synonyms of the first 4 categories were considered. The algorithm used the following fields of the SYNONYMS table: SynGenus, SynSpecies (includes subspecies), SynAuthor, SynStatus, Valid (yes/no), SynRef (reference used to derive the status of the synonym), CurrentGenus, CurrentSpecies (includes subspecies), CurrentAuthor, and Family.

In preparation for running the algorithm, a list of scientific names (binomia and trinomia) is imported into a table that allows additional information to be assigned to each name. The following 5 fields are provided for import: ImportGenus, ImportSpecies (includes subspecies), ImportAuthor (with year and parentheses, if required), ImportFamily, and ImportCommonName. The examination of the names follows four major steps:

Masculine	Feminine	Neuter
-us	-a	-um
-is, no suffix	-is	-e
-er	-era, -ra	-erum, -rum
-ns	-ns	-ns
-x	-x	-x
-as	-as	-as
-ior	-ior	-ius

Table I. - Equivalent suffixes used for finding identical (sub)specific epithets.

Suffix	Replacements
ii	i, no character
i	ii, no character
us	ns, as, is
ns	us, as
as	us, ns

Table II. - Frequent suffix misspellings considered for finding matching (sub)specific epithets.

1. If a submitted name matches directly a valid, synonymous, or misspelled name in FishBase, the algorithm adds the valid name, author and year, family, and the reference on which the status of the valid name or synonym was based and proceeds with the next name (Steps 1 and 25 in Appendix 1).
2. If no direct match is found, the algorithm tries to find unambiguous corresponding names by testing a combination with the valid genus - when the submitted genus was a synonym - and several modifications of the suffix of the (sub)specific epithet. The suffixes listed in table I are replaced with the corresponding suffixes in other genders; the suffixes listed in table II are replaced accordingly to account for common misspellings (Steps 6-24 in Appendix 1). If a match is found, the algorithm adds to the imported record the valid name, author and year, family, and the reference on which the status of the synonym is based.
3. If Steps 1 and 2 have found no matching name, the algorithm checks the validity of the generic name. If a matching valid, synonymous or misspelled generic name is found, the algorithm adds the valid genus and family to the submitted name (Steps 2-5 in Appendix 1). If no matching generic name can be found, the algorithm tries to find similar names. It first tests for misspellings in the last one or two characters of the submitted generic name (Steps 1-5 in Appendix 2). It then tests for misspellings in the second and third character by replacing them with combinations of single wildcards, i.e., a special character (?) that matches any single character (but not zero characters; Steps 6-9 in Appendix 2). Finally it replaces the fourth and following characters with a general wildcard, i.e., a special character (*) that matches zero or more characters. Generic names of fishes range in length from 2 to 23 characters (Eschmeyer, 1990). To limit the number of matching names, only those with similar length as the submitted name are considered (Steps 10-14 in Appendix 2). The resulting comments and possible generic names are printed. Based on these suggestions, obviously misspelled generic names are corrected manually and the algorithm is rerun starting with Step 1 above.
4. If Steps 1-3 have not found a matching name, the algorithm makes suggestions for those names for which at least a valid generic name had been found. Comparisons are limited to (sub)specific epithets within the same family. It first tests for misspellings in the second and third character of the (sub)specific name (Steps 1-5 in Appendix 3). It then replaces the fourth and following characters with a general wildcard. (Sub)specific epithets contained in FishBase range in length from 2 to 20 characters; only matching names of similar length to the submitted name are

accepted (Steps 6-10 in Appendix 3). Comments and possible names are printed. Based on these suggestions, further new combinations, synonyms and misspellings are identified and manually corrected.

The algorithm was tested with the following data sets: a checklist from a recent book on Indo-Pacific fishes (Book); a curatorial database of a small museum collection of Indo-Pacific fishes (Museum 1); a fish collection database of a small museum in North America (Museum 2); a reference list used by a fisheries department in the Pacific (Fisheries); a species list from a trawl survey conducted in 1980 in the Indian Ocean (Trawl 1); and a recent reference list used for trawl surveys in the Atlantic, Pacific and Indian Oceans (Trawl 2). These documents are not further identified here to protect their compilers from undue criticism and to emphasize that the errors they contained are typical of the class of documents they represent.

The scientific names from these data sets were processed and the following statistics were recorded:

⇒ Species: total number of submitted unique scientific names;

⇒ Automatic match: number of submitted names that matched with a valid, synonymous, or misspelled name or that could be automatically assigned to a valid name, as described in Step 2 above;

⇒ Synonyms (Suggestions): number of synonymous names that could be manually assigned to a valid name, using no other information than the provided printout;

⇒ Misspelling (Suggestions): number of misspelled names that could be manually assigned to a valid name, using no other information than the provided printout;

⇒ More research (Suggestions): number of names that required more information than provided in the printout;

⇒ No match: number of names for which no match could be found.

RESULTS

The results of applying the algorithm to six data sets with altogether 10,711 scientific names are summarized in table III. About 60% of these names either directly matched valid names, synonyms or known misspellings, or could be assigned automatically to a valid name. See tables IV and V for examples of the resulting printout. For the remaining 4,385 names, the algorithm resulted in suggestions which led to the identification of 1,218 synonyms, 798 misspellings, and 72 cases that could not be decided based on the information provided by the algorithm. No match was found for 2,146 names, mainly because these species were not yet contained in FishBase. See table VI for an example of the suggestions that were the basis for manually assigning valid names to new misspellings and combinations.

DISCUSSION

A routine roughly similar to that presented here, i.e., capable of comparing scientific names against a reference list, exists for legumes (F. Bisby, pers. comm.). However, to date that approach has not been published and the author was not able to compare it with the approach suggested here. It is planned to provide an algorithm similar to the one presented here for use with the checklist of all known organisms that is currently developed by the SPECIES 2000 initiative (Bisby and Smith, 1996).

Table III. - Results of evaluating six lists of scientific names. See text for more details.

Data sources	Species	Autom. match	Synonyms (Suggestions)	Misspelling (Suggestions)	More research (Suggestions)	No match
Book	2066	1321	44	152	9	481
Museum 1	1032	691	44	85	16	168
Museum 2	3049	1433	361	287	12	956
Fishery	1378	1026	25	98	4	218
Trawl 1	970	159	670	42	15	63
Trawl 2	2216	1696	74	134	16	260
Total	10711	6326	1218	798	72	2146

Table IV. - Example of a printout resulting from algorithm described in Appendix 1. Scientific names submitted for checking are in bold. Current status and author of the name according to the SYNONYMS table in FishBase are given in square brackets. The second line gives valid name, author, and reference on which the status of the synonym is based.

Abantennarius analis [original combination, Goseline 1957]
= *Antennarius analis* (Goseline 1957), Ref. Pietsch, T.W. and D.B. Grobecker, 1987, p. 163
Abantennarius neocaledoniensis [junior synonym, Le Danois 1964]
= *Antennarius coccineus* (Lesson 1830), Ref Pietsch, T.W. and D.B. Grobecker, 1987
Abeona aurora [new combination, (Jordan & Gilbert 1880)]
= *Micrometrus aurora* (Jordan & Gilbert 1880), Ref. Eschmeyer, W.N., E.S. Herald and H. Hammann, 1983, p. 231
Ablabys binotata [misspelling, (Peters 1855)]
= *Ablabys binotatus* (Peters 1855), Ref. Poss, S., 1986, p. 479
Abramis pekinensis [original combination, Basilewsky 1855]
= *Parabramis pekinensis* (Basilewsky 1855), Ref. Berg, L.S., 1964, p. 358
Abramites solarii [junior synonym, (Holmberg 1891)]
= *Abramites hypselonotus* (Günther 1868), Ref. Vari, R.P. and A.M. Williams, 1987, p. 89

Table V. - Example of a printout resulting from the algorithm described in Appendix 2. Generic names from the submitted list that do match synonyms in Eschmeyer's (1990) GENERA database are shown in bold. The author of the name is given within the square brackets.

Abantennarius [Schultz 1957, synonym]	= *Antennarius*, Daudin 1816
Abeona [Girard 1855, synonym]	= *Micrometrus*, Gibbons 1854
Abramocephalus [Steindachner 1869, synonym]	= *Hypophthalmichthys*, Bleeker 1859
Abranches [Smith 1947, synonym]	= *Gobiopsis*, Steindachner 1860
Abrostomus [Smith 1841, synonym]	= *Labeo*, Cuvier 1816
Abryois [Jordan & Snyder 1902, synonym]	= *Pholidapus*, Bean & Bean 1896

The presented algorithm is a preliminary approach to identifying synonyms and misspellings in lists of scientific names. It is straightforward for valid names and re-

corded synonyms and misspellings and its effectiveness for these categories solely depends on the completeness of the underlying GENERA and SYNONYMS table. The approach to assigning valid names automatically to unrecorded misspellings and combinations was a cautious one. According to ICZN, the suffix of a (sub)specific epithet that is or ends in a Latin adjective must agree in gender with the generic name (Article 31(b), International Code of Zoological Nomenclature, 1985). Because of this variability and because the gender of the generic name and the type of the (sub)specific epithet (adjective or noun in apposition) are not always clear, allowed and erroneous variations in the suffix of a (sub)specific epithet are more frequent than in the stem. Since it is the stem and not the

Table VI. - Example of a printout resulting from the algorithm described in Appendix 3. Submitted scientific names and authors are given in bold. Indented lines show similarly spelled specific epithets from within the same family, with their status and the reference on which the status is based. An 'Esch.' indicates that the spelling of the scientific name and author have been checked against a 1996 version of Eschmeyer's PISCES database. The user has to assign manually a valid species to the submitted name. Note that the author name is essential for correct assignments. *Phoxinellus handlirschi* is a new combination of *Acanthorutilus handlirschi* Pietschmann 1933, *Puntius katalo* is a misspelling of *P. katolo* (Herre, 1924), and *Solegnathus dunckeri* (Whiteley, 1927) is a valid name not yet contained in FishBase.

***Phoxinellus handlirschi* (Pietschmann 1933)**

handlirschi, Acanthorutilus Pietschmann 1933, Valid [Robins, C.R., R.M. Bailey, C.E. Bond, J.R. Brooker, E.A. Lachner, R.N. Lea and W.B. Scott, 1991, p. 36] Esch.

hankinsoni, Hybognathus Hubbs 1927, Valid [Page, L.M. and B.M. Burr, 1991, p. 109, 374]

***Puntius katalo* (Herre 1924)**

katangae, Barbus Boulenger 1900, junior synonym of *Barbus trimaculatus* Peters 1852 [Lévêque, C. and J. Daget, 1984, p. 299-300]

katangae, Chelaethiops Poll 1948, junior synonym of *Chelaethiops congicus* (Nichols & Griscom 1917) [Lévêque, C. and J. Daget, 1984, p. 299-300]

katinensis, Xenocypris Tchang 1930, questionable of *Xenocypris macrolepis* Bleeker 1871 [Berg, L.S., 1964, p. 173] Esch.

katolo, Barbodes Herre 1924, original combination of *Puntius katolo* (Herre 1924) [Herre, A.W., 1953, p. 125] Esch.

katolo, Puntius (Herre 1924), Valid [Herre, A.W., 1953, p. 125] Esch.

***Solegnathus dunckeri* (Whiteley 1927)**

dunckeri, Syngnathus Metzelaar 1919, original combination of *Bryx dunckeri* (Metzelaar 1919) [Robins, C.R., R.M. Bailey, C.E. Bond, J.R. Brooker, E.A.Lachner, R.N. Lea and W.B. Scott, 1991, p. 37, 84]

dunckeri, Bryx (Metzelaar 1919), Valid [Robins, C.R., R.M. Bailey, C.E. Bond, J.R. Brooker, E.A. Lachner, R.N. Lea and W.B. Scott, 1991, p. 37, 84]

dunckeri, Halicampus (Chabanaud 1929), Valid [Myers, R.F., 1991, p. 88, 89]

dunckeri, Microphis (Prashad & Mukerji 1929), Valid [Talwar, P.K. and A.G. Jhingran, 1992, p. 772] Esch.

dunckeri, Doryichthys Prashad & Mukerji 1929, original combination of *Microphis dunckeri* (Prashad & Mukerji 1929) [Talwar, P.K. and A. G. Jhingran, 1992, p. 722] Esch.

dunckeri, Micrognathus Chabanaud 1929, original combination of *Halicampus dunckeri* (Chabanaud 1929) [Dawson, C., 1986, p. 451]

suffix that differentiates (sub)specific epithets, certain variations in the suffix can be safely ignored in the matching process.

It was assumed safe to assign valid names automatically if: i) the submitted genus was identical with or a known synonym of the valid genus; ii) the stem of the submitted (sub)specific epithet matched that of the valid name; and iii) the suffix of the submitted (sub)specific epithet matched directly, or was of another gender (Table I), or one of a few, common misspellings (Table II). Note that the sequence in step 8, Appendix 1, ensures that, e.g., the suffix "um" is not found before "erum". If a safe method could be devised to distinguish automatically between the stem and the suffix - including irregular adjectives such as *paluster, palustris, palustre* and *major, majus* or nouns without suffix such as *cuvier* - the suffix might be ignored altogether and the matching of (sub)specific epithets might be based on the stem only.

The automatic matching routine is based only on the submitted binomina or trinomina, and ignores additional provided information on, e.g., families or authors. Such information can theoretically be used, for example, to search for misspelled genera only within a given family, or use the authority to distinguish between identical (sub)specific epithets within a family. However, the assignment of genera to families can follow different taxonomies, and the spelling of names of authorities differs widely. Such additional information is utilized to assist the user in matching the name submitted with similarly spelled names in the provided printout (see table VI).

For generic names, automatic assignment by replacing the last one or two characters with, e.g., single wildcards was not an option, because differences in these characters can actually describe distinct genera, such as in *Acanthogobio* Herzenstein 1892, in Cyprinidae and *Acanthogobius* Gill 1859, in Gobiidae (Eschmeyer, 1990). The algorithm as presented here does not accommodate checking of subgeneric names (tetranomia). However, Eschmeyer's GENERA contains also subgeneric names and treats them as synonyms of the respective genus. Thus, if the submitted generic name was considered subgeneric by Eschmeyer (1990), the scientific name would be accepted and the status of the generic name would be pointed out in the printout resulting from the procedure described in Appendix 2.

It might be asked why the algorithm does not start with the examination of the generic name, as would seem logical. The used sequence ensures that information contained in the synonymy is considered before the generic name is used to assign a family to the submitted name (Step 3 above; Steps 2-5 in Appendix 1). For example, *Chromis niloticus* is a known synonym of *Oreochromis niloticus* in Cichlidae. However, *Chromis* is a valid genus in the family Pomacentridae. Thus, the algorithm makes use of the fact that the binomen contains more information than either the generic name or the specific epithet alone.

The approach used for finding new misspellings could be improved to avoid offering names that are clearly non-matching (see table VI). For example, one could subsequently replace each character with a single wildcard, two single wildcards, and no character to account for mistyped, additional, or missing characters, respectively. However, every such combination has to be compared with sometimes more than 1,000 names, which increases the search time considerably while still not finding cases where more than one character are mistyped or where subsequent characters are inverted, both of which happen quite often. The algorithm found most of such cases without suggesting too many obviously different names. For example, of the 14 different spellings of *Bregmaceros mcclellandi* listed in Bailly and Hureau (1995), two (*B. mcclellandi* (valid) and *B. macclel-*

landii (misspelling)) were contained in the FishBase SYNONYMS table and thus matched directly, nine could be automatically assigned to the valid name, two were suggested as being similar to the valid name, and one spelling (*B. Mac Clellandii*) which is not permitted under ICZN, was not found.

A search using the sound-alike wildcard (~) of two database systems (DataEase, 1990 and Microsoft Access, 1994) did not produce useful results. This, however, could probably be improved if the underlying routine could be trained on scientific names.

Some colleagues have compared the algorithm proposed here to a spell-checker, but this comparison is misleading. A spell-checker makes use of pattern-matching routines to find similar combinations of characters without any consideration of what the words mean. The algorithm presented here makes use of the information contained in the taxonomic hierarchy to restrict the search for a (sub)specific epithet to the corresponding genus or family; it also makes use of the synonymies contained in the latest revisionary work, notably Eschmeyer's recent revision of the generic names. It also makes use of the regulations of the International Code of Zoological Nomenclature (1985) by distinguishing between the stem and the suffix of a (sub)specific epithet.

The six data sets tested stemmed from a variety of sources (see above) and are probably representative of similar compilations. All data sets were recent with the exception of Trawl 1 which was assembled in 1980 and which had the lowest percentage (16%) of directly matching names. Overall, the algorithm was able to assign valid names automatically to about 60% of the names submitted. If we assume that the 2,146 names in the "No match" category are not yet in FishBase and exclude these names from the comparison, the percentage of direct hits was 74%. In other words, with a database containing all species of a group and with a synonymy similar to that of FishBase, the algorithm can be expected to find automatically, on the average, corresponding valid names for three out of four submitted names. Conversely, similar lists of scientific names can be expected to contain 26% new combinations or misspellings.

For the 4,386 names for which no automatic match could be found, the algorithm made 2,016 (46%) acceptable suggestions and identified 72 cases that could be solved with additional information. If we again exclude the "No match" names from this consideration, 90% of the suggestion made by the algorithm led to the correct assignment of a submitted new combination or new misspelling.

Acknowledgements. - Thanks are due to Daniel Pauly and Roger Pullin for their comments on the manuscript and to the FishBase Team for entering and checking the many names in the FAMILIES, SPECIES, and SYNONYMS table. Special thanks to Portia Bonilla for programming the various procedures in Microsoft Access, to Catherine Binudin for translating the abstract, and for Emily Capuli for doublechecking the algorithms presented in Appendices 1-3. More than thanks are due to W.N. Eschmeyer for making his GENERA database available to us, and for compiling it in the first place. His monograph and database on all species of recent fishes, to be published in 1997, will be a major contribution to modern taxonomy. A preliminary version of this database was available in 1996 at the home page of the California Academy of Sciences at:

http://www.calacademy.org/research/ichthyology/species/

REFERENCES

BAILLY N. & J.-C. HUREAU, 1995. - Bases de données en biologie: quelques problèmes liés à la nomenclature et aux références bibliographiques. *Cybium*, 19(4): 323-324.

BERG L.S., 1964. - Freshwater fishes of the U.S.S.R. and adjacent countries. Vol. 2, 4th Edition. Israel Program for Scientific Translations Ltd., Jerusalem.

BISBY F. & P. SMITH, 1996. - Species 2000 Project Plan. 44 p. Univ. Southampton, Southampton, UK.

DATAEASE INTERNATIONAL, 1990. - DataEase reference manual. Trumbull, USA.

DAWSON C., 1986. - Syngnathidae, pp. 445-458. *In:* Smith's Sea Fishes (Smith M.M. and P.C. Heemstra, eds). Springer-Verlag, Berlin.

ESCHMEYER W.N., 1990. - Catalog of the Genera of Recent Fishes. 697 p. California Academy of Sciences, San Francisco.

ESCHMEYER W.N., HERALD E.S. & H. HAMMANN, 1983. - A field guide to Pacific coast fishes of North America. 336 p. Houghton Mifflin Company, Boston, USA.

FROESE R. & D. PAULY, eds, 1995. - FishBase: a biological Database on Fish (Version 1.2). 146 p. ICLARM, Manila, Philippines.

FROESE R. & D. PAULY, eds., 1996. - FishBase 96: concepts, design and data sources. 179 p. ICLARM, Manila, Philippines.

HERRE A.W., 1953. - Check list of Philippine fishes. Res. Rep. U.S. Fish. Wildl. Serv., (20): 977 p.

INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE, 1985. - International Trust for Zoological Nomenclature. 338 p.

LÉVÊQUE C. & J. DAGET, 1984. - Cyprinidae, pp. 217-342. *In:* Check-List of the freshwater Fishes of Africa (CLOFFA). Vol. 1 (Daget J., Gosse J.-P. & D.F.E. Thys van den Audenaerde, eds). ORSTOM, Paris and MRAC, Tervuren.

McCALL R.A. & R.M. MAY, 1995. - More than a seafood platter. *Nature*, 378: 735.

MICROSOFT, 1994. - Microsoft Access language reference. Microsoft Corporation, USA.

MYERS R.F., 1991. - Micronesian reef fishes. Second Edition. 298 p. Coral Graphics, Barrigada, Guam.

NELSON J.S., 1994. - Fishes of the World. Third Edition. 600 p. John Wiley & Sons, New York.

PAULY D. & R. FROESE, 1991. - FishBase: Assembling Information on Fish. *Naga, ICLARM Q.*, 14(4): 10-11.

PIETSCH T.W. & D.B. GROBECKER, 1987. - Frogfishes of the World. Systematics, Zoogeography, and behavioral Ecology. 420 p. Stanford Univ. Press, Stanford, California.

POSS S., 1986. - Tetrarogidae, pp. 479. *In:* Smith's Sea Fishes (Smith M.M. & P.C. Heemstra, eds). Springer-Verlag, Berlin.

ROBINS C.R., BAILEY R.M., BOND C.E., BROOKER J.R., LACHNER E.A., LEA R.N. & W.B. SCOTT, 1991. - Common and scientific Names of Fishes from the United States and Canada. Amer. Fish. Soc. Spec. Publ., (20): 183 p.

ROBINS C.R., BAILEY R.M., BOND C.E., BROOKER J.R., LACHNER E.A., LEA R.N. & W.B. SCOTT, 1991. - World Fishes important to North Americans. Exclusive of Species from the continental Waters of the United States and Canada. Amer. Fish. Soc. Spec. Publ., (21): 243 p.

TALWAR P.K. & A.G. JHINGRAN, 1992. - Inland fishes of India and adjacent countries. Vol. 2. A.A. Balkema, Rotterdam.

VARI R.P. & A.M. WILLIAMS, 1987. - Headstanders of the neotropical anastomid genus *Abramites* (Pisces: Characiformes: Anostomidae). *Proc. Biol. Soc. Wash.*, 100(1): 89-103.

Reçu le 20.09.1996.
Accepté pour publication le 30.01.1997.

APPENDIX 1

Algorithm to automatically assign a valid name to a submitted list of zoological names. See table IV for an example of a resulting printout. Note that we refrained from summarizing steps to facilitate translation into any database query language.

1a		Submitted name matches a valid name, a synonym, or a recorded misspelling ---------- **25**
1b		Submitted name does not match ---------- **2**
2a	(1b).	Generic name valid ---------- **5**
2b	(1b).	Generic name synonym, recorded misspelling, or unknown ---------- **3**
3a	(2b).	Generic name synonym or recorded misspelling ---------- **4**
3b	(2b).	Generic name unknown ---------- **26**
4a	(3a).	Combination of valid Genus and submitted (sub)specific name matches a valid name, a synonym, or a recorded misspelling ---------- **25**
4b	(3a).	Combination of valid Genus and submitted (sub)specific name does not match ---------- **5**
5	(2a, 4b).	**Action:** Add valid Genus and Family to the submitted name ---------- **6**
6a	(5).	Submitted name is a trinomen ---------- **7**
6b	(5).	Submitted name is a binomen ---------- **8**
7a	(6a).	Combination of submitted or valid Genus and subspecific epithet matches a valid name, a synonym, or a recorded misspelling ---------- **25**
7b	(6a).	Combination of submitted or valid Genus and subspecific epithet does not match ---------- **8**
8a	(6b, 7b).	Submitted specific or subspecific name ends on "erum" ---------- **9**
8b	(6b, 7b).	Submitted specific or subspecific name ends on "era" ---------- **10**
8c	(6b, 7b).	Submitted specific or subspecific name ends on "er" ---------- **11**
8 d	(6b, 7b).	Submitted specific or subspecific name ends on "ior" ---------- **12**
8e	(6b, 7b).	Submitted specific or subspecific name ends on "ius" ---------- **13**
8f	(6b, 7b).	Submitted specific or subspecific name ends on "is" ---------- **14**
8 g	(6b, 7b).	Submitted specific or subspecific name ends on "us" ---------- **15**
8h	(6b, 7b).	Submitted specific or subspecific name ends on "um" ---------- **16**
8i	(6b, 7b).	Submitted specific or subspecific name ends on "ii" ---------- **17**
8j	(6b, 7b).	Submitted specific or subspecific name ends on "i" ---------- **18**
8k	(6b, 7b).	Submitted specific or subspecific name ends on "a" ---------- **19**
8l	(6b, 7b).	Submitted specific or subspecific name ends on "e" ---------- **20**
8m	(6b, 7b).	Submitted specific or subspecific name ends on "ns" ---------- **21**
8 n	(6b, 7b).	Submitted specific or subspecific name ends on "as" ---------- **22**
8o	(6b, 7b).	Submitted specific or subspecific name ends on "ra" ---------- **23**

8p	(6b, 7b).	Submitted specific or subspecific name ends on "rum"	**24**
8q	(6b, 7b).	Submitted specific or subspecific name ends on none of the above suffixes	**26**
9a	(8a).	Combination of submitted or valid Genus and modified (sub)specific name with the last four characters replaced by "era" or "er" matches a valid name, a synonym, or a recorded misspelling	**25**
9b	(8a).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
10a	(8b).	Combination of submitted or valid Genus and modified (sub)specific name with the last three characters replaced by "erum" or "er" matches a valid name, a synonym, or a recorded misspelling	**25**
10b	(8b).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
11a	(8c).	Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "era" or "erum" matches a valid name, a synonym, or a recorded misspelling	**25**
11b	(8c).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
12a	(8d).	Combination of submitted or valid Genus and modified (sub)specific name with the last three characters replaced by "ius" matches a valid name, a synonym, or a recorded misspelling	**25**
12b	(8d).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
13a	(8e).	Combination of submitted or valid Genus and modified (sub)specific name with the last three characters replaced by "ior" matches a valid name, a synonym, or a recorded misspelling	**25**
13b	(8e).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
14a	(8f).	Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "e" matches a valid name, a synonym, or a recorded misspelling	**25**
14b	(8f).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
15a	(8g).	Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "a" or "um" or "ns" or "as" matches a valid name, a synonym, or a recorded misspelling	**25**
15b	(8g).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**
16a	(8h).	Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "a" or "us" matches a valid name, a synonym, or a recorded misspelling	**25**
16b	(8h).	Combination of submitted or valid Genus and modified (sub)specific name does not match	**26**

17a (8i). Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "i" or removed (deleted) matches a valid name, a synonym, or a recorded misspelling ---------- **25**

17b (8i). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

18a (8j). Combination of submitted or valid Genus and modified (sub)specific name with the last replaced by "ii" or removed (deleted) matches a valid name, a synonym, or a recorded misspelling ---------- **25**

18b (8j). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

19a (8k). Combination of submitted or valid Genus and modified (sub)specific name with the last character replaced by "us" or "um" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

19b (8k). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

20a (8l). Combination of submitted or valid Genus and modified (sub)specific name with the last character replaced by "is" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

20b (8l). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

21a (8m). Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "us" or "as" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

21b (8m). Combination of submitted or valid Genus and modified(sub)specific name does not match ---------- **26**

22a (8n). Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "ns" or "us" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

22b (8n). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

23a (8o). Combination of submitted or valid Genus and modified (sub)specific name with the last two characters replaced by "er" or "rum" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

23b (8o). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

24a (8p). Combination of submitted or valid Genus and modified (sub)specific name with the last three characters replaced by "er" or "ra" matches a valid name, a synonym, or a recorded misspelling ---------- **25**

24b (8p). Combination of submitted or valid Genus and modified (sub)specific name does not match ---------- **26**

25. **Action:** Add valid name, author and year, family, and reference to the submitted name ---------- **26**

26. **Action:** Proceed with next name in list. If end of list, end procedure. Continue with procedure described in Appendix 2.

APPENDIX 2

Algorithm for suggesting similarly spelled generic names for cases where no matching generic name had been found. Comments and possible genera are printed. The user can correct or replace the submitted generic names and rerun the procedure described in Appendix 1.

1a		Submitted generic name with last character replaced by one single wildcard matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
1b		Modified generic name does not match ---------- **2**
2a	(1b).	Submitted generic name with last character replaced by two single wildcards matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
2b	(1b).	Modified generic name does not match ---------- **3**
3a	(2b).	Length of submitted generic name more than two characters ---------- **4**
3b	(2b).	Length of submitted generic name not more than two characters ---------- **13**
4a	(3a).	Submitted generic name with last two characters replaced by two single wildcards matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
4b	(3a).	Modified generic name does not match ---------- **5**
5a	(4b).	Submitted generic name with last two characters replaced by one single wildcard matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
5b	(4b).	Modified generic name does not match ---------- **6**
6a	(5b).	Length of submitted generic name more than three characters ---------- **7**
6b	(5b).	Length of submitted generic name not more than three characters ---------- **13**
7a	(6a).	Submitted generic name with second and third character each replaced by a single wildcard matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
7b	(6a).	Modified generic name does not match ---------- **8**
8a	(7b).	Submitted generic name with second and third character together replaced by a single wildcard matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
8b	(7b).	Modified generic name does not match ---------- **9**
9a	(8b).	Submitted generic name with second and third character together replaced by three single wildcards matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
9b	(8b).	Modified generic name does not match ---------- **10**
10a	(9b).	Length of submitted generic name more than ten characters ---------- **11**
10b	(9b).	Length of submitted generic name more than five characters---------- **12**
10c	(9b).	Length of submitted generic name less than six characters---------- **13**

11a	(10a).	Submitted generic name with fourth and following characters replaced by a general wildcard and with the matching name of a length equal to the length of the submitted name +/- two characters matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
11b	(10a).	Modified generic name does not match ---------- **13**
12a	(10b).	Submitted generic name with fourth and following characters replaced by a general wildcard and with the matching name of a length equal to the length of the submitted name +/- one character matches one or more valid, synonymous, or misspelled generic name(s) ---------- **14**
12b	(10b).	Modified generic name does not match ---------- **13**
13		**Action:** Print submitted generic name with comment: "No matching or similarly spelled generic name could be found." Proceed with next name. If end of list, end procedure. Continue with procedure described in Appendix 3.
14		**Action:** Print submitted generic name and all matching or similarly spelled generic names, the status of the names, the corresponding valid generic name, and the family. Proceed with next name. If end of list, end procedure. Continue with procedure described in Appendix 3.

APPENDIX 3

Algorithm for suggesting similarly spelled specific or subspecific epithets from within the same family; the routine only considers submitted names that have a family but still no valid name assigned to them; comments and possible scientific names are printed. Based on the printout, the user can correct or replace submitted names and rerun the procedure described in Appendix 1.

1a		Length of submitted (sub)specific epithet equals two characters ---------- **2**
1b		Length of submitted (sub)specific epithet more than two characters ---------- **3**
2a	(1a).	Submitted (sub)specific epithet with second character replaced by a single wildcard matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
2b	(1a).	Modified sub(specific) epithet does not match ---------- **9**
3a	(1b).	Submitted (sub)specific epithet with second and third character each replaced by a single wildcard matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
3b	(1b).	Modified sub(specific) epithet does not match ---------- **4**
4a	(3b).	Submitted (sub)specific epithet with second and third character together replaced by a single wildcard matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
4b	(3b).	Modified sub(specific) epithet does not match ---------- **5**
5a	(4b).	Submitted (sub)specific epithet with second and third character together replaced by three single wildcards matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
5b	(4b).	Modified sub(specific) epithet does not match ---------- **6**
6a	(5b).	Length of submitted (sub)specific epithet more than 10 characters ---------- **7**
6b	(5b).	Length of submitted (sub)specific epithet more than 3 characters ---------- **8**
6c	(5b).	Length of submitted (sub)specific epithet less than 4 characters ---------- **9**
7a	(6a).	Submitted (sub)specific epithet with fourth and following characters replaced by a general wildcard and with the matching name of a length equal to the length of the submitted name +/- two characters matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
7b	(6a).	Modified (sub)specific epithet does not match ---------- **9**
8a	(6b).	Submitted (sub)specific epithet with fourth and following characters replaced by a general wildcard and with the matching name of a length equal to the length of the submitted name +/- one character matches one or more valid, synonymous, or misspelled (sub)specific epithet(s) within the same family ---------- **10**
8b	(6b).	Modified (sub)specific epithet does not match ---------- **9**

9 **Action:** Print submitted name and comment: "No matching or similarly spelled (sub)specific epithet could be found within this family." Proceed with next name. If end of list, end procedure.

10 **Action:** Print submitted name and all matching or similarly spelled names, the status of the names, the corresponding valid name, the reference, and the family.

Proceed with next name. If end of list, end procedure.