ELSEVIER
MOLECULAR
PHYLOGENETICS
& EVOLUTION

Molecular Phylogenetics and Evolution 69 (2013) 39–45

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Barcoding in the dark?: A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge

Sebastian Kvist

Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 11 December 2012
Revised 14 May 2013
Accepted 16 May 2013
Available online 28 May 2013

Keywords:
DNA barcoding
Taxonomy
Animalia
BOLD
NCBI
Databases

ABSTRACT

The functionality of standard zoological DNA barcoding practice (the identification of unknown specimens by comparison of COI sequences) is contingent on working barcode databases with sufficient taxonomic coverage. It has already been established that the main barcoding repositories, NCBI and BOLD, are devoid of data for many animal groups but the specific taxonomic coverage of the repositories across animal biodiversity remains unexplored. Here, I shed light on this mystery by contrasting the number of unique taxon labels in the two databases with the number of currently recognized species for each animal phylum. The numbers reveal an overall paucity of COI sequence data in the repositories (15.13% total coverage across the recognized biodiversity on Earth, and 20.76% average taxonomic coverage for each phylum) and, more importantly, bear witness to the idleness towards numerous phyla, rendering current barcoding efforts either ineffective or inaccurate. The importance of further integrating taxonomic expertise into barcoding practice is briefly discussed and some guidelines, previously mentioned in the barcoding literature, are suggested anew. Finally, the asserted values concerning the taxonomic coverage in barcoding databases for Animalia are contrasted with those of Plantae and Fungi.

1. Introduction

Zoological DNA barcoding continues to grow as a popular means of identifying animal specimens by similarity comparison of partial cytochrome *c* oxidase subunit I (COI or *cox1*) sequences from the mitochondrial genome (Hebert et al., 2003a,b, 2004a,b; Hajibabaei et al., 2006; see also Ratnasingham and Hebert, 2007). In essence, barcoding involves two main elements, each as implicitly vital as the other in allowing unambiguous identification by means of molecular characterization. These are the query and the target. The query is normally represented by a partial COI sequence (~650 basepairs) of unknown origin, whereas the target is a COI-sequence that resides in a database or other repository with its identity previously determined (typically based on morphology and preferably to the species level). Whereas the initial aim of DNA barcoding was merely to allow for the identification of specimens (Hebert et al., 2003a), its application has greater implications. For example, it holds the power to energize taxonomy, promote epidemiology, further ecology, and benefit agriculture and aquaculture (e.g., Frézal and Leblois, 2008; Valentini et al., 2009). The method has already been used to augment border biosecurity (Armstrong and Ball, 2005; Boykin et al., 2012a; Collins et al., 2012), identify contraband bushmeat (Eaton et al., 2010), and advance the discussion on conservation management (Rubinoff, 2006; Ward et al., 2008; Brown and Paxton, 2009; Francis et al., 2010; Krishnamurthy and Francis, 2012).

For the last decade, proponents of DNA barcoding have attempted to build large-scale, rigorously curated DNA barcode storage databases (such as the Barcode Of Life Data System [BOLD]; Ratnasingham and Hebert, 2007), and practitioners of the method wholly rely on the taxonomic coverage of barcodes in these databases. Put bluntly, if the query sequence lacks a conspecific target counterpart in the database, species-level barcoding-based identification of the query will fail or, worse, produce a type I error (e.g., Nielsen and Matz, 2006; Kelly et al., 2007; Virgilio et al., 2010). Yet, in such instances, the sequences may still allow for accurate identification of the query to a level only slightly higher than species (e.g., Whitworth et al., 2007). COI sequences originating from congeners may also possess unique motifs or characteristic attributes that would allow for identification to the genus level; character-based barcoding can be applied by the use of e.g., CAOS (Sarkar et al., 2002, 2008) but this is not yet commonplace in barcoding practice. However, the aim of DNA barcoding is to aid in the resolution of the *species* diversity on Earth (Hebert et al., 2003a) and, for this aim to be realized, the ability of the method for identification to the level of species is essential. Moreover, DNA barcoding promises to shed light on our ever-growing understanding of the

E-mail address: skvist@fas.harvard.edu

1055-7903/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2013.05.012

diversity and abundance of cryptic species (Hebert et al., 2004a; Witt et al., 2006; Pfenniger and Schwenk, 2007). To the extent that traditional taxonomic tools fail to evince morphological characters that define a taxon, molecular techniques can clarify at least its similarity pattern towards closely related species.

In addition to resolving cryptic species complexes, both Hebert and Gregory (2005) and Hajibabaei et al. (2007) argue that barcoding is useful in potentially enhancing the rate of species discovery by revealing new taxa based on their sequence divergence. That is, the lack of a matching target sequence (given specific criteria) in a barcoding database may suggest the novelty of the species from which the query sequence was generated. Both contributions also mention that thorough taxonomic analyses are needed before any definitive statement. If taxonomic investigations were not carried out on flagged specimens, calling them novel would assume full taxonomic coverage of closely related species in the database and would suggest a species concept based on Euclidean neighbor-joining tree distances that does not conform to widely accepted taxonomic concepts (Rubinoff, 2006); specimen identification based on interspecific divergence vs. intraspecific variation has already been shown to be impossible for several taxonomic groups because of the lack of a barcoding gap (Boyer et al., 2007; Whitworth et al., 2007; see discussion below).

Both the exposition of cryptic species and the flagging of potentially new species, as well as standard barcoding procedures for identifying specimens, presuppose a working, extensive database covering the better part (if not all) of organismal life on the planet. Clearly, many organismal groups are underrepresented in currently available databases but, until now, no study has estimated the actual taxonomic coverage in the databases across the recognized biodiversity of life, in a comparative manner. One must recognize the Herculean task at hand for barcoders in sequencing and cataloging life on Earth and the purpose of this paper is chiefly to provide specifics on the current state of the two main barcode repositories, insofar as this is completely essential to the functionality of DNA barcoding. To this end, the present study provides a novel compilation of publicly available data, and is devoted to infusing new information concerning three main questions: (i) on a phylum per phylum basis, what is the actual coverage of DNA barcodes in the main barcode repositories, in relation to the recognized species diversity?, (ii) how can the coverage of the databases be increased in a manner that ensures the stability and correctness of the taxon labels that are associated with new sequences?, and (iii) who maintains the intellectual property to perform such a task?

2. Material and methods

2.1. The status of taxonomic barcode-coverage in BOLD and NCBI

Creating comprehensive COI databases is tedious work and it is unclear if the growth of current databases has kept pace equivalent with the growing popularity of the method. Several barcoding investigations have found specific groups lacking comprehensive taxonomic coverage (e.g., Boykin et al., 2012a,b; Siddall et al., 2012; Kwong et al., 2012) but it is still unknown if this pattern is mirrored across the full spectrum of life. To shed light on this question, the present study, for the first time, compares the taxonomic coverage in the two largest barcode repositories, BOLD v. 3 and GenBank (NCBI), to the current estimation of recognized biodiversity at the species level (Zhang, 2011; and references therein), with slight modifications to the phyletic separations. Specifically, as opposed to Zhang (2011), herein I consider Acanthocephala as part of Rotifera (Lorenzen, 1985), and both Echiura and Sipuncula as nested within Annelida (Rouse and Fauchald, 1997; Struck et al., 2007; Kvist and Siddall, in press). Consequently, 36 phyla are recognized in the present study; note that, whereas he mentions that 40 phyla are currently recognized, only 39 are listed in the table provided by Zhang (2011 pp. 9–11). Data were downloaded from NCBI using a transparent and reproducible strategy to retrieve information about the taxonomic coverage in the database: a "Taxonomy" search of NCBI was performed for each of the phyla and subsequent nested searches used also the keywords "cytochrome c oxidase subunit I", "cytochrome c oxidase I ", "COI" and "cox1". For example, for Annelida, the "COI" search term employed the following string: txid6340[Organism:exp] COI[Gene]. All hits were sent to a file in GenBank format and managed with standard unix language to show the number of hits for each unique taxon label: *grep ORGANISM sequence.gb | cut –d " " –f 4- | awk 'BEGIN{OFS="\t"} {n[$0]++} END {for (i in n) {print I, n[i]}}' > AnnelidacountNCBI.txt.* Data for the taxonomic coverage in BOLD are available directly on the website (http://www.boldsystems.org/) following a "taxonomy" search. Nevertheless, the data were manually copied from the website and arranged to show only unique taxon labels; the number of labels found did not always correspond to that stated by BOLD, even when removing taxa for which there is no accompanying COI sequence. The language used for the retrieval of unique taxon labels from BOLD was as follows: *sort AnnelidacountBOLD.txt | uniq*. Both the original NCBI data files and the BOLD data files are available from the author upon request (these are collectively roughly 2 gigabytes in size). Subsequently, because BOLD often mines data from NCBI, the two data sets were combined and the same script was used to tease out the overall taxonomic coverage from the databases combined. Usage of unique taxon labels of some sequences poses a putative challenge: conspecifics may be assigned different taxon labels by the authors. For example, *Allolobophora chlorotica* L1 and *Allolobophora chlorotica* L2 would each be considered unique taxon labels under the scheme presented here and, furthermore, misspellings like "*Homo sapins*" would result in an artificial increase in the number of unique taxon labels. Also, in several cases, the specific epithets of the taxa are undetermined and are followed by unique codes assigned by the authors even for several individuals included in the same population study and that are most likely the same species (e.g., *Acerophagus* sp. BMNH723311 and *Acerophagus* sp. BMNH723312). Alas, short of checking every record manually, this approach is one of the ways to retrieve a close approximation of the taxonomic coverage present in the databases.

3. Results

The results of these searches (Table 1 and Fig. 1) show that the number of unique taxon labels in NCBI and BOLD jointly cover 15.13% of the recognized biodiversity on Earth (235,013 unique taxon labels as compared to 1,553,399 recognized species [Zhang, 2011]). At the phyletic level, the two databases contain enough unique labels to putatively cover an average of 20.76% of the recognized species diversity within each phylum.

NCBI alone holds enough unique taxon labels to putatively cover, on average, 13.52% of the recognized species within each phylum; a total of 93,328 unique taxon labels are represented in the database (6.33% of the total recognized biodiversity). The best taxonomic coverage occurs in smaller phyla such as Cycliophora (100%) with only two recognized species, and Micrognathozoa (100%) with only a single species recognized. In contrast, no COI representation is present on NCBI for Kinorhyncha (Mud dragons), Loricifera (Girdle wearers), Myxozoa, Orthonectida, Rhombozoa or Placozoa (although a draft genome as well as whole mitochondrial genomes do exist for the latter; Srivastava et al., 2008). BOLD currently holds unique taxon labels with associated barcodes

Table 1
The number and distribution of animal species barcodes in NCBI and BOLD as compared to the known species-diversity of each phylum. NCBI representation and BOLD representation denotes the number of unique taxon labels present in the database; "Total representation" indicates the same for the databases conjoined; "Times fold current effort needed" is the closest integer of the number of "Recognized species" divided by the "Total representation". Databases were accessed on November 6, 2012.

Scientific name	Common name	Recognized species[a]	NCBI representation	BOLD representation[b]	Total representation[c]	Times fold current effort needed
Acoelomorpha	Acoel flatworms	393	79 (20.10%)	6 (1.53%)[d]	81 (20.61%)	5
Annelida[e]	Segmented worms	18,953	2013 (10.62%)	2063 (10.88%)	2882 (15.21%)	7
Arthropoda	Arthropods	1,242,040	69,123 (5.56%)	149,997 (12.08%)	189,319 (15.24%)	7
Brachiopoda	Lamp shells	443	24 (5.42%)	36 (8.13%)	40 (9.03%)	11
Bryozoa	Moss animals	10,941	140 (1.28%)	116 (1.06%)	195 (1.78%)	56
Cephalochordata	Lancelets	33	6 (18.18%)	4 (12.12%)	10 (30.30%)	3
Chaetognatha	Arrow worms	186	22 (11.83%)	24 (12.90%)	28 (15.05%)	7
Cnidaria	Cnidarians	10,105	734 (7.26%)	649 (6.42%)	997 (9.87%)	10
Craniata	Craniates	64,832	16,256 (25.07%)	22,855 (35.25%)[f]	26,899 (41.49%)	2
Ctenophora	Comb jellies	242	4 (1.65%)	3 (1.24%)	7 (2.89%)	35
Cycliophora	Cycliophorans	2	2 (100%)	2 (100%)	2 (100%)	0
Echinodermata	Echinoderms	13,000	648 (4.98%)	1406 (10.82%)	1613 (12.41%)	8
Entoprocta	Goblet worms	169	3 (1.77%)	4 (2.37%)	6 (3.55%)	28
Gastrotricha	Hairy backs	790	51 (6.46%)	0 (0%)	51 (6.46%)	15
Gnathostomulida	Jaw worms	109	8 (7.34%)	8 (7.34%)	9 (8.26%)	12
Hemichordata	Acorn worms	120	2 (1.67%)	4 (3.33%)	4 (3.33%)	30
Kinorhyncha	Mud dragons	179	0 (0%)	0 (0%)	0 (0%)	N/A
Loricifera	Girdle wearers	30	0 (0%)	0 (0%)	0 (0%)	N/A
Micrognathozoa	Micrognathozoans	1	1 (100%)	0 (0%)	1 (100%)	0
Mollusca	Molluscs	117,358	6844 (5.83%)	7180 (6.12%)	9786 (8.34%)	12
Myxozoa	Myxozoans	2402	0 (0%)	0 (0%)	0 (0%)	N/A
Nematoda	Round worms	24,783	720 (2.91%)	351 (1.42%)	848 (3.42%)	29
Nematomorpha	Horsehair worms	351	15 (4.27%)	0 (0%)	15 (4.27%)	23
Nemertea	Ribbon worms	1200	183 (15.25%)	104 (8.66%)	209 (17.42%)	6
Onychophora	Velvet worms	182	59 (32.42%)	55 (30.22%)	73 (40.11%)	2
Orthonectida	Orthonectids	43	0 (0%)	0 (0%)	0 (0%)	N/A
Phoronida	Horseshoe worms	10	2 (20%)	8 (80%)	9 (90%)	0
Placozoa	Placozoans	1	0 (0%)	1 (100%)[g]	1 (100%)	0
Platyhelminthes	Flatworms	29,285	691 (2.36%)	379 (1.29%)	883 (3.02%)	33
Porifera	Sponges	8346	386 (4.62%)	220 (2.64%)	457 (5.48%)	18
Priapulida	Penis worms	19	1 (5.26%)	1 (5.26%)	1 (5.26%)	19
Rhombozoa	Rhombozoans	123	0 (0%)	0 (0%)	0 (0%)	N/A
Rotifera[h]	Wheel animals	2777	123 (4.43%)	278 (10.01%)	347 (12.50%)	5
Tardigrada	Water bears	1157	72 (6.22%)	2 (0.17%)	72 (6.22%)	16
Tunicata	Sea squirts	2792	115 (4.12%)	146 (5.23%)[i]	167 (5.98%)	17
Xenoturbellida	Xenoturbellids	2	1 (50%)	1 (50%)	1 (50%)	1
Total		1,553,399	98,328 (6.33%)	185,903 (11.97%)[j]	235,013 (15.13%)	7
Average		39,831	2521 (13.52%)	5164 (14.62%)	6528 (20.76%)	16

[a] Based on current estimates from Zhang (2011) and references therein.

[b] Including both public and private entries in BOLD. Note that the number of public entries usually is much lower than the total count.

[c] Only unique entry names, i.e., identical entry names occurring in both NCBI and BOLD are only counted once.

[d] Including species of Acoela, no data is available for Nemertodermatida.

[e] Including both Echiura and Sipuncula (Rouse and Fauchald, 1997; Struck et al., 2007); values were for the total estimate.

[f] The value stated for the coverage of Craniata in BOLD encompasses Acanthopterygii, Actinopterygii, Amphibia, Aves, Cephalaspidomorphi, Chondrichthyes, Elasmobranchii, Holocephali, Mammalia, Myxini, Reptilia and Sarcopterygii.

[g] BOLD houses six COI sequences for Placozoa and it is likely that this phylum houses higher species diversity than currently recognized. Nevertheless, until new species are formally described, these six sequences are considered to belong to the same species.

[h] Including Acanthocephala (Lorenzen, 1985); values were added for the total estimate.

[i] Including Appendicularia, Ascidiacea and Thaliacea.

[j] The number stated on the BOLD website for the number of barcoded and formally described species is 126,042 (accessed November 6, 2012).

putatively covering an average of 14.62% of the species for each phylum, with sequences available for 29 out of the 37 phyla (Table 1); no data are available for Gastrotricha (Hairy backs), Micrognathozoa, Myxozoa, Nematomorpha (Horsehair worms), Orthonectida, Loricifera (Girdle wearers), Kinorhyncha (Mud dragons) or Rhombozoa. When comparing the total number of unique taxon labels ($n = 185{,}903$) with the recognized total number of species, BOLD holds 11.97% of the diversity. Note that this value is very similar to the value stated on the BOLD website: 171,383 "Species with Barcodes" resulting in coverage of 11.03% (although BOLD also states that barcodes exist for 126,792 described species [8.16%] and it is difficult to determine which of these values to trust; database accessed February 11, 2013). When removing the outliers mentioned above, for which taxonomic coverage is either unusually high (100%) or unusually low (0%), the total joined taxonomic coverage in NCBI and BOLD increases slightly to an average of 15.43%.

This broad scope of data also provides perspective to some skewed values (Fig. 1). Of the phyla with >10,000 recognized species, Bryozoa (Moss animals) and Platyhelminthes (Flatworms) are the most grossly neglected; the unique taxon labels putatively covering only 1.78% and 3.02% of the species, respectively. If dividing the number of currently recognized species with the number of unique taxon labels present in the databases as a proxy for the current effort devoted to increasing the taxonomic coverage, this would respectively necessitate 56 and 33 times the current COI-sequencing effort to barcode every species within the phylum. Naturally, the mega-diverse Arthropoda is represented by a high number of unique labels ($n = 189{,}319$). However, because of the speciose nature of the phylum, a staggering 1,052,721 of the

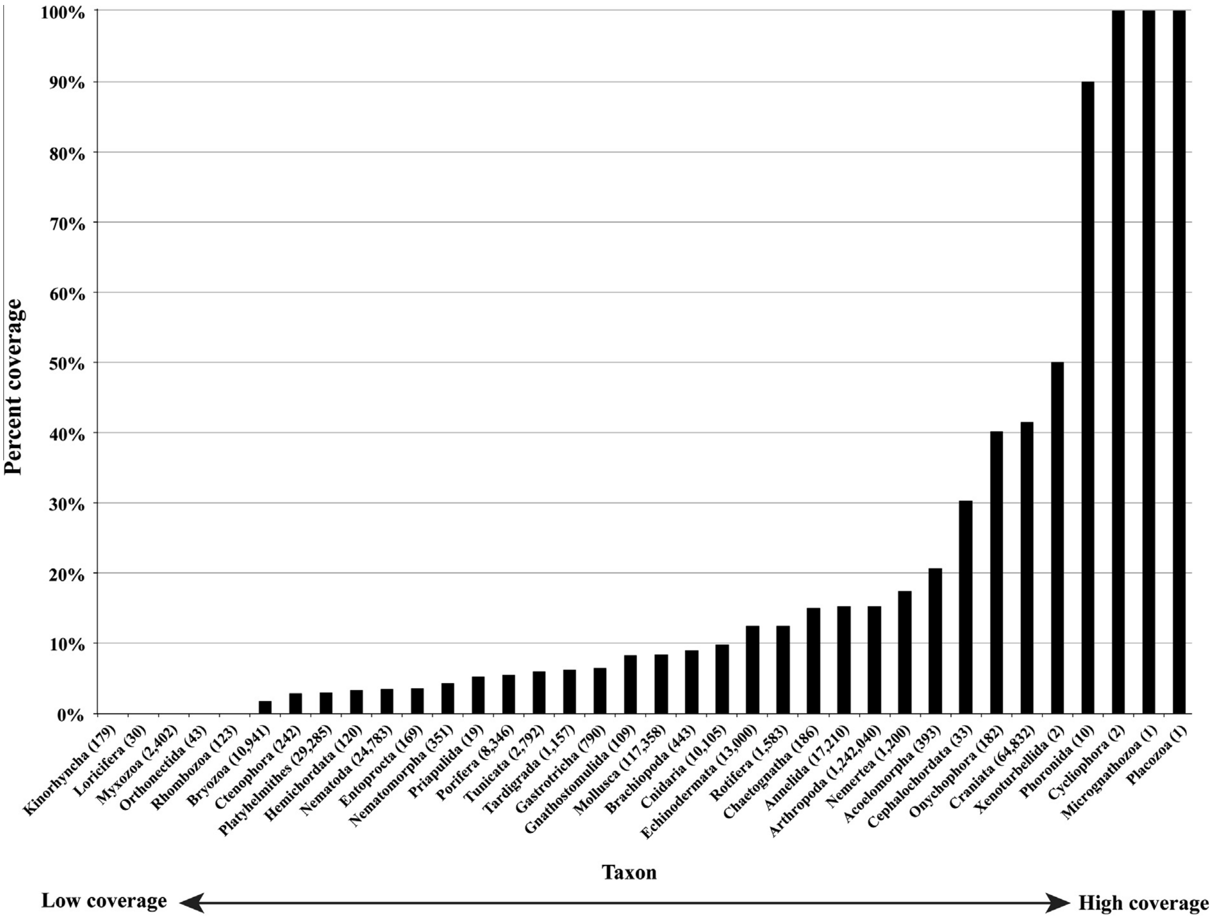

Fig. 1. Chart showing the proportion of effort devoted to the establishment of DNA barcodes for all recognized animal phyla (Zhang (2011) with slight modifications; see text for further details). The *X*-axis shows the various phyla with the number of recognized species in parentheses, and the *Y*-axis shows the percentage of coverage present jointly in BOLD and NCBI when compared to the recognized species diversity. Left-bound taxa have less coverage in the database, in terms of percent barcoded species, than the right-bound taxa.

recognized species (84.76%) are not associated with a COI sequence in the databases, rendering attempts at barcoding futile for the vast majority of taxa within the phylum. In contrast, Craniata, which is highly represented by 26,899 unique labels compared to the 64,832 recognized species (41.49%), only needs about twice the current effort to be completely barcoded. Within the phyla with <10,000 recognized species, Hemichordata (Acorn worms) is largely underrepresented, with only 4 unique taxon labels as compared to the 120 recognized species (3.33%), needing 30 times the current effort to represent the entire phylum with barcoded species. In addition, only 6 unique taxon labels compared to the 169 recognized species (3.55%) within Entoprocta (Goblet worms) are represented by COI sequences in the databases, necessitating 28 times the current effort for full taxonomic coverage. Ctenophora (Comb jellies), with 242 species recognized, is represented by a meager 7 unique labels (2.89%) indicating that 35 times the current effort is needed to completely fill the COI-sequence void. Onycophorans (Velvet worms) are, despite their elusiveness (Bouvier, 1905; Ruhberg, 1985), relatively well-represented by barcodes in the databases; 73 unique taxon labels compared to the 182 recognized species (40.11%) are represented by COI barcodes and only about twice the current effort is needed for full representation.

As mentioned above, the values relating to unique taxon labels stated here are probably very generous when compared to the true number of species represented in the databases, as even the smallest of differences, such as the addition of authors' specimen codes, will consider the sequences as representative of distinct taxon labels. To be explicit, it is likely that, in some cases, the same species was counted more than once by virtue of only minute differences in taxon labels, thereby artificially increasing the coverage in the database as compared to the number of recognized species. A corollary issue is that different species may have been counted as a single unique taxon label if the specific epithet is undetermined; for example, the taxonomic label *Pareiorhaphis* sp. may occur twice on NCBI and may represent two different species but will only be counted once because of the lack of precision.

4. Discussion

4.1. Skewed species coverage

Overall, the paucity of comprehensive representation of so many taxa in the barcoding databases renders barcoding relatively useless for the majority of life on Earth, aside from the charismatic Craniata and low-diversity groups such as Micrognathozoa or Placozoa. For example, for invasive insects, Boykin et al. (2012a,b) showed that BOLD lacked data for 42% of the species examined, and Siddall et al. (2012) noted an absence of COI data in both BOLD and NCBI even for very common nematode species. In an investigation of the DNA barcode coverage for Metazoa, Kwong et al. (2012) found that only 60,930 species out of the estimated 1.7 million species (Marshall, 2005 as referenced by Kwong et al. (2012)) were represented by barcodes in NCBI, whereas BOLD reported barcodes for ca. 150,000 species; however, it is unclear how Kwong et al. (2012) dealt with overestimations due to different taxon labels for the same species and the study does not provide any specific values for lower levels within Metazoa. The same study concludes both that 74% of the barcodes in NCBI are derived from specimens that have only been identified to levels higher than species, and that the growth of the number of barcodes (not necessarily the coverage depending on the target taxa) is rather slow. Importantly, however, Kwong et al. (2012) show that species coverage is better for those taxa involved in specific barcoding campaigns. This might suggest that we will see a continued growth of coverage for target taxa while the idleness will remain for other, perhaps less charismatic groups.

It is evident that certain groups of organisms are either more demanding in terms of DNA extraction or are less perceptive to species-level identification by COI-barcoding. Regardless of this, however, zoological DNA barcoding has come to target COI alone for the purpose of cataloguing biodiversity, and the advancement of taxonomic coverage of problematic groups should not be excused here but, rather, evaluated in light of this target, much like non-problematic groups.

4.2. The holy trinity?: DNA barcoding, taxonomy and geography

The potential future application of DNA barcoding in the areas mentioned in Section 1 (the resolution of cryptic species complexes, flagging of potentially new species, etc.), as well as those areas in which barcoding has become standard practice for specimen identification (border biosecurity, conservation biology, epidemiology, etc.) will most likely have to rely on databases with substantially increased coverage to that currently available. How, then, can the expansion of barcoding databases be accelerated while ensuring the stability and correctness of the taxon labels with which the sequences are affiliated? By incorporating accurately barcoded queries into the target database, the repositories will gain taxonomic coverage, but this only becomes possible after rigorous morphological investigation of the specimen from which the first-round query was derived – otherwise how can it be used as an authoritative target barcode (especially if no voucher material is present) and how can barcoding produce reliable assessments without authoritative target barcodes? Thus, barcoding is not only underpinned by taxonomy but at the mercy of those with taxonomic knowledge of any given group. Barcoding relies on taxonomy for the integration of barcoding results into the body of scientific knowledge. Although similar ideas have been argued in previous contributions (for discussions, see Tautz et al., 2002, 2003; Lipscomb et al., 2003; Moritz and Cicero, 2004; Will and Rubinoff, 2004; Schander and Willassen, 2005; Ebach and Holdredge, 2005; DeSalle et al., 2005; Rubinoff et al., 2006), it seems that there is still a notable separation between the fields of barcoding and taxonomy and it is worth underscoring their importance relative to each other. As emphasized by Rubinoff et al. (2006), this separation does not mean that most morphologists oppose the use of molecular techniques for inferences on the systematics of any taxon but merely that DNA barcoding cannot *replace* traditional taxonomy. Here, I hold the same to be true and suggest that research at the organismal level, rather than perhaps viewing the advancement of barcoding as detrimental to taxonomy, should embrace the possibilities that follow such advancement. Although taxonomists typically possess vast knowledge even beyond the difficulty of delimiting and describing new species, it is obvious that describing all life on Earth is an overwhelmingly large feat in itself. In considering the need for taxonomic expertise in, among other areas, DNA barcoding, 277 years after the arguably first major taxonomic work, *Systema Naturae* (Linnaeus, 1758), we have reached roughly 1.5 million nominal animal species (Zhang, 2011). When extrapolated, and assuming that the world's total animal biodiversity is predicted accurately at 5–15 million species (May, 1988, 1997; Wilson, 2003; Savolainen et al., 2005), 923–2770 years will pass before the current number of taxonomists have completely cataloged the planet's diversity. The main concern of barcoding-opposed taxonomists seems to be the ideological backbone of DNA barcoding and perhaps its overly ambitious goals, given the limitations of the method (see Waugh, 2007; DeSalle, 2007), and, as such, is not mediated by the rather contrived example of the timeline for cataloging the species diversity on Earth. However, given both the considerable resources devoted to DNA barcoding (Ebach and de Carvalho, 2010; Kwong et al., 2012), and its reliance on taxonomic coverage in the barcoding databases, a closer collaboration between barcoding and taxonomy would likely be mutually beneficial. That is, if all authoritative barcodes are associated with specimens that have undergone rigorous taxonomic scrutiny, the deductive power of barcoding will increase immensely (e.g., Will et al., 2005; Hajibabaei et al., 2007; Collins & Cruickshank, 2012) and, at the same time, taxonomists will gain valuable resources (funding, access to specimens, etc.). The argument that barcoding does not propose to infer species delimitations will optimistically be maintained by practitioners, and it can be noted that it has been implicitly stated that species discovery is not mandated by barcoding alone but by the taxonomic community in general and needs to be supported by other lines of evidence (Hebert and Gregory, 2005; p. 854). Whether this is maintained or not is dependent upon the integrity of the investigator.

It is becoming apparent that numerous barcodes for each species are needed to ensure the presence (or absence for that matter) of a discernable barcoding gap between interspecific divergence and intraspecific variation in any given taxon. The lack of such a gap can hinder accurate specimen identification (Meyer and Paulay, 2005; Shearer and Coffroth, 2008) because individuals within the same species will show an average level of variation equivalent to the average divergence when compared to other species. This is a clear-cut case, frequently discussed in the literature (Wiemers and Fiedler, 2007; Meier et al., 2008; Srivathsan and Meier, 2012), in which DNA barcoding must surrender to morphological examinations if and when autapomorphic characters are available to analyze. Because of the relatively frequent lack of a barcoding gap between congeners, some studies suggest that barcoding cannot be used to identify specimens at the taxonomic levels that are the most important (Sperling, 2003; Will and Rubinoff, 2004). On the one hand, in some cases, the presence of a gap may merely be contingent on the paucity of comparable data, and therefore contingent on the apparent insufficiency of current barcoding databases. On the other hand, if type specimens of all recognized species in the world were associated with a barcode, performance of barcoding techniques would have increased functionality and the taxonomic coverage in these databases is directly proportional to the functionality and accuracy of DNA barcoding. As such, barcoding and taxonomy would mutually benefit from a database holding both authoritative barcodes *sensu* Kvist et al. (2010), preferably from types, and other morphologically compatible specimens that may counsel on the presence or absence of a barcoding gap.

In mimicking the rules of the International Code of Zoological Nomenclature, an authoritative target barcode, deposited in some database, would be best acquired from a type specimen. That is, ideally, an authoritative barcode should be a "type-barcode". There are several obstacles pertaining to the use of "type-barcodes" such as the bulk of type specimens in the world proving refractory to DNA sequencing and that using a single "type-barcode" would repudiate intraspecific variation. There are ways to ameliorate these obstacles by following the rationale behind traditional taxonomy – in order of decreasing authoritativeness, this means barcoding holotypes, paratypes, other specimens from the type series, morphologically compatible specimens from the type locality or, when all else fails, morphologically compatible specimens from as geographically close to the type locality as possible (Kvist et al., 2010). The geographic connectivity infers a level of security in that it is more likely that the barcoded specimens belong to the "type-population" and to some extent lessens the risk of inaccurate identification that would follow from barcoding a cryptic variant elsewhere in the world. However, it is becoming increasingly important to note that there is still a risk that cryptic species co-occur at the type locality (e.g., Hebert et al., 2004a; Stuart et al., 2006; Bely and Weisblat, 2006) but in such cases, and when the holotype defies DNA sequencing, neither morphology nor molecules can fully clarify the ambiguity.

4.3. Comparing database coverage across kingdoms

Botanical DNA barcoding has struggled to agree upon universal DNA barcodes that both hold the discriminatory power needed and that are easily amplified across the diversity of Plantae; the same criteria used for choosing COI for animals. Although some may disagree, it seems as though the most likely candidate plant barcode loci are *rbcL* and *matK*, which, when combined, seem to hold high discriminatory power (e.g., Chase et al., 2005; Newmaster et al., 2006; Ford et al., 2009; Kress and Erickson, 2012; but see also Hollingsworth et al., 2011). BOLD states that the repositories stores barcodes for 47,123 species (this is the number of species with barcodes, not the number of specimen records, which is much higher), representing between 12.83% and 18.14% of the estimated 259,721–367,196 species formally recognized (Chapman, 2009 and references therein); it is still unclear what types of barcodes (i.e., what locus) BOLD considers for plants. Thus, if the numbers presented here are correct, botanical barcoding has come no further than zoological barcoding in terms of overall coverage of the kingdom.

Much like those of Plantae, researchers of Fungi have evaluated several loci for their potential as fungal barcodes. Recent studies seem to converge on the agreement that the internal transcribed spacer (ITS) region, which occupies part of the ribosomal cistron, holds the discriminatory power required and is sufficiently easy to amplify across Fungi (Nilsson et al., 2008; Bergerow et al., 2010; Schoch et al., 2012). The kingdom still suffers from the lack of a clear definition as to what constitutes a fungus and the challenges to taxonomy are likely more similar to those of Bacteria than Animalia, as suggested by Seifert (2009). Regardless of this, the species diversity has been estimated to include between 72,000 and 100,000 recognized species (Hawksworth and Rossman, 1997). By contrast, BOLD declares that a meager 331 species (0.33–0.46%) are represented by barcodes in the database. This astonishingly low percentage leads one to believe that most DNA barcodes for Fungi are stored elsewhere, for example NCBI or the International Fungal Working Group (http://www.fungalbarcoding.org). Unfortunately, the international Barcode Of Life project specifically aimed at barcoding Fungi (aptly named Fun-BOL) does not report on the number of barcoded species online. Clearly, more barcodes for Fungi exist in other places and it would be ignorant to say that these values reflect the current state of fungal DNA barcoding. However, they do evince the fact that BOLD does not seems to be the main repository of fungal barcodes.

Although these comparisons are very superficial, they indicate that the pace at which zoological DNA barcoding databases have increased in coverage (see also Kwong et al., 2012) is equivalent to, or greater than, that for Plantae and Fungi.

Acknowledgments

This contribution was funded by the Wenner-Gren Foundations, the Sixten Gemzéus Foundation, Helge Ax:son Johnson's Foundation and the Olle Engkvist Byggmästare Foundation. This manuscript benefitted greatly from insightful comments from Anna Phillips, Alejandro Oceguera-Figueroa, Mark Siddall, Gonzalo Giribet and three anonymous reviewers.

References

Armstrong, K.F., Ball, S.L., 2005. DNA barcodes for biosecurity: invasive species identification. Philos. Trans. Roy. Soc. Lond. B 360, 1813–1823.

Bely, A.E., Weisblat, D.A., 2006. Lessons from leeches: a call for DNA barcoding in the lab. Evol. Dev. 8, 491–501.

Bergerow, D., Nilsson, H., Unterseher, M., Maier, W., 2010. Current state and persepctives of fungal DNA barcoding and rapid identification procedures. Appl. Microbiol. Biotechnol. 87, 99–108.

Bouvier, E.L., 1905. Monographie des Onychophores. Ann. Sci. Nat. Zool. (Ser. 9) 2, 1–384.

Boyer, S.L., Baker, J.M., Giribet, G., 2007. Deep genetic divergences in *Aoraki denticulate* (Arachnida, Opilliones, Cyphophtalmi): a widespread 'mite harvestman' defies DNA taxonomy. Mol. Ecol. 16, 4999–5016.

Boykin, L., Armstrong, K.F., Kubatko, L., De Barro, P., 2012a. Species delimitation and global biosecurity. Evol. Bioinform. 8, 1–37.

Boykin, L., Armstrong, K., Kubatko, L., De Barro, P., 2012b. DNA barcoding invasive insects: database roadblocks. Invert. Syst. 26, 506–514.

Brown, M.F.J., Paxton, R.J., 2009. The conservation of bees: a global perspective. Apidologie 40, 410–416.

Chapman, A.D., 2009. Numbers of Living Species in Australia and the World, second ed. Australian Biodiversity Information Services, Toowoomba, Australia.

Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haidar, N., Savolainen, V., 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans. Roy. Soc. B 360, 1889–1895.

Collins R.A. and Cruickshank R.H., The seven deadly sins of DNA barcoding, Mol. Ecol. Res. 2012, http://dx.doi.org/10.1111/1755-0998.12046

Collins, R.A., Armstrong, K.F., Meier, R.F., Yi, Y., Brown, S.D.J., Cruickshank, R.H., Keeling, S., Johnston, C., 2012. Barcoding and border biosecurity: identifying cyprinid fishes in the aquarium trade. PLoS ONE 7, e28381.

Desalle, R., 2007. Phenetic and DNA taxonomy; a comment on Waugh. BioEssays 29, 1289–1290.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philos. Trans. Roy. Soc. B 360, 1905–1916.

Eaton, M.J., Meyers, G.L., Kolokotronis, S.-O., Leslie, M.S., Martin, A.P., Amato, G., 2010. Barcoding bushmeat: molecular identification of central African and South African harvested vertebrates. Conserv. Genet. 11, 1389–1404.

Ebach, M.C., de Carvalho, M.R., 2010. Anti-intellectualism in the DNA barcoding enterprise. Zoologia 27, 165–178.

Ebach, M.C., Holdredge, C., 2005. DNA barcoding is no substitute for taxonomy. Nature 434, 697.

Ford, C.S., Ayres, K.L., Toomey, N., Haider, N., van Alpen Stahl, J., Kelly, L.J., Wikström, N., Hollingsworth, P.M., Duff, R.J., Hoot, S.B., Cowan, R.S., Chase, M.W., Wilkinson, M.J., 2009. Selection of candidate coding DNA barcoding regions for use on land plants. Bot. J. Linn. Soc. 159, 1–11.

Francis, C.M., Borisenko, A.V., Ivanova, N.V., Eger, J.L., Lim, B.K., Guillén-Servent, A., Kruskop, S.V., Mackie, I., Hebert, P.D.N., 2010. The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. PLoS ONE 5, e12575.

Frézal, L., Leblois, R., 2008. Four years of DNA barcoding: current advances and prospects. Infect. Genet. Evol. 8, 727–736.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D.N., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA 103, 968–971.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N., Hickey, D.A., 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23, 167–172.

Hawksworth, D.L., Rossman, A.Y., 1997. Where are all the undescribed fungi? Phytopathology 87, 888–891.

Hebert, P.D.N., Gregory, T.R., 2005. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 852–859.

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003a. Biological identifications through DNA barcodes. Proc. Roy. Soc. Lond. B Biol. Sci. 270, 313–321.

Hebert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003b. Identification of birds through DNA barcodes. PLoS Biol. 270, S96–S99.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H., Hallwachs, W., 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA 101, 14812–14817.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004b. Identification of birds through DNA barcodes. PLoS Biol. 2, e312.

Hollingsworth, P.M., Graham, S.W., Little, D.P., 2011. Choosing and using a plant DNA barcode. PLoS ONE 6, e19254.

Kelly, R.P., Sarkar, I.N., Eernisse, D.J., DeSalle, R., 2007. DNA barcoding using chitons (genus *Mopalia*). Mol. Ecol. Res. 7, 177–183.

Kress, W.J., Erickson, D.L., 2012. DNA barcodes: methods and protocols. Methods Mol. Biol. 858, 3–8.

Krishnamurthy, P.K., Francis, R.A., 2012. A critical review on the utility of DNA barcoding in biodiversity conservation. Biodivers. Conserv. 21, 1901–1919.

Kvist, S., Siddall, M.E., in press. Phylogenomics of Annelida revisited: a cladistics approach using genome-wide EST data mining and examining the effects of missing data. Cladistics. http://dx.doi.org/10.1111/cla.12015 (in press).

Kvist, S., Oceguera-Figueroa, A., Siddall, M.E., Erséus, C., 2010. Barcoding, types and the *Hirudo* files: using information content to critically evaluate the identity of DNA barcodes. Mitochondrial DNA 21, 198–205.

Kwong, S., Srivathsan, A., Meier, R., 2012. An update on DNA barcoding: low species coverage and numerous unidentified sequences. Cladistics 28, 639–644.

Linnaeus, C., 1758. Systema Naturae: sive regna tria naturae systematice proposita per classes, ordines, genera et species, 10th Ed. Theodorum Haak, Leiden, the Netherlands.

Lipscomb, D., Platnick, N., Wheeler, Q., 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends Ecol. Evol. 18, 64–66.

Lorenzen, S., 1985. Phylogenetic aspects of pseudocoelamate evolution. In: Conway Morris, S., George, J.D., Gibson, R., Platt, H.M. (Eds.), The Origins and Relationships of Lower Invertebrates. Clarendon Press, Oxford, UK, pp. 210–223.

Marshall, E., 2005. Will DNA bar codes breathe life into classification? Science 307, 1037.

May, R.M., 1997. The dimensions of life on earth. In: Raven, P.H. (Ed.), Nature and Human Society: The Quest for a Sustainable World. National Academy Press, Washington, DC, pp. 30–45.

May, R.M., 1988. How many species are there on Earth? Science 241, 1441–1449.

Meier, R., Zhang, G., Ali, F., 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. Syst. Biol. 57, 809–813.

Meyer, C.P., Paulay, G., 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3, e422.

Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2, e354.

Newmaster, S.G., Fazekas, A.J., Ragupathy, S., 2006. DNA barcoding in land plants: evaluation of *rbcL* in a multigene tiered approach. Can. J. Bot. 84, 335–341.

Nielsen, R., Matz, M., 2006. Statistical approaches for DNA barcoding. Syst. Biol. 55, 162–169.

Nilsson, R.H., Kristiansson, E., Ryberg, M., Hallenberg, N., Larsson, K.-H., 2008. Intraspecific ITS variability in the kingdom *Fungi* as expressed in the International Sequence Database and its implications for molecular species identification. Evol. Bioinform. Online 4, 193–201.

Pfenniger, M., Schwenk, K., 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. BMC Evol. Biol. 7, 121.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system. Mol. Ecol. Res. 7, 355–364, <http://www.barcodinglife.org>.

Rouse, G.W., Fauchald, K., 1997. Cladistics and polychaetes. Zool. Scr. 26, 139–204.

Rubinoff, D., 2006. Utility of mitochondrial DNA barcodes in species conservation. Conserv. Biol. 20, 1026–1033.

Rubinoff, D., Cameron, S., Will, K., 2006. A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J. Hered. 97, 581–594.

Ruhberg, H., 1985. Die Peripatopsidae (Onychophora). Systematik, ekologie, chorologie und phylogenetische aspekte. Zoologica 137, 1–183.

Sarkar, I.N., Planet, P.J., Bael, T.E., Stanley, S.E., Siddall, M., DeSalle, R., Figurski, D.H., 2002. Characteristic attributes in cancer microarrays. J. Biomed. Inform. 35, 111–122.

Sarkar, I.N., Planet, P.J., DeSalle, R., 2008. CAOS software for use in character-based DNA barcoding. Mol. Ecol. Res. 8, 1256–1259.

Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K., Lane, R., 2005. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. Philos. Trans. Roy. Soc. B 360, 1805–1811.

Schander, C., Willassen, E., 2005. What can biological barcoding do for marine biology? Mar. Biol. Res. 1, 79–83.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., et al., 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc. Natl. Acad. Sci. USA 109, 6241–6246.

Seifert, K.A., 2009. Progress towards DNA barcoding of fungi. Mol. Ecol. Res. 9 (Suppl. 1), 83–89.

Shearer, T.L., Coffroth, M.A., 2008. Barcoding corals: limited by interspecific divergence, not intraspecific variation. Mol. Ecol. Res. 8, 247–255.

Siddall, M.E., Kvist, S., Phillips, A., Oceguera-Figueroa, A., 2012. DNA barcoding of parasitic nematodes: is it Kosher? J. Parasitol. 98, 692–694.

Sperling, F., 2003. DNA barcoding. Deux et machina. Newsl. Biol. Surv. Can. (Terrestrial Arthropods), Opin., 22, <http://www.biology.ualberta.ca/bsc/news22_2/contents.htm>.

Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., et al., 2008. The *Trichoplax* genome and the nature of placozoans. Nature 454, 955–960.

Srivathsan, A., Meier, R., 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28, 190–194.

Struck, T.H., Schult, N., Kusen, T., Hickman, E., Bleidorn, C., McHugh, D., Halanych, K.M., 2007. Annelid phylogeny and the status of Sipuncula and Echiura. BMC Evol. Biol. 7, 57.

Stuart, B.L., Inger, R.F., Voris, H.K., 2006. High level of cryptic species diversity revealed by sympatric lineages of South Asian forest frogs. Biol. Lett. 2, 470–474.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2002. DNA points the way ahead in taxonomy. Nature 418, 479.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18, 71–74.

Valentini, A., Pomanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. Trends Ecol. Evol. 24, 110–117.

Virgilio, M., Backeljau, T., Nevado, B., De Meyer, M., 2010. Comparative performances of DNA barcoding across insect orders. BMC Bioinform. 11, 206.

Ward, R.D., Holmes, B.H., White, W.T., Last, P.R., 2008. DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. Mar. Freshwater Res. 59, 57–71.

Waugh, J., 2007. DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29, 188–197.

Whitworth, T.L., Dawson, R.D., Magalon, H., Baudry, E., 2007. DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). Proc. Roy. Soc. B 274, 1731–1739.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Front. Zool. 4, 8.

Will, K., Rubinoff, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. Cladistics 20, 47–55.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844–851.

Wilson, E.O., 2003. The encyclopedia of life. Trends Ecol. Evol. 18, 77–80.

Witt, J.D.S., Threloff, D.L., Hebert, P.D.N., 2006. DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Mol. Ecol. 15, 3073–3082.

Zhang, Z.-Q., 2011. Animal biodiversity: an introduction to higher-level classification and taxonomic richness. Zootaxa 3148, 7–12.