

Grey nomenclature needs rules

ALESSANDRO MINELLI

Department of Biology, University of Padova, Via Ugo Bassi 58 B, I 35131 Padova, Italy; E-mail: alessandro.minelli@unipd.it

Received 15 November 2016 | Accepted 18 January 2017 | Published online 20 January 2017.

Abstract

Intentional departures from Linnaean nomenclature are common, and are due to different causes, including deliberate refusal to obey a specific rule of the Code, full rejection of Linnaean nomenclature in favour of an alternative system, use of formulae for special kinds of organisms and especially the use of informal names, or formulae, for kinds of organisms provisionally recognized as corresponding to still undescribed species, whose actual description and naming are reserved for a subsequent study. Special attention should be paid to the mixed usage of names and formulae in the BOLD and GenBank databases. Save for BINs in the BOLD databases, that are an excellent example of how to create and use non-Linnaean formulae for segments of biological diversity, the other kinds of grey nomenclature are seriously faulty in three respects: first, the lack of rules for the creation and usage of these names or formulae precludes unambiguous understanding of what the name or formula is for; second, it is often unclear, especially in database entries, if the alphanumeric specifiers added to a genus name, or even to a Linnaean binomen, are intended to provide unambiguous labelling of a single specimen, or of a taxon, or both; third, and most important, because of the subjectivity of the criteria according to which these non-Linnaean names are created and used, it is generally impossible to compare them across studies, or databases. When using names (or formulae) other than Linnaean binomens, it should be always made clear what the label is intended for. This requires an international agreement on a small set of simple, clear principles, fixing the standard format for each of the objects we need to distinguish, starting with the definition of a number of standard formats for the different kinds of objects (or hypotheses) we need to label, such as individual, species, undescribed new species.

Key words: zoological nomenclature; BOLD; GenBank; grey nomenclature.

Non-Linnaean names and other kinds of grey nomenclature in current use

It is generally taken for granted that animal names used in the scientific literature and in professional databases are in agreement with the rules of the International Code of Zoological Nomenclature (International Commission on Zoological Nomenclature 1999; hereinafter, the Code), except for occasional exceptions due to oversight or sloppiness. Intentional departures from Linnaean nomenclature, however, are common, and are due to different causes, the most important of which are listed and briefly discussed in the following sections.

Infrageneric or infraspecific ranks excluded or not strictly ruled by the Code

Years ago, Tyler (1991) provided the following list of terms occasionally or regularly used for infrageneric ranks other than subgenus, species and subspecies: chromosomal race, clone, cryptic species, cytotype,

ecospecies, forma, incipient species, infra-genus, klepton, morph, morphospecies, morphotype, race, sibling species, species aggregate, species assemblage, species clade, species complex, species group, superspecies, synklepton, variant, variety. The Code, however, does not accept names for taxa at all the many infrageneric levels that researchers sometimes need to recognize. The only opportunities offered by the Code for coping with some of these needs are found in the following rule:

Art. 6.2. Names of aggregates of species or subspecies. A specific name may be added in parentheses after the genus-group name, or be interpolated in parentheses between the genus-group name and the specific name, to denote an aggregate of species within a genus-group taxon; and a subspecific name may be interpolated in parentheses between the specific and subspecific names to denote an aggregate of subspecies within a species.

together with a recommendation that allows for some degree of freedom in the application of this rule:

Recommendation 6B. Taxonomic meaning of interpolated names. An author who wishes to denote an aggregate at either of the additional taxonomic levels mentioned in Article 6.2 should place a term to indicate the taxonomic meaning of the aggregate in the same parentheses as its interpolated species-group name on the first occasion that the notation is used in any work.

The name of a superspecies¹, when recognized, is thus usually placed in parenthesis between the generic name and the specific epithet, that is, in the same position as the name of a subgenus. An example (from Seraphin 2012): *Agrias (amydon) amydonius klugi*. It is not clear why Dubois and Raffaëlli (2009) put instead in parentheses the names of the species they recognize within an infrageneric group for which they use the term “supraspecies”, as for the newt genus *Ichthyosaura* within which these authors recognize a supraspecies *Ichthyosaura alpestris* with *alpestris* and *reiseri* treated as species, with several subspecies each. The latter take names such as *Ichthyosaura alpestris (alpestris) apuana* and *I. alpestris (reiseri) carpathica*.

At any rate, even in the case of small orthographic deviations from the use prescribed by the Code, what is intended by these names can be easily understood by reference to the traditional framework of Linnaean nomenclature.

Deliberate refusal to obey a specific rule of the Code

Some articles of the Code do not meet with universal favour and a few of them are intentionally violated by a number of zoologists. This is the case of the following rule:

Art. 31.2. Agreement in gender. A species-group name, if it is or ends in a Latin or latinized adjective or participle in the nominative singular, must agree in gender with the generic name with which it is at any time combined.

Many zoologists regard Art. 31.2 as a leftover from the past, when most naturalists had familiarity with Latin, a condition progressively restricted to a smaller and smaller percentage of zoologists. I will not discuss here comparatively arguments pro and against this rule; I will only mention, that gender agreement is systematically rejected by a sensible number of lepidopterists – not a symptom of inadequate familiarity with the Code, but the consequence of a deliberate rejection of this article. For example, in their list of the gelechiid Lepidoptera of the Italian fauna, Huemer & Karsholt (1995) listed four species of the genus *Isophrictis* with specific epithets in feminine form and another two in masculine form, e.g. *Isophrictis kefersteiniellus* and *Isophrictis anthemidella*, while the first of these species had been originally described as *Ypsolophus kefersteiniellus*, the latter as *Cleodora anthemidella*.² Let’s remark here, marginally, that one of the benefits of gender agreement is that you always (usually) know the ending of the specific name.

¹ Mayr (1931) introduced the term superspecies, corresponding to Rensch’s (1928) *Artenkreis*, for a set of closely related allopatric species (the semispecies of Mayr (1940)).

² Due to the nature of this article, I will not give author and name in association with Linnaean binomens, although I strongly urge for their specification in ordinary zoological literature.

Full rejection of the Linnaean nomenclature, in favour of an alternative system

Very different is the case of names created in agreement with a set of rules presented as a radical alternative to the Code. The best example is offered by the names of the taxa called LITUs or Least Inclusive Taxonomic Units that Pleijel & Rouse (1999) proposed as terminal taxa within the - at the time - still tentatively formulated PhyloCode – essentially, a nomenclature without ranks, intended as the optimal way to name monophyletic clades of any degree of inclusiveness. Due to several difficulties, practical as well as conceptual (the debatable requirement of monophyly for the taxa at the level of species, or equivalent), the official version of the PhyloCode eventually issued (Cantino & De Queiroz 2010) does not include rules for creating names of terminal (species) taxa. In the meantime, however, several names have been introduced for LITUs, e.g. by Pleijel & Rouse (2000) and by Pleijel (2000). These names are uninominal, such as *Lyonsi*, *Xiamenensis* and *Bidentata*, the names of three polychaete LITUs.

Formulae for special kinds of organisms

Within traditional Linnaean nomenclature it has proven difficult or plainly impossible, even in the presence of obvious phenotypic differences, to accommodate animals (see Table 1) reproducing by metasexual processes. This collective term (Scali *et al.* 2003) is applied to three different mechanisms that represent deviations from the typical bisexual system. One of these mechanisms is *gynogenesis*: here, the role of the sperm cell is reduced to the activation of the female gamete, that is, to stimulate it to complete meiosis and subsequently start embryogenesis, but the male gamete does not contribute its nucleus to the offspring. In *androgenesis*, to the contrary, the female nucleus is discarded and only the sperm cell's nucleus, now included within the egg's cytoplasm, provides genetic information to the offspring. A third metagenetic mechanism of reproduction is *hybridogenesis*, in which sperm cells produced by males of species A fertilize the eggs of a closely related species B, and both the maternal and the paternal genomes are retained in all the (hybrid) somatic cells of the offspring; however, during the gametogenesis of the hybrid one of the parental genomes is discarded; as a consequence, all gametes produced by the hybrid carry a haploid copy of one of the two genomes (say, A) and the hybrid survives through the generations only by continuing hybridization between the hybrid itself and the parental species (B) whose genome is discarded during the hybrid's gametogenesis.

The more or less overt hybrid origin of metagenetic animals is not, per se, a condition excluding them from the application of the Code. According to Art. 1.3, "Excluded from the provisions of the Code are names proposed [...] for hybrid specimens as such," but in the case of metagenetic animals the hybrid (or quasi-hybrid) condition is not confined to individual specimens, but extend to whole lineages, thus suggesting that Art. 17.2. will apply, according to which "The availability of a name is not affected even if [...] it is applied to a taxon known, or later found, to be of hybrid origin." Problems are due instead to the widespread desire to have their hybrid origin somehow reflected in their names. Opinions and operational solutions are different (e.g., Lowcock *et al.* 1987), sometimes including suggestions for a system of nomenclature external, or parallel, to the Code. For example, Schultz (1969) suggested the use of hyphens between the names of the parental species and (in the case of polyploids) the prefix of a number in front of the species name of the parent contributing a multiple set of chromosomes. Following this approach, Cook & Gorham (1979) renamed *Ambystoma 2laterale-jeffersonianum* a hybrid newt previously known as *Ambystoma tremblayi* and *A. laterale-2jeffersonianum* the hybrid that had been already named *A. platineum*.

Informal names, or formulae, for kinds of organisms provisionally recognized as corresponding to still undescribed species, whose actual description and naming are however reserved for a subsequent study

Papers containing the simultaneous description of large numbers of closely related (or at least congeneric) new species are not confined to the old literature, where huge numbers of putatively new taxa were diagnosed in very few words and described in a couple of sentences. Excellent recent examples of publications in which a great many closely related species are described based on both morphological and molecular evidence are Riedel *et al.*'s (2013, 2014) articles on the weevils of the genus *Trigonopterus*. Previous to the most recent researches, this genus included 91 described species ranging from Sumatra to Samoa and from the Philippines to New Caledonia. Of these, 50 species of *Trigonopterus* had been described from New Guinea, the centre of the genus's diversity. But new targeted samplings in seven localities across

Table 1. Metasexual animals and their reproductive mechanisms (data after Tilquin 2014, rearranged and corrected).

'Rotifera' Monogononta	
<i>Keratella quadrata</i>	Gynogenesis
<i>Brachionus urceolaris</i>	
Platyhelminthes	
<i>Schmidtea polychroa</i>	Gynogenesis
<i>Dugesia benazzii</i>	
<i>Polycelis nigra</i>	
<i>Polycelis tenuis</i>	
Bivalvia	
<i>Corbicula</i> spp.	Androgenesis
Acarina	
<i>Dermanyssus gallinae</i>	Gynogenesis
Phytoseiidae, many spp.	
Collembola	
<i>Onychiurus procampatus</i>	Gynogenesis
Phasmatodea	
<i>Bacillus rossi-grandii benazzii</i>	Gynogenesis
<i>Bacillus rossius-grandii</i>	Hybridogenesis, androgenesis
Homoptera	
<i>Muellerianella fairmairei-brevipennis</i>	Gynogenesis
<i>Ribautodelphax pungens</i>	
Lepidoptera	
<i>Alsophila pomataria</i>	Gynogenesis
<i>Luffia lapidella</i>	
Coleoptera	
<i>Ptinus clavipes-mobilis</i>	Gynogenesis
<i>Ips</i> spp.	
Hymenoptera	
<i>Nasonia vitripennis</i>	Gynogenesis
<i>Apis mellifera</i>	Androgenesis
<i>Wasmannia auropunctata</i>	
<i>Vollenhovia emeryi</i>	
Osteichthyes	
<i>Poeciliopsis</i> spp.	Gynogenesis
<i>Poecilia formosa</i> complex	
<i>Menidia clarkhubbsi</i> complex	
<i>Cobitis</i> cplx	
<i>Carassius auratus gibelio</i> + <i>C. a. langsdorfi</i>	
<i>Tropidophoxinellus alburnoides</i> complex	
<i>Misgurnus anguillicaudatus</i>	
<i>Phoxinus eos-neogaeus</i> complex	Gynogenesis, hybridogenesis
<i>Poeciliopsis monacha</i> -sp	Hybridogenesis
<i>Tropidophoxinellus alburnoides</i> complex	
<i>Hypseleotris</i> spp.	
Amphibia	
<i>Rana klepton esculenta</i>	Hybridogenesis, rarely gynogenesis
<i>Ambystoma</i> spp.	Gynogenesis

New Guinea have resulted in the recognition of 279 *Trigonopterus* species, most of which new to science; of these, a first set of 101 species have been described by Riedel *et al.* (2013). Another 98 new species of *Trigonopterus* have been described in a paper (Riedel *et al.* 2014) devoted to materials recently collected in Indonesia (Sumatra, Java, Bali, Palawan, Lombok, Sumbawa, Flores), a large area from where only one species of *Trigonopterus* was previously known.

In an increasing number of instances, however, articles in which the hypothesis of the occurrence of a diversity of species-level taxa is put forward do not include formal description and naming of these (cf. Minelli 2015). A few examples follow.

Barcoding of the biting midges (Ceratopogonidae) of the county of Finnmark in northern Norway suggests that out of a total of 54 recorded species, 14 are likely to be new to science (Stur & Borkent 2014), but none of these has been actually described and named. The same goes for the 33 ‘provisional species’ recognized within a clade of amphipods living in desert spring of the southern Great Basin of California and Nevada, USA, hitherto referred to the one species, *Hyaella azteca* (Witt *et al.* 2006).

Modern revisitation based on molecular evidence and sophisticated methods of species delimitation often reveal an unsuspected cryptic diversity even within what were hitherto regarded as taxonomically unproblematic, and even popular, species. Raupach *et al.* (2014) have recently tested the efficiency of DNA barcoding for the Heteroptera of Central Europe. Their study has provided evidence for the putative existence of cryptic species, e.g. within the anthocorid *Orius niger*, but no descriptions or names are provided for the would-be new taxa.

In a forthcoming article, Martinsson & Erséus (2017) report on their study of an extensive sample of earthworms corresponding to the common European morphospecies *Lumbricus rubellus*. Based on one nuclear (H3) and one mitochondrial (COI) marker and using a Bayesian multi-locus species delimitation method, as well as single gene haplotype networks and gene trees, these authors identify seven well supported cryptic species; however, they do not describe them formally, but provisionally refer to them as *Lumbricus rubellus* A, B, G, H, J, K and M.

Again, based on the mitochondrial cytochrome-*b* gene, several new species of the bat genus *Miniopterus* have been recently described from Madagascar and the neighbouring Comoros archipelago, but at least seven out of the 18 species-level taxa recognized in the most recent study still require formal taxonomic treatment (Christidis *et al.* 2014).

In another study (Bittencourt-Silva *et al.* 2016) the new species-level taxa identified by molecular markers in the *Notophryne* toads of the Afromontane regions of Malawi and Mozambique have been simply labelled by adding geographical tags to the genus name, thus *Notophryne* Inago, *N.* Namuli, *N.* Pese, *N.* Ribáuè, *N.* Taratibu.

Table 2. Cryptic diversity discovered within some polychaete morphospecies by recent molecular investigations (data compiled from Nygren, 2014, Table S1).

genus	number of species	
	inferred from molecular studies	currently available species-level names
<i>Archinome</i>	5	<i>A. jasoni</i> , <i>A. tethyana</i> , <i>A. levinae</i> , <i>A. rosacea</i> , <i>A. storchi</i>
<i>Branchiomma</i>	11	<i>B. spp.</i>
<i>Capitella</i>	12+	<i>C. capitata</i>
<i>Eumida</i>	11	<i>Eu. sanguinea</i>
<i>Harmothoe</i>	6	<i>H. imbricata</i>
<i>Leitoscoloplos</i>	5	<i>L. pugettensis</i>
<i>Marenzelleria</i>	5	<i>M. viridis</i> , <i>M. bastropi</i> , <i>M. neglecta</i> , <i>M. wireni</i> , <i>M. arctia</i>
<i>Marphysa</i>	5	<i>M. sanguinea</i>
<i>Ophryotrocha</i>	14	<i>O. labronica</i>
<i>Owenia</i>	5	<i>O. fusiformis</i>
<i>Palola</i>	16	<i>P. spp.</i>
<i>Sabellastarte</i>	7	<i>S. spp.</i>
<i>Scoloplos</i>	5–6	<i>S. armiger</i>
<i>Syllis</i>	5	<i>S. alternata</i>

In polychaetes, cryptic species crop up with virtually every accurate study. The detailed review published by Nygren (2014) includes several dozen examples, of which only the most conspicuous ones (those with ≥ 5 cryptic species inferred to be present within a taxon currently treated as a single species) are listed in Table 2. The taxonomic complexity revealed by this study is probably nothing more than the tip of a huge iceberg of species diversity in the annelids. Most of the cryptic diversity discovered to date in polychaetes is still formally undescribed, one of the few exceptions being the five species of *Archinome* listed in the Table.

In terms of nomenclature, the treatment of the putative new taxa is far from uniform, even within one paper. For example, Blagoev *et al.* (2016), while listing the new species discovered among the Canadian spiders diagnosed by barcoding, fix some of these at the genus level only, while others are provisionally named with reference to the closest named species. Examples are ‘*Agyneta* sp. 1GAB’ and ‘*Agyneta darrelli* cf.’, respectively.³

A revision of the hydroids of the family Aglaopheniidae, based on different molecular markers, revealed the presence of 36 species-level taxa, 30 of which can be unequivocally referred to previously named Linnaean species. Interestingly, for the remaining six Postaire *et al.* (2016) used three different kinds of notation: (a) *Aglaophenia* sp. 1, *Aglaophenia* sp. 2, *Lytocarpia* sp. 1; (b) *Macrorhynchia* nov. sp.; (c) *Macrorhynchia phoenicea* morpho-type A; *Macrorhynchia phoenicea* morpho-type B.

In other papers, authors do not say clearly which samples among those they have compared actually correspond, in their opinion, to still undescribed taxa. For example, in discussing phylogenetic relationships among the Philippine sun skinks of the genus *Eutropis*, Barley *et al.* (2013) list some of them with Linnaean binomens or trinomens (*E. indepressa*, *E. cumingi*, *E. bontocensis*, *E. multicarinata multicarinata*, *E. m. borealis*), others instead as geographically labelled taxon without species epithet (*Eutropis* sp. Paulau) and still others just as clades without any reference to formally named taxa (simply referred to as clade A through clade G).

Harbach’s (2004) conspectus of the mosquito species of the genus *Anopheles* included 444 formally named and 40 provisionally (and informally) designated species, with the intent “to aid researchers and students who are interested in analysing species relationships, making group comparisons and testing phylogenetic hypotheses” irrespective of the fact that these entities have been formally described and named, or not. Here are examples of the far from uniform treatment of those *Anopheles* species lacking a code-compliant Linnaean name. Harbach included under the binomen *Anopheles subpictus* four species simply listed as species A, B, C and D; remarking that, based on the banding patterns of polytene chromosomes, Sarala *et al.* (1994) had recognized three species within *A. fluviatilis*, provisionally listed as species S, T and U; and reported that “*Hunt & Coetsee* (1991) provided cytogenetic evidence for two species of *A. marshallii* [sic]”. Unclear is the taxonomic status of what are described as multiple karyotypic forms of the same (named) species, such as forms A and B of each of the following taxa: *Anopheles argyropus*, *A. sinesi*, *A. crawfordi*, *A. nigerrimus*, *A. jamesii* and *A. vagus*, while other clusters of forms are still more complex, each of them including three karyotypic forms (A, B and C) such as in *A. aconitus*, *A. karwari* and *A. subpictus*, or four (A, B, C and D, as in *A. barbirostris* and *A. jeyporiensis*), while the “*Crucians Complex*” is described as including *A. bradleyi* and five species provisionally designated as *A. crucians* A, B, C, D and E.

The names of taxonomic units in sequence-based taxonomies

MOTUs vs. species names

The increasing use of sequence data in the discovery and taxonomic arrangement of biodiversity, as suggested for example by Tautz *et al.* (2003), opened the problem of the equivalence between the discovered clusters of sequence diversity and the species recognized and named in traditional taxonomy. It was soon realized that such an equivalence cannot be taken for granted without further scrutiny, even if the (increasingly automated) partitioning of sequence diversity is calibrated in order to maximize the correspondence between the resulting units and the Linnaean species. In addition, there is still a widespread

³ Oddly enough, one of the species (*Alopecosa koponeni*, correctly listed under its Linnaean name) is also given as new, despite the fact that it had been already described two years before (Blagoev & Dondale 2014).

although far from universal opinion, that Linnaean species should not be still diagnosed using molecular characters only.

Following a suggestion already floated online by Mark Blaxter for two years at least, Floyd *et al.* (2002) proposed the acronym MOTU (Molecular Operational Taxonomic Unit) as a term for any diagnosable and operationally useful set of sequences recognized within the barcode sequences. Recognizing a MOTU does not imply that this is a kind of ‘molecular hypothesis’ for the existence of a Linnaean species (Blaxter *et al.* 2005).

To some extent, this implies instead the existence of two parallel taxonomies. There are no simple or universal rules to follow, to move from MOTUs to species. Nevertheless, there are fine examples, such as the following one, that show how this may be possible.

Jörger *et al.* (2012) produced a remarkable study of the tiny meiofaunal sea-slugs of the genus *Pontohedyle* from 28 localities worldwide. Only two morphospecies are recognizable according to morphology, whereas a multi-marker genetic approach revealed a much larger, cryptic diversity. Authors recognized twelve MOTUs, for which a double labelling was used, thus making clear that molecular operational taxonomic units are other than taxonomic species, even in the case of a one-to-one correspondence between a MOTU and a species. In this paper, three MOTUs were actually identified with previously described species (MOTU VI = *Pontohedyle verrucosa*; MOTU XI = *P. milaschewitchii*; MOTU XII = *P. brasiliensis*), while the remaining ones were taxonomically referred to as *Pontohedyle* sp. 1 to 9. Remarkably, the numbering of MOTUs and tentative species are not always aligned, for example MOTU VIII is *Pontohedyle* sp. 6. This stresses the concept that MOTUs must not be regarded as species disguised under another name. *Pontohedyle* species 1 through 9 of Jörger *et al.* (2013) have been eventually described and named by Jörger *et al.* (2013) who have followed the rules of the Code, despite the lack of morphological traits differentiating these cryptic new species from one another; the new taxa have been established based on diagnostic nucleotides in DNA sequences of four genetic markers (mitochondrial cytochrome c oxidase subunit I, 16S rRNA, nuclear 28S and 18S rRNA) and DNA samples have been preserved as holotypes (accompanied, whenever possible, by fixed specimens as paratypes).

Mixed usage of names and formulae in the BOLD and GenBank databases

The largest sequence databases, in particular those of the Barcode of Life Data Systems (BOLD; <http://www.barcodinglife.org/>; cf. Ratnasingham & Hebert 2007) and GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) are by far the most important places where a tremendous diversity of formal and informal names and formulae are used to identify specimens, operational units and formal taxa. The main problem with this diverse set of labels is not the fact that many of them (such as those for MOTUs) do not follow the precepts of the Code, but the fact that most of these terms follow no standard, and are often ambiguous, or silent, about the kind of information they are intended to carry.

These labels form a *grey nomenclature* of which I will provide here examples, to be followed by suggestions for a radical improvement on the current situation.

As a sample of the terms used in BOLD, I have analyzed those contained in the taxonomic lists for four groups: Mammalia, Cestoda, ‘Turbellaria’ and Clitellata.⁴ As shown in Table 3, Linnaean binomials are 84.7% of all ‘names’ in the case of mammals, very similar to the 86.7% of the ‘names’ of cestodes, but the names in agreement with the Code are only 74.3% for the ‘turbellarians’ and just 54.3% for the clitellates.⁵ The worst aspect of this grey nomenclature in the BOLD database is its very heterogeneous nature, as reported in Table 3.

⁴ I do not object here to the fact that Turbellaria is not a natural group, neither to the fact that in the BOLD database acoels are still included within the Turbellaria, and therefore within the Platyhelminthes, as in the past; to some extent, this choice can be justified.

⁵ An additional problem with BOLD is the inadequate curation of the list of Linnaean names of the included taxa. By searching some partial list to get the information discussed in this article, I came across several instances of species listed twice under different binomials, including the following: Mammalia: *Hylobates syndactylus* and *Symphalangus syndactylus*, *Taurotragus oryx* and *Tragelaphus oryx*, *Manis temminckii* and *Smutsia temminckii*, *Manis tetradactyla* and *Phataginus tetradactyla* (sic), *Mustela vison* and *Neovison vison*, *Vampyressa brocki* and *Vampyriscus brocki*, *Cynopterus luzoniensis* and *Rousettus luzoniensis*; Clitellata: *Biwadrilus bathybates* and *Criodrillus bathybates*, *Dinodriloides beddardi* and *Proandricus beddardi*, *Allolobophora caliginosa* and *Aporrectodea caliginosa*, *Chamaedrillus cognettii* and *Cognettia cognettii*, *Allolobophora rosea* and *Aporrectodea rosea*; ‘Turbellaria’: *Geoplana multicolor* and *Paraba multicolor*, *Dugestia tigrina* and *Girardia tigrina*; Cestoda: *Hydatigera taeniaeformis* and *Taenia taeniaeformis*, *Taenia mustelae* and *Versteria mustelae*, *Andrya dentata* and *Anoplocephaloides dentata*.

Table 3. Statistics of Linnaean names and different kinds of names or formulae (grey nomenclature) included in the taxonomic lists of four groups of animals in the BOLD databases (<http://www.barcodinglife.org/>) as of 28 September 2016. Here, formula means a number, or group of letters, or a mixed alphanumeric string.

	Mamm	Clit	'Turb'	Cest	Examples
Linnaean binomen	2814	11056	176	157	
Id. + formula	38				Mammalia <i>Trachops cirrhosus</i> PS1
Tentative species-level identification ('aff.', 'cf.' or 's.l.')	146	13		11	Mammalia <i>Akodon</i> aff. <i>cursor</i> <i>Akodon</i> cf. <i>boliviensis</i> Clitellata <i>Haplotaxis</i> cf. <i>gordioides</i> Cestoda <i>Echinobothrium</i> cf. <i>chisholmae</i>
Id. + formula		120		6	Clitellata <i>Marionina minutissima</i> PDW2010 Cestoda <i>Paranoplocephala</i> cf. <i>omphalodes</i> III
Genus-level identification	78	101	3	3	Mammalia <i>Akodon</i> sp. Clitellata <i>Dendrobaena</i> sp. 'Turbellaria' <i>Geoplana</i> sp. Cestoda <i>Taenia</i> sp.
Id. + MOTU identifier		56			Clitellata <i>Ocnerodrilus</i> MOTU_35
Id. + non-MOTU formula	248	424	36	10	Mammalia <i>Akodon</i> sp. 1 <i>Cricetomys</i> sp. 3 PG2014 Clitellata <i>Pheretima</i> sp. 1-DGF 'Turbellaria' <i>Girardia</i> sp. 2 <i>Girardia</i> sp. AW2014 <i>Stenostomum</i> sp. longpit <i>Bipalium</i> sp. Kawakatsu <i>Diopisthoporus</i> sp. UJ-2011-nsp-2 Cestoda <i>Anoplocephaloides</i> sp. Q47
Family level identification		2			Clitellata Enchytraeidae gen. sp.
Id. + formula		168	7		Clitellata Lumbricidae gen. sp. DPEW104296 'Turbellaria' Geoplanidae gen. sp. 11
Total number of entries	3324	1944	237	181	

Similar problems with grey nomenclature are also present in GenBank, about which I will give here one example. Of the 306 entries in the taxonomic list for Lumbricidae (earthworms) in GenBank [accessed 28 09 2016], only 115 (37.6%) are standard binomens; another 15 entries are binomens followed by letters and/or numbers, e.g. *Allolobophora chlorotica* L5 (in two cases, the species epithet is preceded by ‘aff.’ of ‘cf.’); two entries are for species complexes, e.g. *Lumbricus rubellus* complex; 59 entries are for a genus name followed by a formula, e.g. *Lumbricus* sp. SL-2003; another 6 for genus name only; and 109 entries are for samples identified to family level only, and just distinguished by a formula that in most cases does not correspond to any recognizable standard, e.g. Lumbricidae sp. Esik50. A small number of formulae (e.g. *Eophila* sp. BOLD:ACJ0004 and Lumbricidae sp. BOLD:AAV0357) point to corresponding entries in the BOLD database, which are expressed in the rigorous format discussed in the next section.

The BINs in BOLD

The BINs, the Barcode Index Numbers of the BOLD database, are the best example of how a non-Linnaean nomenclature of animals (or plants) should be planned, defined and used. As explained by Ratnasingham and Hebert (2013), BINs are dynamically generated partitions of the total diversity of the sequences deposited with the BOLD database. The trees obtained by applying to its contents at any given time a dedicated algorithm called RESL (Refined Single Linkage) are segmented into clusters of sequences according to rules that optimally balance sequence distances within each cluster with sequence distances between each cluster and its nearest neighbour. Parameters are calibrated in such a way as to obtain a partition of the sequences in the database into clusters that are quite often in agreement with species-level taxonomic units. However, similar to MOTUs – as said – and as very clearly stressed by Ratnasingham & Hebert (2013), the alphanumerical codes (BINs) assigned to these clusters are definitely *not* intended as substitutes for species names. BINs are given an easily recognizable alphanumeric code (three capital letters and four digits, in the order) preceded by ‘BOLD:’. Two examples:

BOLD:AAA3964, currently (2 October 2016) clustering together 118 sequences from specimens identified, in Linnaean nomenclature, as *Mus musculus*

BOLD:AAA4089, corresponding (same date) to 57 sequences from bat specimens, of which 3 from specimens identified as *Hipposideros grandis*, 13 as *H. larvatus* and 31 tentatively referred to the latter species, as *Hipposideros cf. larvatus*

These examples deserve two remarks. First, the presence in the same BIN of sequences from specimens currently referred to two different species does not necessarily mean that these should be immediately regarded as synonymous. Second, sequences from specimens identified as belonging to the same species are often clustered in more than one BIN. This applies also to the two examples just given. In the case of mice, 46 sequences are found in BOLD:ACE3469 rather than in BOLD:AAA3964. In the case of the *Hipposideros* bats, while the sequences from specimens identified as *H. grandis* are all in BOLD:AAA4089, those from specimens identified as *H. larvatus* are distributed between this BIN (13 sequences, as said above) and BOLD:AAA4093 (34 sequences); in the case of specimens identified as *H. cf. larvatus*, sequences are distributed in as many as nine BINs. To some extent, this distribution could be due to the existence of a number of still undescribed species (Kruskop 2015).

The future of the grey nomenclature

BINs are an excellent example of how to create and use non-Linnaean formulae for segments of biological diversity: BINs are immediately recognizable and their meaning is unequivocally defined. These qualities are sadly lacking in the other kinds of grey nomenclature discussed above. Specifically, most of the grey nomenclature is seriously faulty in three respects.

First, the lack of rules for the creation and usage of these names of formulae precludes unambiguous understanding of what the name or formula is for. For example, what does ‘*Lumbricus* sp. A’ mean? Just an incomplete (genus-level only) identification, for specimens regarded as probably different at the species level from others, which are labelled, for the same reason, as *Lumbricus* sp. B? We have no standard of reference (not to mention a code) to which to refer to answer this question. We can therefore suppose that some authors will give exactly this meaning to that kind of names/formulae, but others may use them in the sense of ‘*Lumbricus* sp. n. A’, that is, a species recognized as new, but to be described at a later date.

Second, it is often unclear, especially in database entries, if the alphanumeric specifiers added to a genus name, or even to a Linnaean binomen, are intended to provide unambiguous labelling of a single specimen, or of a taxon, or both.

Third, and most important, because of the subjectivity of the criteria according to which these non-Linnaean names are created and used, it is generally impossible to compare them across studies, or databases (cf. Pante *et al.* 2015; Morard *et al.* in press), except perhaps by the researchers who are responsible for them and in a sense remain their ‘owners’. This is very far from those objectives of universality and stability to the service of which the rules of the Code have been devised.

It should be clear by now, that, when using names (or formulae) other than Linnaean binomens, it should be always made clear what the label is intended for. Most cases would result in one of the following alternatives:

(1) identification limited to the genus (or the family) level due to technical reasons such as (i) damaged or poorly preserved material, (ii) lack of specimens belonging to the stage or sex bearing morphological characters upon which a more precise identification would have been possible, or (iii) unavailability of competent specialists, for whatever reason. As a rule, there would be little reason, if any, to try a more precise identification of the same material at a later time, therefore, even if the preservation of voucher specimens is generally to be recommended, this may not be required. No further information content should be attached to the genus- or family-level names as used

(2) identification at the species level uncertain, mainly for the same reasons as (i) to (iii) under (1). However, this may be due to the lack of a modern revision of the group, in which case the specimens are better preserved in view of future taxonomic revisions. At any rate, the tentative identification is not associated with a hypothesis of the occurrence of a hitherto unrecognized (unnamed) taxon

(3) identification at the species level uncertain, because of the suspected occurrence of one or more hitherto unrecognized (unnamed) taxa in the genus, more frequently in a particular species group within a genus

(4) identification limited to the species group, or even to the genus, because of the discovery of hitherto unnamed species, the formal description of which is however reserved for future studies.

It would be nonsense to fight against the use of these non-Linnaean names and formulae. The point is another: this *grey nomenclature needs its rules*. I mean, a small set of simple, clear principles, fixing the standard format for each of the objects we need to distinguish.

It is vital to achieve an international and authoritative agreement on these matters. Not just for zoology, although the analysis presented in this article has been limited to animal names. At any rate, limiting here the discourse to zoology, responsibility for establishing rules for the grey nomenclature should reside first of all with the International Commission on Zoological Nomenclature, but other partners should be necessarily involved, especially those in whose hands huge numbers of these terms are handled, such as the major databases GenBank and BOLD.

Now, when looking for a way to standardize this grey nomenclature, where should we better look for inspiring principle: in the Linnaean nomenclature, or in BOLD’s BINs?

A fundamental difference between BINs and Linnaean names is the dynamic nature of the former as opposed to the unchanging nature of the latter. These opposite qualities are not a reason to prefer BINs over Linnaean names, or vice versa; to the contrary, this is one of the main reasons why we may take profit from both. Let’s spell out the argument in full. Each Linnaean name is intrinsically bound to its type (usually, a single specimen serving as its holotype, lectotype, or neotype): this is the ultimate material voucher for the taxon bearing that name. On the contrary, BINs have no types. Each BIN is a hypothesized operational unit of (molecular) biological diversity extracted from the total BOLD database at any precise time the latter is mined for this kind of information. BINs do not have fixed contents, but ‘evolve’ continuously, following the steady inflow of sequences.

All other kinds of names and formulae discussed in this article have been created instead with reference to a fixed set of specimens (often, just one) and/or sequences, analyzed with algorithms whose identity is often, but not always, specified in full. In the absence of detailed information such as normally provided with the description of a new species, the future use of these new names will depend on the fulfilment of the following conditions:

- definition of a number of standard formats for the different kinds of objects (or hypotheses) we need to label, such as individual species, undescribed new species, or other
- preservation (and exhaustive labelling) of material voucher (whole specimen(s), DNA sequence(s))

- permanent association of the ‘grey name’ with source information such as author and year for names introduced in a publication, or equivalent information, in suitable format to be specified, for unpublished database entries.

Pending the future of the registration policy for Linnaean names, it might be sensible also to register the informal names for hypothesized and still undescribed species. These informal names will not take precedence over the name eventually published with compliance with the Code rules, but will remain in their synonymy, thus ensuring a correct transfer of information from the phase in which the new taxon is recognized as such to the following phase in which it is eventually described and named, as in the case of the *Pontohedyle* species mentioned above.

Acknowledgements

I am grateful to an anonymous reviewer for useful suggestions.

References

- Barley, A. J., White, J., Diesmos, A. C. & Brown, R. M. (2013) The challenge of species delimitation at the extremes: diversification without morphological change in Philippine sun skinks. *Evolution*, 67, 3556–3572.
- Bittencourt-Silva, G., Conradie, W., Siu-Ting, K., Tolley, K. A., Channing, A., Cunningham, M., Farooq, H. M., Menegon, M. & Loader, S. P. (2016) The phylogenetic position and diversity of the enigmatic mongrel frog *Nothophryne* Poynton, 1963 (Amphibia, Anura). *Molecular Phylogenetics and Evolution*, 99, 89–102.
- Blagoev, G.A. & Dondale, C.D. (2014) A new species of *Alopecosa* (Araneae: Lycosidae) from Canada: a morphological description supported by DNA barcoding of 19 congeners. *Zootaxa*, 3894, 152–160.
- Blagoev, G. A., deWaard, J. R., Ratnasingham, S., deWaard, S. L., Lu, L., Robertson, J., Telfer, A. C. & Hebert, P. D. N. (2016) Untangling taxonomy: a DNA barcode reference library for Canadian spiders. *Molecular Ecology Resources*, 16, 325–341.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Robin Floyd, R. & Abebe, E. (2005) Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 360, 1935–1943.
- Cantino, P. D. & de Queiroz, K. (2010) *PhyloCode: International Code of Phylogenetic Nomenclature (Version 4c)*. <http://www.ohio.edu/phylocode>.
- Christidis, L., Goodman, S. M., Naughton, K. & Appleton, B. (2014) Insights into the evolution of a cryptic radiation of bats: dispersal and ecological radiation of Malagasy *Miniopterus* (Chiroptera: Miniopteridae). *PLoS ONE*, 9(3): e92440.
- Cook, F. R. & Gorham, S. W. (1979) The occurrence of the triploid form in populations of the Blue-spotted Salamander, *Ambystoma laterale*, in New Brunswick. *Journal of the New Brunswick Museum*, 1979, 154–161.
- Dubois, A. & Raffaëlli, J. (2009) A new ergotaxonomy of the family Salamandridae Goldfuss, 1820 (Amphibia, Urodela). *Alytes*, 26, 1–85.
- Floyd, R., Eyualem, A., Papert, A. & Blaxter, M. (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, 11, 839–850.
- Harbach, R. E. (2004) The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bulletin of Entomological Research*, 94, 537–553.
- Humer, P. & Karsholt, O. (1995) Gelechiidae, in Baldizzone, G., Gozmány, L., Huemer, P., Karsholt, O., Lvovsky, A., Parenti, U., Passerin d’Entrèves, P., Riedl, T., Varalda, P. G. & Zangheri, S., Lepidoptera Gelechioidea, in: Minelli, A., Ruffo, S. & La Posta, S. (eds.) *Checklist delle specie della fauna italiana*. Calderini, Bologna, 86, 28–41.
- Hunt, R. H. & Coetzee, M. (1991) Cytogenetic evidence for a new species within the taxon *Anopheles* (*Cellia*) *marshallii* (Theobald) (Diptera: Culicidae). *Mosquito Systematics*, 23, 191–194.
- International Commission on Zoological Nomenclature (1999) *International Code of Zoological Nomenclature, Fourth Edition*. London: The International Trust for Zoological Nomenclature, xxix+306 pp.

- Jörger, K. M. & Schrödl, M. (2013) How to describe a cryptic species? Practical challenges of molecular taxonomy. *Frontiers in Zoology*, 10: 59.
- Jörger, K. M., Norenburg, J. L., Wilson, N. G. & Schrödl, M. (2012) Barcoding against a paradox? Combined molecular species delineations reveal multiple cryptic lineages in elusive meiofaunal sea slugs. *BMC Evolutionary Biology*, 12: 245.
- Kruskop, S. V. (2015) Dull and bright: cryptic diversity within the *Hipposideros larvatus* group in Indochina (Chiroptera: Hipposideridae). *Lynx, n. s. (Praha)*, 46, 29–42.
- Lowcock, L. A., Licht, L. E. & Bogart, J. P. (1987) Nomenclature in hybrid complexes of *Ambystoma*: no case for the erection of hybrid “species”. *Systematic Zoology*, 36, 328–336.
- Martinsson, S. & Erséus, C. (2017) Cryptic speciation and limited hybridization within *Lumbricus* earthworms (Clitellata: Lumbricidae). *Molecular Phylogenetics and Evolution*, 106, 18–27.
- Mayr, E. (1931) Notes on *Halcyon chloris* and some of its subspecies. *American Museum Novitates*, 469, 1–10.
- Mayr, E. (1940) Speciation phenomena in birds. *American Naturalist*, 74, 249–278.
- Minelli, A. (2015) Taxonomy faces speciation: the origin of species or the fading out of the species? *Biodiversity Journal*, 6, 123–138.
- Morard, R., Escarguel, G., Weiner, A. K., André, A., Douady, C. J., Wade, C. M., Darling, K. F., Ujiie, Y., Sears, H. A., Quillévéré, F., De Garidel-Thoron, T., De Vargas, C. & Kucera, M. (2016) Nomenclature for the nameless: a proposal for an integrative molecular taxonomy of cryptic diversity exemplified by planktonic Foraminifera. *Systematic Biology*, 65, 925–940.
- Nygren, A. (2014) Cryptic polychaete diversity: a review. *Zoologica Scripta*, 43, 172–183.
- Pante, E., Puillandre, N., Viricel, A., Arnaud-Haond, S., Aurelle, D., Castelin, M., Chenuil, A., Destombe, C., Forcioli, D., Valero, M., Viard, F. & Samadi, S. (2015) Species are hypotheses: avoid connectivity assessments based on pillars of sand. *Molecular Ecology*, 24, 525–544.
- Pleijel, F. & Rouse, G. W. (1999) Least-inclusive taxonomic unit: a new taxonomic concept for biology. *Proceedings of the Royal Society of London B*, 267, 627–630.
- Pleijel, F. & Rouse, G. W. (2000) A new taxon, *capricornia* (Hesionidae, Polychaeta), illustrating the LITU ('least-inclusive taxonomic unit') concept. *Zoologica Scripta*, 29, 157–168.
- Pleijel, F. (2000) Phylogenetic taxonomy, a farewell to species, and a revision of Heteropodarke (Hesionidae, Polychaeta, Annelida). *Systematic Biology*, 48, 755–789.
- Postaire, B., Magalon, H., Bourmaud, C. & Bruggemann, J. H. (2016) Molecular species delimitation methods and population genetics data reveal extensive lineage diversity and cryptic species in Aglaopheniidae (Hydrozoa). *Molecular Phylogenetics and Evolution*, 105, 36–49.
- Ratnasingham, S. & Hebert, P. D. N. (2007) BARCODING, BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, 7, 355–364.
- Ratnasingham, S. & Hebert, P. D. N. (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLoS ONE*, 8(8): e66213.
- Raupach, M. J., Hendrich, L., Küchler, S. M., Deister, F., Morinière, J. & Gossner, M. M. (2014) Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. *PLoS ONE*, 9(9): e106940.
- Rensch, B. (1928) Grenzfälle von Rasse und Art. *Journal für Ornithologie*, 76, 222–231.
- Riedel, A., Sagata, K., Surbakti, S., Tänzler, R. & Balke, M. (2013) One hundred and one new species of *Trigonopterus* weevils from New Guinea. *ZooKeys*, 280, 1–150.
- Riedel, A., Tänzler, R., Balke, M., Rahmadi, C. & Suhardjono, Y. R. (2014) Ninety-eight new species of *Trigonopterus* weevils from Sundaland and the Lesser Sunda Islands. *ZooKeys*, 467, 1–162.
- Sarala, K. S., Nutan, N., Vasantha, K., Dua, V. K., Malhotra, M. S., Yadav, R. S. & Sharma, V. P. (1994) Cytogenetic evidence for three sibling species in *Anopheles fluviatilis* (Diptera: Culicidae). *Annals of the Entomological Society of America*, 87, 116–121.
- Scali, V., Passamonti, M., Marescalchi, O. & Mantovani, B. (2003) Linkage between sexual and asexual lineages: genome evolution in *Bacillus* stick insects. *Biological Journal of the Linnean Society*, 79, 137–150.
- Schultz, R. J. (1969) Hybridization, unisexuality and polyploidy in the teleost *Poeciliopsis* (Poeciliidae) and other vertebrates. *American Naturalist*, 108, 605–619.
- Seraphin, G. (2012) Le complexe *Agrias amydon* dans le haut bassin Amazonien (Equateur, Pérou, Brésil, Bolivie) (Lep. Nymph.). *Bulletin de la Société Entomologique de France*, 117, 461–472.

- Stur, E. & Borkent, A. (2014) When DNA barcoding and morphology mesh: Ceratopogonidae diversity in Finnmark, Norway. *ZooKeys*, 463, 95–131.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. & Vogler, A. (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, 18, 70–74.
- Tilquin, A. (2014) *Je t'aime, moi non plus: what evolutionary fate for sexual parasites?* Université Claude Bernard Lyon 1, Master 1, Ecosciences, Microbiologie.
- Tyler, M. J. (1991) Biological nomenclature, classification and the ethnozoological specieme. *The Journal of the Polynesian Society, Memoirs*, 48, 164–167.
- Witt, J., Threlloff, D. S., Doug, L. & Hebert, P. D. N. (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. *Molecular Ecology*, 15, 3073–3082.