

The problem of hemihomonyms and the on-line hemihomonyms database (*HHDB*)

Alexey SHIPUNOV

Department of Biology, Minot State University, 500 University Ave. W, Minot, North Dakota 58707, USA.
<dactylorhiza@gmail.com>.

Abstract

Hemihomonyms (same nomina which are used for taxa from different nomenclature jurisdictions) are an overlooked but genuine nuisance in biological nomenclature. We compiled the first list of hemihomonyms for nomina in bacteriological, botanical and zoological nomenclatures and prepared an on-line database, the “Hemihomonym database” or *HHDB* (<<http://herba.msu.ru/shipunov/os/homonyms/index.php>>). *HHDB* now includes 1164 nomina, including 12 triple hemihomonyms. A simple suffix-based solution (like “*Oenanthe* (z)” for *Oenanthe* in zoology) could be used in case of hemihomonymy. More effort should be afforded towards the resolution of long-standing nomenclature confusing situations such as hemihomonymy, including regarding the nomina of higher taxa, nomina of intermediate ranks and ambiregual nomina.

Keywords: hemihomonyms, homonyms, biological nomenclature, databases

Introduction

If a scientific name or nomen is used for more than one species, genus or other taxon, this nomen is considered to be a homonym. The common opinion is that homonyms are invalid in biological nomenclature. However, historical development of biological taxonomy led to the establishment of different Codes of nomenclature. Homonyms are “illegal” within every Code (i.e., “incorrect” in botanical nomenclature, “invalid” in zoological nomenclature” or “inadequate” according to the terminology of Dubois 2011b), but what happens if the nomina in question are under the jurisdiction of different Codes? This situation is not regulated by any Rules and therefore the same nomina for different taxa are not homonyms in the strict sense. Starobogatov (1991) proposed the term “hemihomym” for such situations. Hemihomonyms (like the plant generic nomen *Oenanthe* and the bird generic nomen *Oenanthe*) are often considered as nomenclatural curiosities which probably was the right approach in previous centuries. *The International Code of Zoological Nomenclature* (Anonymous 1999) simply states that “*The name of an animal taxon identical with the name of a taxon which has never been treated as animal is not a homonym for the purposes of zoological nomenclature*”. However, contemporary large-scale databases and search engines revive the problem of hemihomonyms. The simple experiment with Google image search for *Oenanthe* will immediately show the problem: whereas scientific names are often considered to be unique identifiers, the hemihomonyms will spoil the result: the user will retrieve images for both the plant and the bird. Therefore, as long as hemihomonyms exist, and the result of such a search is not unambiguous, we cannot achieve the ultimate goal of nomenclature, i.e., a one-to-one relation between nomen and taxon. In large databases, hemihomonyms will not only hamper the effectiveness of the system, but could also be misleading. Computer-based tools do not

“understand” the jurisdictions of the Codes, so any database which contains nomina from different Codes will always be a potential “security hole”. Databases like *USDA Plant*, *IPNI* or *Index Fungorum* become now commonly used scientific tools, and without a hemihomonyms check, they may be a source of mistakes similar to well-known cases related with the use of office software (Zeeberg *et al.* 2004).

The level of hemihomonyms knowledge is, however, surprisingly small. Before this work, nobody knew the number of hemihomonyms or even its order of magnitude, and no published list of hemihomonyms did exist. This could have the same reason as the existence of hemihomonyms themselves: interdisciplinary researches employing many nomina following different Codes are still rare. Nevertheless, I am almost positive that, as these investigations will become more common in the near future, the problem of hemihomonyms should be solved as soon as possible.

It is worth mentioning here that Codes like those of nomenclature of bacteria (Lapage *et al.* 1992), animals (Anonymous 1999) and plants and fungi (McNeill *et al.* 2006) have slightly different approaches regarding “normal” homonyms. For example, the zoological Code distinguishes between primary and secondary homonyms, and also strictly regulates the use of similarly spelled nomina in some cases (“parahomonyms”). This Code regulates only nomina up to the level of superfamily, which means that nomina of higher taxa (e.g., orders and classes) may in theory be “legal” homonyms. Among different Codes, the bacteriological one has probably the most advanced approach saying that “*The name should not be a later homonym of a previously validly published name of an alga, bacterium, fungus, protozoon, or virus...*” This has, however, only limited consequences since other Codes do not do the same, and even the bacteriological Code does not consider the nomina of higher plants or animals.

The idea of making a single Code for all biological nomina (the “*BioCode*”) have a tough history. The most recent proposal (Greuter *et al.* 2011) does not deal with hemihomonyms, but if the *BioCode* was to be finally implemented, clarification of hemihomonyms would become unavoidable. Instead of a published text, the future *BioCode* could also become an overlaying Web service maintaining all Codes’ Rules together (Shipunov *et al.* 2009), and in this case the service will need a hemihomonyms database.

Materials and methods

The starting list of nomina was compiled from different and unequal sources: the *Catalogue of Life 2008 Checklist* (Bisby *et al.* 2008), the *Wikispecies List of Valid Homonyms* (Anonymous 2009b), the *Taxacom archives* (Anonymous 2009a) and a variety of personal communications and individual submissions directly to me.

All lists were normalized: converted to comma separated text format with standardized columns (nomen, ancestry, rank, source and ID) and merged. For merging, a specific R (Anonymous 2011) program was written. The database itself uses flat text file as a data source, PHP interface to the flat list plus DataTables JavaScript application (<<http://datatables.net/index>>), which produces the spreadsheet-like output interface. There is also a simple submission page which accepts comments from users, and rudimentary API allowing the checking of single nomina with two possible answers (“yes” if the nomen is in the database, and “no” if it is not).

We also found that several large name databases like *CU*STAR* (<<http://starcentral.mbl.edu>>) or *GNI* (<<http://www.globalnames.org/GNI>>) were not simple to analyze, probably due to the vast amount of misspellings listed. These databases were not used for the primary list.

Nevertheless, since not all reliable nomen sources were analyzed, we cannot be sure that our database is complete. If a nomen is not in the database, there still is a possibility that it may be a hemihomonym. Another approach to the hemihomonymy problem is now being implemented in *IRMNG* (<<http://www.cmar.csiro.au/datacentre/irmng/homonyms.htm>>).

Results

As for 1st August 2011, the “Hemihomonym database” (*HHDB*; <<http://herba.msu.ru/shipunov/os/homonyms/index.php>>; mirror: <<http://ashipunov.info/shipunov/os/homonyms/index.php>>) has 1164 nomina. Most hemihomonyms are results of clashes between the botanical and zoological Codes (1113 nomina, i.e. 96 %), and much less are between bacteriological and botanical nomina and between zoological and bacteriological nomina (8 nomina, i.e. 1 % and 31 nomina, i.e. 3 %, respectively). Twelve nomina (1 %) are triple hemihomonyms.

The shortened variant of *HHDB* including only nomina is given here in Table 1. On the Web site, it is also possible to retrieve so-called “ancestry”, i.e., the data about the taxonomical position of the nominal taxon.

Discussion

Since the hemihomonym database is now established, it is possible to identify the situations of hemihomonymy between the three Codes considered. In order to avoid ambiguity, I propose here that whenever a nomen is a hemihomonym, it should be followed with a postfix “(b)”, “(p)” or “(z)” for nomina covered by the bacteriological, botanical and zoological Codes of nomenclature, respectively. To check the possibility of being a hemihomonym, one could use Table 1, the main Web page of *HHDB*, or the proposed *API* (<<http://herba.msu.ru/shipunov/os/homonyms/index.php>>).

A few additional situations which may cause ambiguity in nomenclatural research are not yet resolved:

(1) Nomina of higher taxa. This is an “Achilles heel” of modern nomenclature. Multiple efforts were done towards a resolving of higher taxa problem, rule- or database-based (Reveal 2008; Shipunov 2009; Kluge, 2010 and his earlier Russian publications; Dubois, 2011a and his other publications) but the current Codes still almost completely ignore the problem.

(2) Nomina of “intermediate levels” of nomenclatural hierarchy. It is common for most databases and sometimes even monographs to skip tribal, section or other levels intermediate between the “main” ones. This is often considered as a fundamental problem for all rank-based nomenclature (R. Olmstead, pers. comm.). As a result, we simply do not know how many homonyms are hidden here.

(3) The nomen could be an “unstable” hemihomonym if it is unclear which Code should be used for it (“ambiregnal names”; see Patterson 1991). Many protists’ nomina are ambiregnal.

(4) Three other Codes of nomenclature exist which may also contain hemihomonyms: the Codes for viruses (Anonymous 2002) and cultivated plants (Brickell *et al.* 2009; it does not control nomenclature at the generic level), and the *PhyloCode* (Cantino & de Queiroz 2010; it is not an officially recognized Code). The nomina recognized under these Codes may later be added to the *HHDB*.

Acknowledgements

I am very grateful to David Patterson, Patrick Leary and Dmitry Mozzherin (EOL) for their invaluable help in the creation of HHDB and to Alain Dubois for his great assistance in the manuscript preparation.

TABLE 1. Hemihomonyms from realms of three main codes of nomenclature (according to HHDB, last accessed 1st August 2011).

TABLE 1. (continued).

TABLE 1. (continued).

Oceanites	Oceanites
Odontobius	Odontophorus
Odontostoma	Odontostoma
Odontura	Odontura
Oenanthie	Oenanthe
Oenia	Oenia
Olea	Olea
Olfersia	Olfersia
Oligochaeta	Oligochaeta
Olyra	Olyra
Omania	Omania
Ommatophora	Ommatophora
Oncotylus	Oncotylus
Oncylophora	Oncylophora
Operculta	Operculta
Ophidion	Ophidion
Ophiocephalus	Ophiocephalus
Orania	Orania
Orchideum	Orchideum
Oristas	Oristas
Orgia	Orgia
Oriocnemis	Oriocnemis
Oriola	Oriola
Oriina	Oriina
Ormosia	Ormosia
Oroya	Oroya
Orthosia	Orthosia
Oryctes	Oryctes
Osa	Osa
Ourococcus	Ourococcus
Ovidia	Ovidia
Owenia	Owenia
Oxyphyllum	Oxyphyllum
Oxyrrhynchus	Oxyrrhynchus
Ozophora	Ozophora
Padiellina	Padiellina
Pachystylus	Pachystylus
Palmeira	Palmeira
Palminota	Palminota
Palumbina	Palumbina
Pandiesia	Pandiesia
Pandora	Pandora
Panus	Panus
Papillaria	Papillaria
Paracephala	Paracephala
Paradymonia	Paradymonia
Paragonia	Paragonia
Paragorgia	Paragorgia
Parasenella	Parasenella
Parisia	Parisia
Parikia	Parikia
Paritaria	Paritaria
Pamela	Pamela
Parthenope	Parthenope
Passerina	Passerina
Patellina	Patellina
Paulinella	Paulinella
Pavillia	Pavillia
Pediaria	Pediaria
Pentilia	Pentilia
Pelaea	Pelaea
Pennella	Pennella
Pentacheila	Pentacheila
Pentagramma	Pentagramma
Pentodon	Pentodon
Pembaea	Pembaea
Pentraphaga	Pentraphaga
Perigrapta	Perigrapta
Penicilla	Penicilla

TABLE 1. (continued).

<i>Trichoneura</i>	<i>Trichopeplatum</i>	<i>Trichopteryx</i>
<i>Trichopterum</i>	<i>Tridens</i>	<i>Tridens</i>
<i>Trigonaspis</i>	<i>Trigonaspis</i>	<i>Trigonaspis</i>
<i>Trigonalium</i>	<i>Trigonalium</i>	<i>Trigonalium</i>
<i>Trimenia</i>	<i>Trimenia</i>	<i>Trimenia</i>
<i>Trinacia</i>	<i>Trinacia</i>	<i>Trinacia</i>
<i>Trinaria</i>	<i>Trinaria</i>	<i>Trinaria</i>
<i>Tricodia</i>	<i>Tricodia</i>	<i>Tricodia</i>
<i>Triphora</i>	<i>Triphora</i>	<i>Triphora</i>
<i>Trisira</i>	<i>Trisira</i>	<i>Trisira</i>
<i>Trochus</i>	<i>Trochus</i>	<i>Trochus</i>
<i>Trolliella</i>	<i>Trolliella</i>	<i>Trolliella</i>
<i>Tropidia</i>	<i>Tropidia</i>	<i>Tropidia</i>
<i>Turcataella</i>	<i>Turcataella</i>	<i>Turcataella</i>
<i>Tubaginia</i>	<i>Tubaginia</i>	<i>Tubaginia</i>
<i>Turbinaria</i>	<i>Turbinaria</i>	<i>Turbinaria</i>
<i>Tydemania</i>	<i>Tydemania</i>	<i>Tydemania</i>
<i>Tylotus</i>	<i>Tylotus</i>	<i>Tylotus</i>
<i>Typha</i>	<i>Typha</i>	<i>Typha</i>
<i>Ufua</i>	<i>Ufua</i>	<i>Ufua</i>
<i>Uvaria</i>	<i>Uvaria</i>	<i>Uvaria</i>
<i>Urbania</i>	<i>Urbania</i>	<i>Urbania</i>
<i>Uceolina</i>	<i>Uceolina</i>	<i>Uceolina</i>
<i>Udon</i>	<i>Udon</i>	<i>Udon</i>
<i>Urenoma</i>	<i>Urenoma</i>	<i>Urenoma</i>
<i>Urvillea</i>	<i>Urvillea</i>	<i>Urvillea</i>
<i>Vadiva</i>	<i>Vadiva</i>	<i>Vadiva</i>
<i>Valonia</i>	<i>Valonia</i>	<i>Valonia</i>
<i>Vanhoefenia</i>	<i>Vanhoefenia</i>	<i>Vanhoefenia</i>
<i>Venturia</i>	<i>Venturia</i>	<i>Venturia</i>
<i>Vermicularia</i>	<i>Vermicularia</i>	<i>Vermicularia</i>
<i>Vescularia</i>	<i>Vescularia</i>	<i>Vescularia</i>
<i>Victoria</i>	<i>Victoria</i>	<i>Victoria</i>
<i>Vitalia</i>	<i>Vitalia</i>	<i>Vitalia</i>
<i>Viguierella</i>	<i>Viguierella</i>	<i>Viguierella</i>
<i>Vicia</i>	<i>Vicia</i>	<i>Vicia</i>
<i>Virginia</i>	<i>Virginia</i>	<i>Virginia</i>
<i>Vittia</i>	<i>Vittia</i>	<i>Vittia</i>
<i>Vossia</i>	<i>Vossia</i>	<i>Vossia</i>
<i>Waltonia</i>	<i>Waltonia</i>	<i>Waltonia</i>
<i>Wardia</i>	<i>Wardia</i>	<i>Wardia</i>
<i>Wenyrigia</i>	<i>Wenyrigia</i>	<i>Wenyrigia</i>
<i>Werneria</i>	<i>Werneria</i>	<i>Werneria</i>
<i>Weymouthia</i>	<i>Weymouthia</i>	<i>Weymouthia</i>
<i>Willardia</i>	<i>Willardia</i>	<i>Willardia</i>
<i>Williamsonia</i>	<i>Williamsonia</i>	<i>Williamsonia</i>
<i>Wilsonia</i>	<i>Wilsonia</i>	<i>Wilsonia</i>
<i>Womersleyella</i>	<i>Womersleyella</i>	<i>Womersleyella</i>
<i>Woodfordia</i>	<i>Woodfordia</i>	<i>Woodfordia</i>
<i>Woodsia</i>	<i>Woodsia</i>	<i>Woodsia</i>
<i>Xenococcus</i>	<i>Xenococcus</i>	<i>Xenococcus</i>
<i>Xenus</i>	<i>Xenus</i>	<i>Xenus</i>
<i>Xeropityllum</i>	<i>Xeropityllum</i>	<i>Xeropityllum</i>
<i>Xizangia</i>	<i>Xizangia</i>	<i>Xizangia</i>
<i>Xylococcus</i>	<i>Xylococcus</i>	<i>Xylococcus</i>
<i>Xylophylus</i>	<i>Xylophylus</i>	<i>Xylophylus</i>
<i>Xylothiza</i>	<i>Xylothiza</i>	<i>Xylothiza</i>
<i>Yersinia</i>	<i>Yersinia</i>	<i>Yersinia</i>
<i>Yestia</i>	<i>Yestia</i>	<i>Yestia</i>
<i>Zenkerella</i>	<i>Zenkerella</i>	<i>Zenkerella</i>
<i>Zeus</i>	<i>Zeus</i>	<i>Zeus</i>
<i>Zilia</i>	<i>Zilia</i>	<i>Zilia</i>
<i>Zonaria</i>	<i>Zonaria</i>	<i>Zonaria</i>

References

- Anonymous [International Commission on Zoological Nomenclature] (1999) *International code of zoological nomenclature*. Fourth edition. London (International Trust for zoological Nomenclature): i–xxix + 1–306.
- Anonymous [International Committee on Taxonomy of Viruses] (2002) *The international code of virus classification and nomenclature*. <http://www.ictvonline.org/codeOfVirusClassification_2002.asp>. [Accessed 1st January 2011].
- Anonymous [Taxacom] (2009a) *The Taxacom archives*. <<http://mailman.nhm.ku.edu/pipermail/taxacom/>>. [Accessed 1st January 2009].
- Anonymous [Wikispecies] (2009b) *List of valid homonyms*. <http://species.wikimedia.org/wiki/List_of_valid_homonyms>. [Accessed 1st January 2009].
- Anonymous [R Development Core Team] (2011) *R: a language and environment for statistical computing*. Vienna, Austria (R Foundation for Statistical Computing). <<http://www.r-project.org>>. [Accessed 1st January 2009].
- Bisby, F. A., Roskov, Y. R., Orrell, T. M., Nicolson, D., Paglinawan, L. E., Bailly, N., Kirk, P. M., Bourgoin, T. & van Hertum, J. (ed.) (2008) *Species 2000 & ITIS Catalogue of Life: 2008 Annual Checklist*. Reading, U.K. (Species 2000), CD-ROM.
- Brickell, C. D., Alexander, C., David, J. C., Hetterscheid, W. L. A., Leslie, A. C., Malécot, V., Xiaobai Jin & Cubey, J. J. (ed.) (2009) International code of nomenclature for cultivated plants. Eighth edition. *Scripta Horticulturae*, Leuven (International Society for Horticultural Science), **10**: 1–204.
- Cantino, P. D. & de Queiroz, K. (2010) *International code of phylogenetic nomenclature. Version 4c*. <<http://www.ohio.edu/phylocode/preface.html>>. [Accessed 1st January 2011].
- Dubois, A. (2011a) The *International Code of Zoological Nomenclature* must be drastically improved before it is too late. *Bionomina*, **2**: 1–104.
- Dubois, A. (2011b) A zoologist viewpoint on the *Draft BioCode*. *Bionomina*, **3**: 45–62.
- Greuter, W., Garrity, G., Hawksworth, D. L., Jahn, R., Kirk, P. M., Knapp, S., McNeill, S., Michel, E., Patterson, D. J., Pyle, R. & Tindall, B. J. (2011) Draft BioCode (2011). Principles and Rules regarding the naming of organisms. New draft, revised in November 2010. *Bionomina*, **3**: 26–44.
- Kluge, N. J. (2010) Circumscriptional names of higher taxa in Hexapoda. *Bionomina*, **1**: 15–55.
- Lapage, S. P., Sneath, P. H. A., Lessel, E. F., Skerman, V. B. D., Seeliger, H. P. R. & Clark, W.A. (ed.) (1992) *International code of nomenclature of Bacteria (Bacteriological code 1990 revision)*. Washington, D.C. (American Society of Microbiology): i–xlvi + 1–189.
- McNeill, J., Barrie, F. R., Burdet, H. M., Demoulin, V., Hawksworth, D. L., Marhold, K., Nicolson, D. H., Prado, J., Silva, P. C., Skog, J. E., Wiersema, J. H. & Turland, N. J., (ed.) (2006) *International code of botanical nomenclature (Vienna Code) adopted by the Seventeenth International Botanical Congress, Vienna, Austria, July 2005*. Königstein (Koeltz): i–viii + 1–568.
- Patterson, D. J. & Larsen, J. (1991) Nomenclatural problems with protists. In: D. L. Hawksworth (ed.), *Improving the stability of names: needs and opinions*, Königstein (Koeltz Scientific Books): 197–208.
- Reveal, J. L. (2008) *Index nominum supragenericorum plantarum vascularium*. <<http://www.plantsystematics.org/reveal/pbio/WWW/supragen.html>>. [Accessed 1st January 2009].
- Shipunov, A. B. (2009) *Systema Naturae* or the outline of living world classification. *Protistology*, **6** (1): 3–13.
- Shipunov, A. B., Mozzherin D. & Patterson, D. J. (2009) Towards building code compliance within biodiversity informatics. In: *The Systematics Association 7th Biennial Meeting, Leiden, August 2009*. <http://ashipunov.info/shipunov/author/shipunov_eusyst2009.pdf>. [Accessed 1st December 2011].
- Starobogatov, Y. I. (1991) Problems in the nomenclature of higher taxonomic categories. *Bulletin of zoological Nomenclature*, **48** (1): 6–18.
- Zeeberg, B. R., Riss, J., Kane, D. W., Bussey, K. J., Uchio, E., Linehan, W. M., Barrett, J. C. & Weinstein, J. N. (2004) Mistaken identifiers: gene name errors can be introduced inadvertently when using Excel in bioinformatics. *BMC Bioinformatics*, **5**: 80.