Mitochondrial DNA, December 2010; 21(6): 198–205

informa
healthcare

Barcoding, types and the *Hirudo* files: Using information content to critically evaluate the identity of DNA barcodes

SEBASTIAN KVIST[1,2], ALEJANDRO OCEGUERA-FIGUEROA[2,3], MARK E. SIDDALL[2,4], & CHRISTER ERSÉUS[5]

[1]*Richard Gilder Graduate School, American Museum of Natural History, New York, USA,* [2]*Division of Invertebrate Zoology, American Museum of Natural History, New York, USA,* [3]*Department of Biology, The Graduate Center, City University of New York, New York, USA,* [4]*Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, USA, and* [5]*Department of Zoology, University of Gothenburg, Göteborg, Sweden*

(*Received 13 June 2010; revised 4 October 2010; accepted 4 October 2010*)

Abstract
Species identifications based on DNA barcoding rely on the correct identity of previously barcoded specimens, but little attention has been given to whether deposited barcodes include correspondence to the species' name-bearing type. The information content associated with *COX1* sequences in the two most commonly used repositories of barcodes, GenBank and the Barcode of Life Data System (BOLD), is often insufficient for subsequent evaluation of the robustness of the identification procedure. We argue that DNA barcoding and taxonomy alike will benefit from more information content in the annotations of barcoded specimens as this will allow for validation and re-evaluation of the initial specimen identification. The aim should be to closely connect specimens from which reference barcodes are generated with the holotype through straight-forward taxonomy, and geographical and genetic correlations. Annotated information should also include voucher specimens and collector/identifier information. We examine two case studies based on empirical data, in which barcoding and taxonomy benefit from increased information content. On the basis of data from the first case study, we designate a barcoded neotype of the European medicinal leech, *Hirudo medicinalis*, on morphological and geographical grounds.

Keywords: *DNA barcoding, taxonomy, type series, geographical data,* Hirudo medicinalis, *neotype*

Introduction

The growth and increased frequency of DNA barcoding campaigns emphasizes the importance of validating the databases upon which species identifications are based. In terms of actual practice, relatively little attention is given by submitters and users of barcodes to the layers of validation that would ensure proper identifications. One conspicuous aspect of this shortcoming is that the two main repositories of barcodes [GenBank/NCBI and the Barcode of Life Data System (BOLD)] do not fully recognize the distinction between the reference database entries and subsequently generated barcodes. The former should be predicated on authoritative entries providing a baseline for association with the latter. In other words, the taxonomic identities of reference barcodes need to be disambiguated through information content and critical evaluation prior to any sensible identity inference of query barcodes. This scenario is often overlooked in the growing barcode databases (although somewhat alleviated in BOLD; see below) such that all barcodes are collectively lumped into a single bin upon which subsequent identifications are based. In addition, there is a lack of information annotated in the barcode submissions that, if present, would allow for discrimination of the validity of the taxonomic labels of the barcodes in the bin. This has led to uncritical majority rules in actual barcoding practice; the determination of a result from barcoding is based on the number of high scoring hits that are

Correspondence: S. Kvist, Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA. Tel: 1 212 313 7635. Fax: 1 212 769 5277. E-mail: skvist@amnh.org

ISSN 1940-1736 print/ISSN 1940-1744 online
DOI: 10.3109/19401736.2010.529905

encountered in the database. This is even more disconcerting given that taxonomic labels from consistently wrongly identified specimens are free to spread throughout the public sequence repositories (Nilsson et al. 2006; Ross and Murugan 2006). Had holotypes for all species ever described been sequenced for the barcode region, or at least been provided with a DNA voucher, subsequent barcoding would be greatly improved. Yet, several issues such as multiple copies of *COX1* in the mitochondrion, nuclear pseudogenes, or the fact that different species have almost identical *COX1* sequences (Williams and Knowlton 2001; Wiemers and Fiedler 2007) would still be problematic. A sensible approach to ameliorating the problem of non-sequenceable type specimens would be major barcode repositories retaining and providing information about barcoded non-types regarding georeferenced locality data, collector, identifier of record, dates of collection and identification, as well as morphological and genetic voucher catalogue numbers. Insofar as *COX1* sequences (i.e. barcodes) are available for rather few type specimens, considerable care must be taken to provide maximal information content relating to specimens actually being used to generate the reference barcode database.

The barcoding process is intended to more rapidly allow specimen identification through comparisons of short DNA sequences without the laborious process of morphological identifications (Hebert and Gregory 2005). To the extent that reference barcodes are useful for such a rapid procedure, they must actually be predicated on just such a laborious process. It has been rightly argued that DNA barcoding itself is a poor tool for species discovery and delimitation (Will and Rubinoff 2004; DeSalle et al. 2005; Ebach and Holdredge 2005; Will et al. 2005). Standard delimitation of species requires rigorous morphological analyses and, in some cases, genetic investigations. Typically, this is achieved by comparisons with type material or voucher specimens as well as specialized literature such as original descriptions, monographs, and/or taxonomic keys. Instead, barcoding should be used as a tool for identifying specimens belonging to a species already represented in the database or as an initial, crude way of "flagging" a potentially new species that needs further investigation (DeSalle 2006). However, the decreased attention to morphological attributes of specimens from which query barcodes are generated strongly increases the need for taxonomic validity of specimens from which the reference sequences were obtained. Furthermore, it increases the need for a connection between the barcode sequence and morphology or other descriptive characteristics (like geography) that are common in standard taxonomy. We recognize the initial need for momentum to generate as many DNA barcodes for as many species as possible for the idea of barcoding to gain traction. For the approach to mature, however, it is now more critical that authoritative barcode reference sequences be created for each species, with specifications as to which pre-existing barcode(s) have the highest information content relative to its connection to type material. Here, we contemplate this issue by discussing the scientific value added by the use of information-rich voucher specimens and the inherent need to create barcode reference databases that allow for critical evaluation of their validity.

Location, location, location

Unlike GenBank, which is not designed to aid barcoding initiatives, BOLD does distinguish between "validated" and "unvalidated" barcodes. This distinction might readily give the user a sense of security that some barcoding-based identification of a query sequence carries the weight of scientific authority if only made in reference to the "validated" subset. Such complacency is misplaced as the criteria for inclusion in BOLD's "validated" barcode database are only that three or more minimum length representatives with the same taxonomic label are included in a cluster and are collectively less than 2% different (Ratnasingham and Hebert 2007). Given these criteria for validity, the taxonomic label of any consistently misidentified species will remain faulty (e.g. Siddall et al. 2009). Barcode databases are not heavily policed, nor do any submissions necessarily have the benefit of peer review. Such procedures, while they would allow for an approach to validation, might also severely hamper the progress of barcoding initiatives. One possible counterbalance to the presently unknowable taxonomic validity of even the "validated" BOLD sequences is to ameliorate the scarcity of information content regarding geographical location of a specimen's collection site, its defining morphological characteristics, as well as collection dates, collector, identifier and the like. In particular, too little attention has been paid to the significance of type localities. A schematic layout of the information content establishment for reference barcodes is presented in Figure 1.

The best approach to DNA barcoding would be that a single yardstick for each species would be based on the highest possible information content; that is a barcode from the holotype (or, when applicable, the lectotype or neotype) that would forever represent a species as the reference barcode. Any identity of subsequent sequences, on the other hand, would only be determined based on their similarity to that reference barcode. However, much of the type material in museums or other scientific collections is too old, too degenerate, or stored in media not allowing for DNA sequencing. As such, the second highest level of certainty when creating a reference barcode would come from sequencing specimens from the remainder of the type series (e.g. paratypes, and paralectotypes). If these also prove refractory to

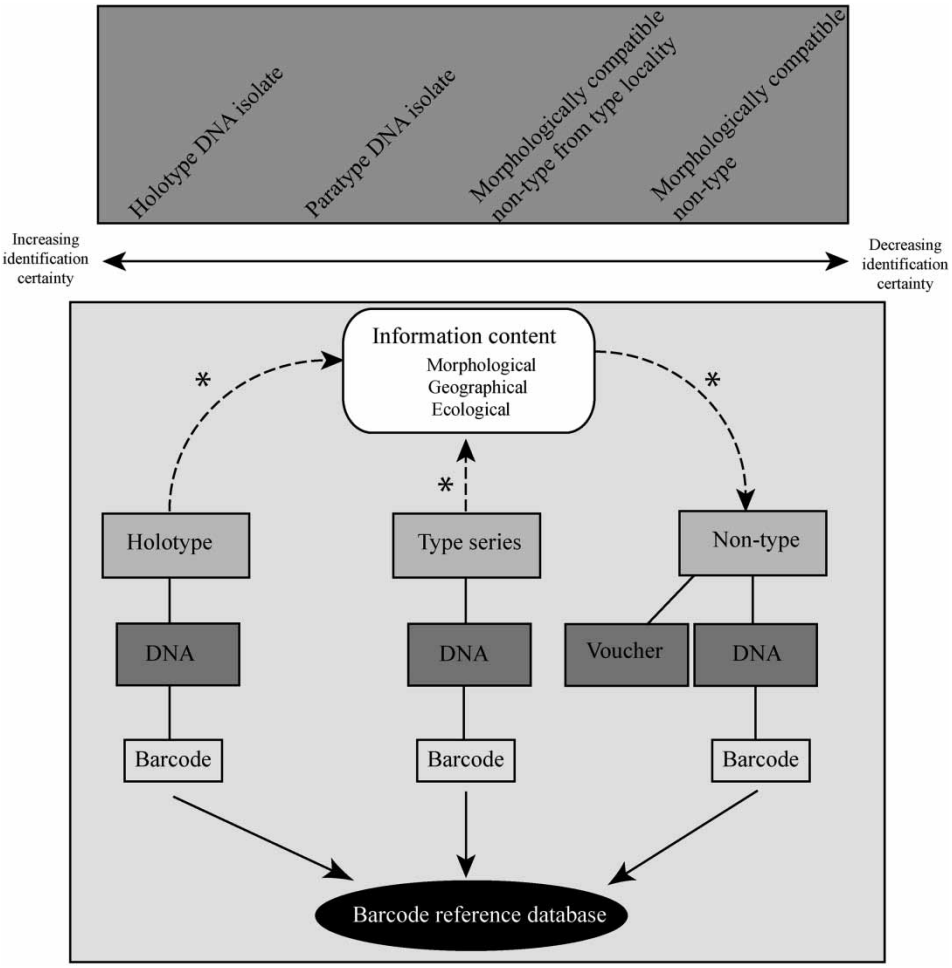

Figure 1. Schematics of the process of establishing information content for reference barcodes. The largest certainty in the taxonomic identity of a specimen comes from following the left path (using the holotype), the second largest from following the middle path (using remainder of type series), and the third largest from following the right path (using nontype specimens from the type locality with the addition of voucher specimens). *Information carry-over (morphological, geographic and ecological) from the type series to nontype specimens.

sequencing, it is reasonable to suggest that the third highest level of certainty would be brought through an established morphological and geographical relatedness between the specimen from which the reference barcode is to be obtained and the holotype of the species. As with any taxonomic survey program, this should include information regarding some correspondence to the holotype itself or its description. That is, beyond being morphologically compatible, information content would be increased if there were some geographic correspondence between specimens from which the reference sequence is obtained and the type specimen. Collecting specimens from the type locality (the collection site of the holotype) has two main benefits. To the extent that both biotic and abiotic factors have shaped species distributions, these factors are information embodied by the type locality. A link between species, their geographic distribution, and a reference sequence would greatly improve the argument that the reference barcode obtained from a specimen from the type locality belongs to the same species as the holotype. Furthermore, in light of the many morphologically cryptic species being uncovered (Bickford et al. 2007), in large part eventuated by barcoding initiatives themselves, geography can assist in determinations. Present DNA databases largely consist of specimens for which identifications have either not been accurately validated in these ways, or for which such critical evaluation is not possible. This is not to ignore the possibility of several morphologically cryptic species co-occurring at a type-locality (Hebert et al. 2004; Bely and Weisblat 2006) although, in such cases, if viable holotype DNA is absent, neither genetics nor morphology can solve the ambiguity.

With a more information-rich database, different levels of certainty (or actual validity) would be discernable from the annotations of individual barcodes. Even without the benefit of peer review for multitudinous reference sequences, users would be empowered to evaluate identifications in more rigorous ways. If no morphological information is available for a barcode submission or if it came from other than the type locality, one need not completely disregard a submission's identification, putative as it

may be. However, any vague association between such a putative reference sequence and the type series should be readily indicated in some way. This more careful strategy for establishing a reference barcode library mirrors the best practices of taxonomy, wherein a holotype has precedence over the remainder of the type series and in which neotypes are best designated from the type locality (International Commission on Zoological Nomenclature 1999, Article 75.3.6). This practice is also followed in comprehensive phylogenetic revisions where the authors spend considerable time and resources collecting specimens of type species of type genera from type localities (e.g. Oceguera-Figueroa et al. 2005; Borda et al. 2008). Even in cases where a type locality has been destroyed, geographic information allows for making the best of a bad situation: a reference sequence should be taken from a specimen with inferred biogeographic affinity to the former type locality. This information content must take precedence over the current "validation" system provided by BOLD.

To the extent that we have focused on type localities as a proxy for creating reference barcodes, this should be expanded to include several other inherent biological realities; for example, host–parasite or plant-associated insect systems. For example, any reference barcode sequence for a spiral valve tapeworm taken from a thorny skate (*Amblyraja radiata* Donovan, 1808) would lack the taxonomic authority inherent in a morphologically indistinguishable cestode originally described from the winter skate type-host (*Leucoraja ocellata* Mitchell, 1815), even if the former was collected from the type locality where both skates co-occur. The study of these conjoined systems is highly dependent on accurate identifications of both associates (Besansky et al. 2003). It is reasonable to suggest that both a parasite and its host should be subjected to simultaneous DNA barcoding, applying the same guidelines concerning the validation of reference barcodes as detailed above. These added efforts in turn provide for enhanced scientific inquiry to the extent that they will allow barcoding databases to provide proper identifications of juveniles or of various life-history stages in intermediate hosts.

The use of voucher specimens

Voucher specimens necessarily are stored in natural history collections at museums or universities and can be any part of an organism that is needed to disambiguate taxonomic identity (e.g. whole specimens, microscope slides, herbarium sheets, and digital photographs; see Miller 2007; Pleijel et al. 2008). Our appeal for the use of voucher specimens in barcoding is that they allow for subsequent re-evaluation. Concerns about the taxonomic identity of specimens become increasingly urgent with the growing use of molecular data (Goldblatt et al. 1992; Schander and Willassen 2005; Nilsson et al. 2006; Pleijel et al. 2008). A voucher specimen of any barcoded organism will enable later cross-referencing between genetics and morphology as well as provide for critical re-evaluation of the identity of the specimen. That is, voucher specimens are at their best when anyone can use them to control the labels associated with a DNA sequence. Pleijel et al. (2008) argued that some organisms are easily recognized but deciphering the taxonomic identity of others still requires the attention of a small number of taxon-specific experts. The growing body of taxonomic dilettantes further underscores the need for enabling subsequent re-evaluation of the identity of specimens used both for creating the reference library and for subsequent queries. In an attempt to quantify the incidence of voucher use, Pleijel et al. (2008) found that only 46% of the DNA sequence submissions they examined include references to voucher depositions. Although a formatted field for voucher catalogue numbers is present in both GenBank and BOLD, it seems as though voucher deposition is far from standard protocol for sequence submission to GenBank. While the vast majority of barcode submissions in BOLD are annotated with a storing institution, considerably fewer include catalogue numbers. We acknowledge that there are several cases where the deposition of voucher specimens becomes impractical or even impossible (environmental samples or destructive extraction protocols). These issues are rather frequently discussed (e.g. Neigel et al. 2007; Rowley et al. 2007; Hunter et al. 2008) and we are confident that progress is being made towards ameliorating them. Regardless, in the event that barcode submissions cannot provide information-rich background data, none of them ought to be considered authoritative, much less validated.

Case studies and the use of reference barcodes

To underscore the importance of the morphological and geographic components to specimen identification, we provide empirical examples that illustrate how barcoding can connect to name-bearing types or type localities of species.

The notorious European medicinal leech, *Hirudo medicinalis* Linnaeus, 1758, is frequently employed in, for example, leech therapy (Whitaker et al. 2004), developmental biology (Fernández and Stent 1982) and neurobiology (Muller et al. 1981). Accurate interpretations of both bioactive compounds, such as anticoagulants, and developmental and neurophysiologic characteristics presuppose accurate specimen determinations (Siddall et al. 2007). Siddall et al. (2007) showed that commercially available medicinal leeches are in fact not *H. medicinalis* but rather a close relative, *Hirudo verbana* Carena, 1820. While this

result is less than astonishing owing to the similarity of morphological features between these two leech species, it can largely affect their usage in medicine because of their putatively different repertoire of anticoagulants. Linnaeus (1758) did not leave any type material for *H. medicinalis*, which makes subsequent identifications more complicated. In addition to a morphological predicament, specimens from which publicly available *COX1* sequences have been attained were collected in areas that are largely separate, both geographically and ecologically, from what was, no doubt, a Swedish type locality. Linnaeus did not specify the type locality in his description but, because he was very active in his native area in and around the province of Uppland in Sweden (Reid 2009), we assume that the specimens used to describe the species were collected in that area. For the present study, we collected specimens from both Uppland and Gotland (the Swedish island in the Baltic Sea) (Table I). The geographic localities for *H. medicinalis* specimens used for genetic studies include Slovenia, Ukraine, Germany, France, and Sweden (Jördens et al. 2004; Trontelj and Utevsky 2005; Siddall et al. 2007; present study). DNA was extracted and the *COX1* region amplified and sequenced following standard protocols described elsewhere (De Wit et al. 2009). All publicly available *COX1* sequences ($n = 41$) for *H. medicinalis*, *H. verbana*, and *Hirudo orientalis* Trontelj and Utevsky, 2005 were downloaded from GenBank and aligned with the newly generated sequences using the ClustalW2 algorithm (Larkin et al. 2007) on the European Bioinformatics Institute web server applying default settings; three *COX1* sequences annotated as *H. medicinalis* (EU100093, AF003272, AY364862) were excluded as these have been shown to not be *Hirudo* species (Trontelj and Utevsky 2005; personal observation). In addition, a *COX1* sequence of *Hirudo nipponia* Whitman, 1886 was used to root the tree. PAUP* 4.0b10 (Swofford 2002) was used to construct a neighbor-joining tree using the Kimura two-parameter correction model (Kimura 1980). As an aside, the Kimura two-parameter model is frequently employed by DNA barcoders without explicit justification. Here, it is used to minimize the disparity between this and other barcoding studies, but we note that a formal justification for using the Kimura two-parameter model in barcoding studies is greatly needed. The resulting neighbor-joining tree is presented in Figure 2. The *COX1* sequences downloaded from GenBank and labeled "*H. medicinalis*" are identical to or extremely similar to the newly acquired specimens, suggesting that these also conform to *H. medicinalis* Linnaeus, 1758. Importantly, sequences of the specimens collected in the more southern part of Sweden (Gotland) are also very similar or identical to those of the Uppland specimens, suggesting that both localities are inhabited by the same population. As such, the geographical affinity between our specimens and those of Linnaeus increases the information content annotated in the barcodes. Below, we assign a neotype of *H. medicinalis* based on morphological characteristics and underscored by our understanding of the locality at which Linnaeus' original material was collected. Also, we associate the neotype with a barcode generated from the specimen, and this barcode will enjoy authoritative power with respect to subsequent identifications.

ARYNCHOBDELLIDA Blanchard, 1894
HIRUDINIDAE Whitman, 1886
Hirudo medicinalis Linnaeus, 1758

Material examined: Two specimens collected 24 July 2008 in Fräkensjön Lake, Hållnäshalvön Peninsula, Uppland, Sweden by Stefan Lundberg. An additional three specimens collected 13 July 2009 in Gåsavätar Swamp, Tingstäde, Gotland, Sweden by Sara Eliasson. All specimens are lodged at the Swedish Museum of Natural History (SMNH), Stockholm.

Neotype: Whole leech fixed in 80% ethanol. Collected by Stefan Lundberg on 24 July 2008 (SMNH Type-8027; GenBank accession number HQ333516).

Neotype locality: Fräkensjön Lake NW of Hållnäs, Hållnäshalvön Peninsula, Tierp, Uppland, Sweden N 60°34′13″ E 17°52′57″.

Other barcoded specimens from Sweden: One whole leech specimen fixed in 80% ethanol; collected by Stefan Lundberg at neotype locality on 24 July 2008 (SMNH111546). Three whole leech specimens

Table I. List of specimens used in the *Hirudo medicinalis* case study.

Voucher	DNA isolate	GenBank accession number	Collector	Collection locality in Sweden	Latitude/longitude	Collection date
SMNH111543	CE6842	HQ333519	S. Eliasson	Gotland, Gåsavätar Swamp	N 57°40′52″ E 18°35′35″	13 July 2009
SMNH111544	CE6844	HQ333518	S. Eliasson	Gotland, Gåsavätar Swamp	N 57°40′52″ E 18°35′35″	13 July 2009
SMNH111545	CE6849	HQ333517	S. Eliasson	Gotland, Gåsavätar Swamp	N 57°40′52″ E 18°35′35″	13 July 2009
SMNH Type-8027	**CE7347**	**HQ333516**	**S. Lundberg**	**Uppland, Fräkensjön Lake**	**N 60°34′13″ E 17°52′57″**	**24 July 2008**
SMNH111546	CE7349	HQ333515	S. Lundberg	Uppland, Fräkensjön Lake	N 60°34′13″ E 17°52′57″	24 July 2008

Notes: Bold denotes the neotype. Linnaeus (1758) is the authority for all specimens. Specimens were identified by C. Erséus and were not feeding on a host.

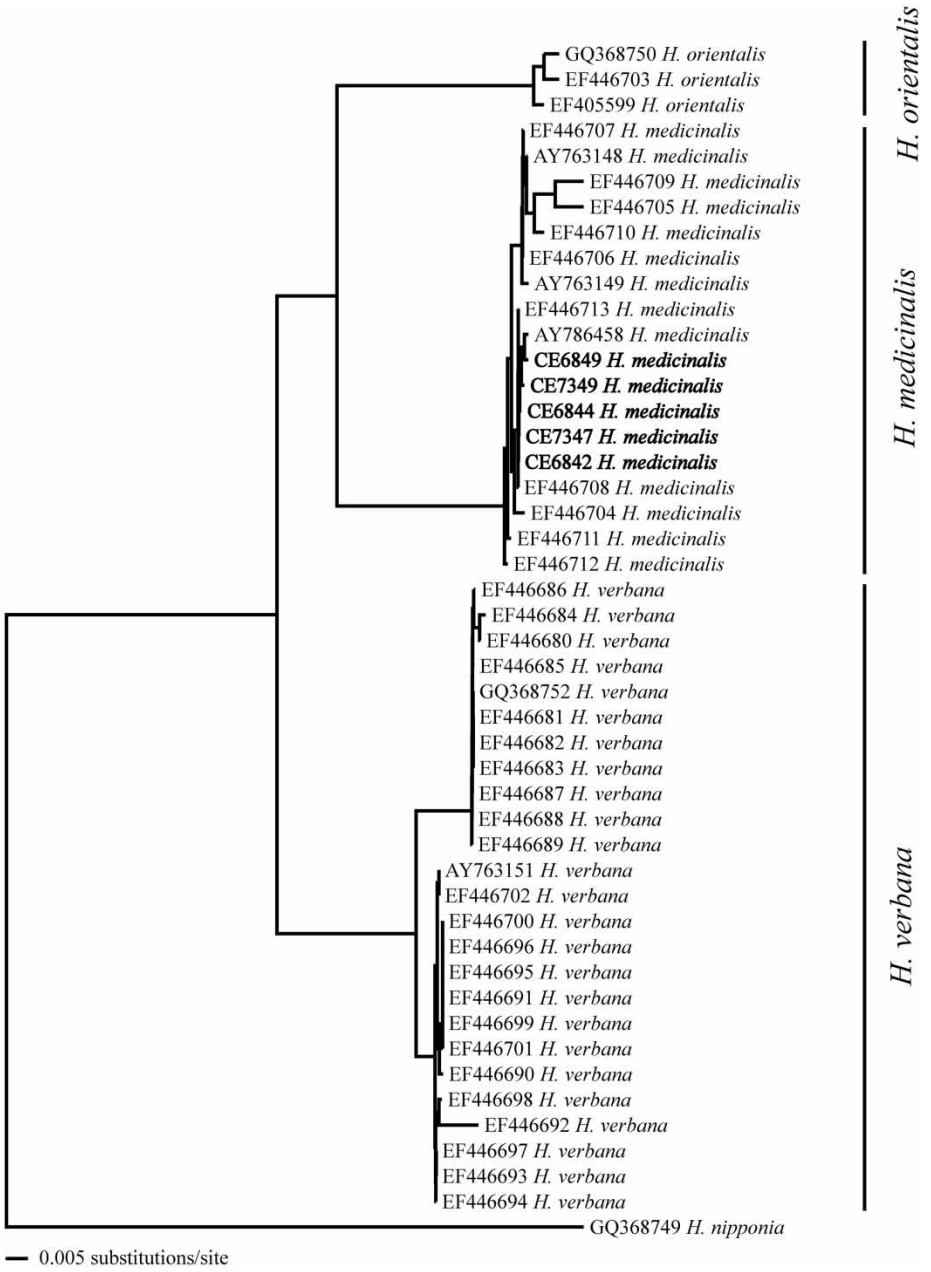

Figure 2. Neighbor-joining tree of the *COX1* locus showing three distinct clades representing *H. orientalis, H. medicinalis* and *H. verbana*, respectively. Specimens in bold were collected near the assumed type locality of the species (see text for further discussion). CE7347 denotes the neotype.

fixed in 80% ethanol; collected at Gåsavätar Swamp, Tingstäde, Gotland, Sweden (N 57°40′52″ E 18°35′35″) by Sara Eliasson on 13 July 2009 (SMNH111543-111545).

Description: See Linnaeus (1758) as re-described by Moquin-Tandon (1827) and verified by Siddall et al. (2007).

Remarks: Siddall et al. (2007) showed that the dorsal color patterns associated with each of *H. medicinalis, H. verbana*, and *H. orientalis* are species specific. Furthermore, the dorsal pattern of the *H. medicinalis* neotype corresponds to both the description by Linnaeus (1758) and that of Moquin-Tandon (1827). The latter did distinguish between *H. medicinalis* and *H. verbana* based on the dorsal color pattern. The consensus between these three investigations regarding the species-specific color patterns of *H. medicinalis* corresponds to our observations of the specimens collected in or close to the assumed type locality (Table I).

As a second example, we examined the marine oligochaete genus *Tubificoides* Lastockin, 1937. Specimens of these small worms are traditionally dehydrated in an ethanol/xylene series and mounted whole in Canada balsam on microscope slides

(e.g. Kvist et al. 2008), both of which prevent DNA sequencing of previously designated type material. *Tubificoides amplivasatus* Erséus, 1975 was originally described from the west coast of Norway. Morphologically identical specimens from the Atlantic coast of Spain were recently identified by Kvist et al. (2010) and were initially thought to belong to the same species. Genetic analyses revealed that these two populations differ by a mean of 12.74% in *COX1*. Corroborating with nuclear data, Kvist et al. (2010) showed that these populations represent different, yet closely related, species and that the Spanish specimens of *T.* cf. *amplivasatus* (labeled *T. amplivasatus II* in that study) represent an undescribed species. Kvist et al. (2010) also collected specimens from the type locality of *T. amplivasatus sensu stricto*, which enabled this inference. However, had no specimens of *T. amplivasatus sensu stricto* previously been barcoded and no attention been given to the type locality, the Spanish specimens would probably be labeled *T. amplivasatus* Erséus, 1975, and their *COX1* sequences would be the basis for subsequent identifications. This not only illustrates the power of barcoding in guiding taxonomic studies, but also shows the dangers of neglecting to establish a connection in information content between the type material, type locality, and some query specimen. Indeed, unbeknownst to the authors, this scenario may have already been manifest in a staggering number of other studies as it only requires a morphologically cryptic species, now known to be common in several animal groups, to be present anywhere in the world.

Increasing information content in the barcode submissions will allow for subsequent evaluation of the strength of those barcodes as specimen identifiers through critical evaluation of the annotated information. With the increase in information, barcoding will not only approach rigorous traditional taxonomy but may also re-vitalize this important yet unfortunately diminishing science.

Acknowledgements

The authors thank Rob DeSalle for encouraging the writing of this paper. They are grateful to Anna Phillips, Ed Stanley, Jairo Arroyave and two anonymous reviewers for providing helpful comments that much improved an early draft of the manuscript. The authors are also grateful to Sara Eliasson and Stefan Lundberg for providing leeches and to Anna Ansebo for assistance with sequencing.

Declaration of interest: This contribution was funded by the Richard Gilder Graduate School fellowship and the Swedish-American Foundation fellowship to S.K., CONACyT scholarship 172322 and the Graduate Center Science fellowship (CUNY) to A.O.-F., NSF grant DEB-0640463 to M.E.S., and Swedish Research Council grant 621-2007-5313 to C.E. The authors alone are responsible for the content and writing of the paper.

References

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

Besansky NJ, Severson DW, Ferdig MT. 2003. DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. Trends Parasitol 19:545–546.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I. 2007. Cryptic species as a window on diversity and conservation. Trends Ecol Evol 3:148–155.

Borda E, Oceguera-Figueroa A, Siddall ME. 2008. On the classification, evolution and biogeography of terrestrial haemodipsoid leeches (Hirudinida: Arhyncobdellida: Hirudiniformes). Mol Phylogenet Evol 46:142–154.

DeSalle R. 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. Conserv Biol 20: 1545–1547.

DeSalle R, Egan MG, Siddall ME. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. Philos Trans R Soc B 360:1905–1916.

De Wit P, Rota E, Erséus C. 2009. *Grania* (Annelida: Clitellata: Enchytraeidae) of the great barrier reef, Australia, including four new species and a re-description of *Grania trichaeta* Jamieson, 1977. Zootaxa 2165:16–38.

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

Fernández J, Stent GS. 1982. Embryonic development of the hirudinid leech *Hirudo medicinalis*: Structure, development and segmentation of the germinal plate. J Embryol Exp Morphol 72: 71–96.

Goldblatt P, Hoch PC, McCook LM. 1992. Documenting scientific data: The need for voucher specimens. Ann MO Bot Gard 79: 969–970.

Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. Syst Biol 54:852–859.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwach W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc Natl Acad Sci USA 12:14812–14817.

Hunter SJ, Goodall TI, Walsh KA, Owen R, Day JC. 2008. Nondestructive DNA extraction from blackflies (Diptera: Simuliidae): Retaining voucher specimens for DNA barcoding projects. Mol Ecol Resources 8:56–61.

International Commission on Zoological Nomenclature. 1999. International code of zoological nomenclature. 4th ed. London: International Trust for Zoological Nomenclature. p 306.

Jördens J, Struck T, Purschke G. 2004. Phylogenetic inference regarding Parergodrilidae and *Hrabeiella periglandulata* ("Polychaeta", Annelida) based on 18S rDNA, 28S rDNA and *COX1* sequences. J Zool Syst Evol Res 42:270–280.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Kvist S, Dreyer J, Erséus C. 2008. Two new species of *Tubificoides* (Annelida: Clitelata: Naididae) from the Blake Ridge methane seep in the northwest Atlantic Ocean. Proc Biol Soc Wash 121: 531–540.

Kvist S, Sarkar IN, Erséus C. 2010. Genetic variation and phylogeny of the cosmopolitan marine genus *Tubificoides* (Annelida: Clitellata: Naididae: Tubificinae). Mol Phylogenet Evol 57:687–702.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Linnaeus C. 1758. Systema naturae. Regnum animale Vol. 1. 10th rev. ed. L. Salvii, Holmiae.

Miller SE. 2007. DNA barcoding and the renaissance of taxonomy. Proc Natl Acad Sci USA 104:4775–4776.

Moquin-Tandon A. 1827. Monographie de la famille des Hirudinées. Montpellier. France: Maison de Commerce. p 141.

Muller KJ, Nicholls JG, Stent GS, editors. 1981. Neurobiology of the leech. Cold spring harbor. New York: Cold Spring Harbor Laboratory Press. p 255.

Neigel J, Domingo A, Stake J. 2007. DNA barcoding as a tool for coral reef conservation. Coral Reefs 26:487–499.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Köljalg U. 2006. Taxonomic reliability of DNA sequence databases: A fungal perspective. PLoS ONE 1:e59.

Oceguera-Figueroa A, Léon-Règagnon V, Siddall ME. 2005. Phylogeny and revision of Erpobdelliformes (Annelida, Arhyncobdellida) from Mexico based on nuclear and mitochondrial gene sequences. Rev Mex Biod 76:191–198.

Pleijel F, Jondelius U, Norlinder E, Nygren A, Oxelman B, Schander C, Sundberg P, Tholleson M. 2008. Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. Mol Phylogenet Evol 48:369–371.

Ratnasingham S, Hebert PDN. 2007. BOLD: The barcode of life data system. Mol Ecol Notes 7:355–364.

Reid GM. 2009. Carolus Linnaeus (1707–1778): His life, philosophy and science and its relationship to modern biology and medicine. Taxon 58:18–31.

Ross HA, Murugan S. 2006. Using phylogenetic analyses and reference datasets to validate the species identities of cetacean sequences in GenBank. Mol Phylogenet Evol 40:866–871.

Rowley DL, Coddington JA, Gates MW, Norrbom AL, Ochoa RA, Vandenberg NJ, Greenstone MH. 2007. Vouchering DNA-barcoded specimens: Test of a nondestructive extraction protocol for terrestrial arthropods. Mol Ecol Notes 7:915–924.

Schander C, Willassen E. 2005. What can biological barcoding do for marine biology? Mar Biol Res 1:78–93.

Siddall ME, Trontelj P, Utevsky SY, Nkamany M, Macdonald KS, III. 2007. Diverse molecular data demonstrate that commercially available medicinal leeches are not *Hirudo medicinalis*. Proc R Soc B 274:1481–1487.

Siddall ME, Fontanella FM, Watson SC, Kvist S, Erséus C. 2009. Barcoding bamboozled by bacteria: Convergence to metazoan primer targets by marine microbes. Syst Biol 58:445–451.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (*and other methods), ver. 4.0b10. Sunderland, MA: Sinauer Associates.

Trontelj P, Utevsky SY. 2005. Celebrity with a neglected taxonomy: Molecular systematics of the medicinal leech (genus *Hirudo*). Mol Phylogenet Evol 34:616–624.

Whitaker IS, Izadi D, Oliver DW, Monteath G, Butler PE. 2004. *Hirudo medicinalis* and the plastic surgeon. Br J Plast Surg 57: 348–353.

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

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

Williams ST, Knowlton N. 2001. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. Mol Biol Evol 18:1484–1493.

Will KW, Mishler BD, Wheeler QD. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst Biol 54: 844–851.