The unholy trinity: taxonomy, species delimitation and DNA barcoding

Rob DeSalle*, Mary G. Egan and Mark Siddall

Division of Invertebrate Zoology, American Museum of Natural History, 79th Street at Central Park West, New York, NY 10024, USA

Recent excitement over the development of an initiative to generate DNA sequences for all named species on the planet has in our opinion generated two major areas of contention as to how this 'DNA barcoding' initiative should proceed. It is critical that these two issues are clarified and resolved, before the use of DNA as a tool for taxonomy and species delimitation can be universalized. The first issue concerns how DNA data are to be used in the context of this initiative; this is the DNA barcode reader problem (or barcoder problem). Currently, many of the published studies under this initiative have used tree building methods and more precisely distance approaches to the construction of the trees that are used to place certain DNA sequences into a taxonomic context. The second problem involves the reaction of the taxonomic community to the directives of the 'DNA barcoding' initiative. This issue is extremely important in that the classical taxonomic approach and the DNA approach will need to be reconciled in order for the 'DNA barcoding' initiative to proceed with any kind of community acceptance. In fact, we feel that DNA barcoding is a misnomer. Our preference is for the title of the London meetings—Barcoding Life. In this paper we discuss these two concerns generated around the DNA barcoding initiative and attempt to present a phylogenetic systematic framework for an improved barcoder as well as a taxonomic framework for interweaving classical taxonomy with the goals of 'DNA barcoding'.

Keywords: DNA barcoding; taxonomy; species delimitation; muntjac; leeches; sturgeon

1. INTRODUCTION: BUILDING A BETTER DNA BARCODER

One of the major issues concerning the inclusion of molecular information into taxonomic aspects of biology that has yet to be discussed in detail in the commentaries on this subject is concerning the best way to read the barcodes. There are two separate tasks to which DNA barcodes are currently being applied. The first is the use of DNA data to distinguish between species (equivalent to species identification or species diagnosis) and the second is the use of DNA data to discover new species (equivalent to species delimitation, species description). These two activities differ in the types and amount of data required. Below we highlight some of the issues that may limit the utility of current DNA barcoding endeavours (especially those used for species discovery) and suggest a framework for the development of a barcoder that addresses these issues.

(a) The barcoder engine: distances or characters?

A major issue that needs to be resolved is how to read the organismal barcode once it is generated. Most recently published approaches to DNA barcoding have utilized distance measures to make the inference as to species designation (Hebert *et al.* 2003*a*,*b*, 2004*a*,*b*). Distances are used in two major approaches; the first is a simple BLAST (Altschul *et al.* 1990) approach where a raw similarity score is used to determine the nearest neighbour to the query sequence. The second approach utilizes distances in tree building (Hebert *et al.* 2003*a*,*b*). We point out the following shortcomings with these approaches and further suggest that character based approaches are more appropriate for DNA barcoding both for theoretical and for practical reasons.

A major shortcoming of using distances in DNA barcoding is that all classical studies and taxonomic schemes that accomplish the same thing that barcodes are meant to accomplish are character based, making the union of classical and DNA barcoding a difficult process if the use of distances is continued in barcoding studies (see below). This shortcoming also is related to the need for diagnostic characters that classical studies use to validate the existence of a species. A second shortcoming is that similarity scores often do not give the nearest neighbour as the closest relative (Koski & Golding 2001). Nevertheless, similarity scores will always give a nearest neighbour. Character based methods have the logical advantage that when diagnostic character data are lacking, they will fail to diagnose, allowing for a degree of hypothesis testing not available when using distances. A third shortcoming involves the lack of an objective set of criteria to delineate taxa when using distances. For example, a universal similarity cut-off to determine species status will simply not exist, because of the broad overlap of inter- and intra-specific distances (Goldstein *et al.* 2000). Researchers will have to constantly revise their similarity cut-offs from group to group. We suspect that

distance-based criteria for different species groups within genera often will have different parameters, making the delineation of species using distances fairly subjective.

We suggest that an alternative approach including character based phylogenetic analysis is more appropriate for establishing or 'printing' barcodes. The character based approach is compatible with classical approaches allowing the combination of classical morphological and behavioural information. Character based approaches sidestep the nearest neighbour problems of distances because they can reconstruct hierarchical relationships where common ancestry is inferred when two entities share derived characters. Neither BLAST (Altschul *et al.* 1990) nor neighbour joining (NJ; Saitou & Nei 1987) tree building approaches allow for character-by-character diagnoses on branches of trees. Any such diagnosis would need to be Parsimony or Maximum Likelihood based. Furthermore, the diagnosis of two separate entities in nature can be accomplished by the existence of a single character shared by a group of organisms to the exclusion of others, whether it be a DNA character or a morphological character.

(b) *The barcoder engine: to tree or not to tree, that is the question*

Given that character based information is a viable alternative to the distance-based approaches already implemented in barcoding studies, the question arises which approach to analysis of characters is more appropriate for barcoding—the non-tree based population aggregation analysis approach (PAA; Davis & Nixon 1992) or a tree based approach. There are several drawbacks to the use of tree-building approaches to species identification. The first relates to the use of distances (described above) to construct trees. But problems with tree building are not limited to trees constructed with distance data. Rather, the second drawback is in the use of single gene trees as evidence of phylogenetic relationships. Several studies have demonstrated both theoretically (Kluge 1989) and empirically (Rokas *et al.* 2003; Gatesy *et al.* 2004) that combined analyses of multiple data partitions yield better representations of evolutionary history than single gene trees. The combined analysis approach has the additional advantage that it allows for the exploration of the character contribution of data partitions to the combined tree and can reveal character support for the combined tree that was not evident in separate analyses of individual gene partitions (Baker & DeSalle 1997; Baker *et al.* 1998; Gatesy *et al.* 1999, 2002, 2003). From these advantages it could be argued that a corroborated total evidence tree could be used as a guide tree for identifying the phylogenetic affinities of an unknown individual's sequence, assuming that the query sequence is one of the gene regions used to construct the total evidence tree. There are bioinformatics tools to aid in the placement of a query sequence based on the presence of shared characters that are diagnostic for nodes on the tree (Sarkar *et al.* 2002). However, there is a third drawback to the use of a tree building approach to species identification. This relates to the use of

hierarchical methods (tree building) and terminology (monophyly as a criterion for species delimitation) when the underlying system (of individuals and populations) is not a hierarchical system of ancestor–descendant relations, redefinitions of monophyly as reciprocal monophyly (Avise *et al.* 1987; Avise 1989; Avise & Ball 1990) or exclusivity (Baum 1992; Baum & Donoghue 1995; Baum & Shaw 1995) notwithstanding (reviewed in Goldstein & DeSalle 2000). A more practical alternative is the exploration of character diagnostics in the sequences themselves without reference to trees. This mirrors the two-step procedure of traditional taxonomic studies in which relationships among species are assessed only after the terminals in the analysis (in this case, species) are first identified by diagnostic characters. In this approach as formulated by Davis & Nixon (1992), sequences are examined using PAA (Davis & Nixon 1992). Diagnostics are accepted if they are fixed and different from aggregate to aggregate of organisms—such diagnostics are termed 'pure' (Sarkar *et al.* 2002). This approach and its relevance to species delimitation has been discussed at length (Davis & Nixon 1992; Goldstein & DeSalle 2000; Goldstein & DeSalle 2003; Goldstein *et al.* 2000; Nixon & Wheeler 1990) and its relevance to diagnosing entities in nature has been discussed both from the technical and theoretical standpoints (Cracraft 1983, 1989; Frost & Kluge 1994; McKitrick & Zink 1988). Some tree based methods attempt to aggregate terminals on the basis of character distribution (Brower 1999) or on tree topology (Wiens & Penkrot 2002) and these are an improvement over distance based tree methods; however, for the DNA barcoding with multiple individuals within a species we feel it inappropriate to use a tree based approach (Davis & Nixon 1992; Goldstein & DeSalle 2000).

(c) *The barcoder database: is cox1 enough?*

A controversial aspect of the DNA barcoding initiative has been which molecular tool to use to generate the DNA barcodes (Prendini 2005). The published efforts so far in animal systems have used the cytochrome *c* oxidase subunit I gene (*cox1*) of the animal mitochondrion. One of the major criticisms of this approach is that a single molecular probe such as *cox1* will not necessarily provide sufficient information to deliver the resolution needed to diagnose the large number of species targeted by the initiative. In arguing for the sufficiency of *cox1* (or any other single molecular marker), Hebert *et al.* (2003*b*) pointed out that just 15 variable sites in the cox1 gene offers 1 billion different combinations of bases giving more than enough possible barcode 'patterns' at the DNA level. Missing from that assertion is a recognition that relatively few of those combinations could ever result in a viable translated protein observable in an extant species. Based on a study of birds (Hebert *et al.* 2004*b*) it was suggested that *cox1* might have broader utility across the animal kingdom and that a universal distance cut-off of 10 times the distance within species could be used to distinguish between species. However, even among the birds surveyed for *cox1*, there were anomalous taxa that showed greater than expected within species divergences. In addition, studies of copepods
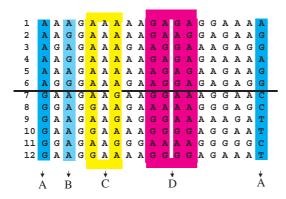
Figure 1. Hypothetical example of character based diagnosis (Davis & Nixon 1992) in action. The twelve sequences represent two populations of six individuals each. The solid line through the middle of the matrix represents a geographical barrier between the two populations. A. DNA sequence attributes in these columns are purely diagnostic characters (sensu Davis & Nixon 1992). B. DNA sequence attributes in this column are not purely diagnostic, but rather the G in the three individuals in the top population are 'private' to that population. C. The DNA sequence attributes in the two columns by themselves constitute two private DNA positions. However, in combination these two columns provide a 'pure' diagnostic combination (AA versus AG or GA; 'compound pure' character in the terminology of Sarkar *et al.* 2002). D. The four columns marked by the shading for D are neither diagnostic nor private. Yet in combination the four columns provide a diagnostic system for the top population versus the bottom. The top population is diagnosed by GA, AG/GA, AG for the four columns.

(Edmands 2001; Goetze 2003) have found high levels of *cox1* variation (up to 20%) even among conspecifics. Having to leave aside these outliers argues against the sufficiency or the universality of the gene region. While the fact that these taxa showed great divergence in genetic distance is suggestive that there may be unrecognized taxonomic diversity present, to test that hypothesis would require more than one line of evidence.

So we examine here the character based approach to diagnosis and the power of character-based approaches. Sarkar *et al.* (2002) recognized that combinations of attributes that are not 'pure' diagnostics could indeed be used to develop compound 'pure' diagnostics. The simplest compound diagnostic is when two attributes that are 'private' for aggregates (found only in one aggregate but not fixed; e.g. if aggregate 1 is fixed for all individuals at position 1 of a sequence with a G and aggregate 2 has 5 individuals with a G in position 1 and 5 individuals with an A in position 1, the A in aggregate 2 is a private diagnostic for aggregate 2) are combined to produce a 'pure' diagnostic. Even more complex combinations can be found if two or more aggregates are defined. For instance, figure 1 shows four positions in a hypothetical sequence that are polymorphic (i.e. neither fixed nor private for the alternative character states), which when combined together create a pure diagnostic. This approach has been used to evaluate character diagnostics in sturgeon species. The system examined was generated from comparisons of over 150 Acipenseridae individuals in two species, *Acipenser gueldenstaedti* and *A. baerii* (Doukakis *et al.* 1999). While the molecular

probe used was 700 base pairs (bp) of cytochrome B gene (cytB) of the mitochondrion, this example will suffice to demonstrate the power of finding diagnostics using this approach. Between these two species for the 700 bp region of cytB, we observed 36 variable sites, of which three were 'pure' diagnostics for the two species. Nearly half of the sites were 'private' to one species and over 1000 combinations of two sites produced compound 'private' sites (i.e. the two sites together were 'private' to one of the species). More interestingly, there were seven combinations of these 15 singly 'private' sites and the 1000 two position compounds that produced 'pure' diagnostics.

The answer to 'Is cox1 enough?' is then yes and no. Cox1 is certainly not enough to delineate phylogenetic relationships of organisms. However, it may be enough to generate suites of characters that can and will diagnose aggregates of organisms as entities in nature.

(d) *The barcoder database: how many individuals are enough?*

Another controversial aspect of the DNA barcoding initiative relates to the number of individuals of each putative species to include in the analysis. Classical taxonomic endeavours screen numerous individuals from multiple localities across the range of a given species to distinguish variation within a species from variation between species in order to identify those characters that are uniquely shared among all members of a species. One or only a few individuals may not be representative of the species as a whole, especially for taxa with widespread distributions (Davis & Nixon 1992; Goldstein *et al.* 2000; Walsh 2000). The necessity for adequate numbers of individuals applies to both distance and character based methods and it is not likely that there will be a universal sample size that will be appropriate for all species. Neither is a universal geographic distance likely to provide a reasonable proxy for determining the appropriate sampling strategy. As with gene region choice, sampling sufficient numbers of individuals to capture representative within-species variation will require pilot studies and the use of background information on life history, dispersal ability and mating patterns, among other information.

2. A CHARACTER BASED BARCODER PROPOSED

Recent work and commentary on the barcoding initiative in the literature has stimulated concern and excitement both from taxonomists and from those who are based in molecular approaches. Concerns range from the philosophical to the technical. These commentaries could loosely be separated into the taxonomy perspective (Agosti 2003; Dunn 2003; Lipscomb *et al.* 2003; Proudlove & Wood 2003; Seberg *et al.* 2003), the molecular perspective (Baker *et al.* 2003; Blaxter & Floyd 2003; Ronquist & Gardenfors 2003; Tautz *et al.* 2003) and commentaries sympathetic to both (Mallet & Willmott 2003; Wilson 2003). The pro-taxonomy commentaries strongly deride the lack of consideration of the intellectual content of classical taxonomy and can be summarized as in the following quote from Lipscomb *et al.* (2003) p. 65: 'advocates of

DNA taxonomy seem not to understand the peerless intellectual content of taxonomy based on all available information, or the hypothesis-driven basis of modern revisionary work.'

It is clear to us that genomic information should be an active component of modern taxonomy, but DNA should not be the sole source of information retrieval. 'Fashionable DNA bar-coding methods are a breakthrough for identification, but they will not supplant the need to formulate and rigorously test species hypotheses.' (Wheeler *et al.* 2004, p. 285). A barcode should incorporate diagnostic characters both from the classical morphological approach and from the newer molecular approaches; one without the other misses the synergy that an integrated taxonomy is capable of attaining (Godfray 2002). We see a major strength to an integrated approach in that descriptive taxonomy and phylogenetic taxonomy together produce a synergy of resolution that neither can attain in the current fragmented 'tower of babel' (Mallet & Willmott 2003). It should also be clear that integration of the 'fashionable' molecular approaches with the classical taxonomic approach is a critical component of reconciling both camps and to move towards the use of barcodes in modern biology. Consequently we present an operational, integrative approach to taxonomy that attempts to reconcile molecular information with other sources of characters.

(a) The taxonomic circle; breaking out

We offer figure 2 as a heuristic for how modern taxonomy can be viewed. While any diagram describing the workings of taxonomy would suffer from over-simplification of the intellectual process that taxonomists use in plying their trade, we feel figure 2 captures many elements of modern taxonomy—hypothesis testing, corroboration, reciprocal illumination and revision. The main problem that needs to be addressed in any attempt to determine the boundary of a species and hence raise the entity to species status is to avoid circular or tautological reasoning. Breaking out of the circle of inference (figure 2, central diagram) in species delineation work is one descriptive way to describe the job of the taxonomist and hence the role of DNA sequence information (and barcoding) in taxonomy.

Figure 2 shows a highly simplified version of several taxonomic problems that have faced systematists and DNA barcoders. The classical process of using morphology in taxonomy is shown first (figure 2, panel A). In this diagram the data points on the 'taxonomy' circle consist of geographical, morphological, ecological, reproductive and behavioural information. In most morphological taxonomic studies an initial hypothesis based on geography is made. The taxonomist then crosses over the interior of the circle to either ecological characters or to morphological characters to test the geographical hypothesis. If morphological, behavioural, reproductive or ecological information relevant to the geographical hypothesis assist in rejecting the null hypothesis that there is no differentiation of the two geographical entities, then the taxonomist can 'break out' of the circle.

Cryptic species detected by DNA approaches is shown next (figure 2, panel B). In this case we add

DNA sequence information to the circle. Initially a geographical hypothesis is formulated, a null hypothesis established and tested with the classical tools of the taxonomist. In this case, none of the classical tools—reproductive biology, morphological, behavioural or ecological characters—can reject the null hypothesis. The taxonomist can turn to DNA sequences where the null hypothesis based on geography is rejected because of fixed DNA differences among the aggregates hypothesized by geography. In essence, the aggregates contain morphologically cryptic species that are only detected at the DNA sequence level, which allows the taxonomist to break out of the circle.

The third panel in figure 2 (panel C) shows the case of lack of ability by all methods to lead to rejection of a geographical null hypothesis. In this case, the putative species entities suggested by geography cannot be corroborated and hence the taxonomist is constrained to remain in the circle. The conclusion by the taxonomist should be that there is a single taxonomic entity. The fourth panel (figure 2, panel D) represents the power of integrating novel methods into this operational scheme. In this case, several individuals within a single geographic area show morphological differences. Because these individuals are considered to reside in the same geographic region, a geographical hypothesis cannot be made. But in this case the morphologically different entities can be aggregated and tested for fixed differences with other sources of data. In the case in the diagram, we imply that DNA sequence information can be used and if fixed DNA differences corroborate the morphological hypothesis then the conclusion of the analysis is that two species exist in sympatry and can be delineated by morphological differences.

The converse situation is also possible—a researcher could examine a 'population' of organisms with morphology and see no morphological differences. When the genomes of the organism are examined, the researcher might discover a DNA sequence polymorphism that clearly separates the single population into individuals with one haplotype and individuals with a distinct second haplotype. The only way to break out of the circle here would be to re-examine morphology or to move on to some other source of information. If no corroboration of the molecular aggregation can be found then the conclusion should be that a single population with two clearly distinct haplotypes exist. If corroboration is attained, then two distinct entities should be concluded to exist.

3. THE CHARACTER BASED (BREAKOUT) BARCODER IMPLEMENTED

We feel that a formalized method for inclusion of molecular information into taxonomy will clarify the intellectual content of taxomony from a molecular perspective, but it will also clarify how DNA sequence information can most efficiently be used in the DNA barcoding initiative. The following section uses a mammalian case study (the genus *Muntiacus* or barking deer), a fish case study (the Acipenseridae or sturgeons—the source of caviar) and invertebrate examples (leeches)—that have manageable numbers
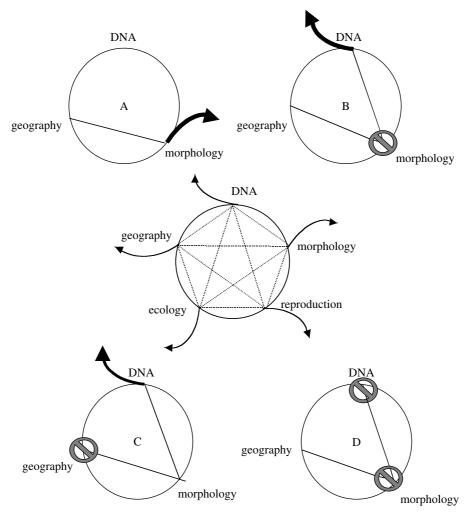
Figure 2. The taxonomic circle. The dotted lines that traverse the inner part of the circle indicate experimental routes that can be taken in taxonomic endeavour to accomplish corroboration of taxonomic hypotheses. The only way to delineate a new taxon is to break out of the circle (the solid arrows emanating from the circle). In our scheme, it only takes one traversal of the interior of the taxonomic circle where corroboration occurs in order for the taxonomist to 'break out' of the circle and designate a taxon. Examples of use of the taxonomic circle using hypothetical examples. A. Classical morphological taxonomy; a taxonomic hypothesis is established on the basis of organisms appearing to be similar at a particular geographic locality. The taxonomic hypothesis is tested with morphological information and corroborated with the morphological attributes. The morphological attributes then become diagnostic characters if they corroborate the geographical hypothesis. B. Cryptic species in taxonomy; a geographical hypothesis is posed and tested with morphology. The morphological attributes collected do not corroborate the geographical hypothesis and hence the taxonomist cannot 'break out' of the circle. Retaining the geographical hypothesis the taxonomist then examines the aggregates established using the geographical hypothesis with DNA sequence data and corroboration ensues with DNA sequence characters being diagnostic. C. Sympatric species in taxonomy; morphological differences are recognized among a group of organisms. A hypothesis of aggregation is posited based on the morphological information. When geographical distributions are used to test the aggregation patterns, there is no geographic pattern to the distribution of the different morphological types. The taxonomist then uses DNA sequence attributes to test the morphological hypothesis of aggregation and corroborates the morphological hypothesis and the taxonomist 'breaks out' of the circle. D. Failure to detect a new taxon; in this example, a geographic hypothesis is made and tested with morphological information. The morphological information fails and the geographical and any morphological hypotheses of aggregation are then tested with DNA sequence information. DNA sequence information fails to reject the hypothesis of no new taxa and, hence, the taxonomist cannot 'break out' of the circle and the inference is that there are no differentiable aggregates and hence only a single taxon.

of taxa and unique taxonomic problems to examine the incorporation of molecular data into taxonomical issues. Specifically, we first examine the use of type specimens in barcoding of muntjac and the impact such type specimens will have on future DNA taxonomic efforts. Second we examine the use of DNA barcoding in the leech genus *Hirudo* to demonstrate the importance of broad scale sampling of groups in taxonomic surveys. *Hirudo* also serves as a strong example of the effect of hybridization on how taxonomy is done. Finally, we use the commercially important

fish family Acipenseridae to examine a cryptic species problem. Each of these examples will demonstrate that while DNA is an important factor in all three, the interaction of DNA sequence data with other kinds of characters produces a more precise taxonomical framework.

(a) *Muntjac barcoding:* **Muntiacus rooseveltorum** *example*
The muntjac study (Amato *et al.* 1999) used DNA barcoding for species discovery in a framework that is

compatible with classical taxonomy. The study highlights some of the issues related to barcoding such as the use of independently identified vouchers, sample size and gene region choice.

It began with field reports of what may have been representatives of a new species of muntjac in Laos for which the only material available at the time consisted of dried tissue samples. In order to devise a method to explore the question of the species status of these muntjac using DNA, the study looked to established practices in taxonomy for guidelines. Sample size and gene region choice are factors that can affect the ability to discern true diagnostic data. In order for DNA data to have the potential to be used in a species discovery process, all species in the group must be included in the analysis. In addition, the numbers of individuals sampled for each species must be large enough to be representative of the variation in a given gene region for the species as a whole and the gene region must be variable enough to detect true differences between species. Given a large enough and representative sample size, the problem of a too variable gene region leading to incorrectly rejecting the null can be minimized.

The number of available samples for the putative new species was low ($n=10$). Given the low number of individuals, a pilot study was undertaken that explored several gene regions for representatives of each species in the group and chose a gene region (in this case 16S mt rDNA) to balance the potential for two types of errors: (1) mistaking individual variation for species level variation by using too few individuals and a highly variable gene region; or (2) failing to identify true species differences, by using a conserved gene region sequenced for too few individuals to recover sufficient variation.

A diagnosis matrix of 114 individuals representing all species in the genus plus outgroup taxa was constructed. For the more widespread species, this included larger sample sizes from several localities across the range of those species as in the case of *Muntiacus muntjak* ($n=49$). In addition, in order to be able to associate newly collected material with species names, for each species DNA sequence was obtained from named, morphologically examined specimens from museum collections and included in the diagnostic matrix. The sequences of the putative new species compared to other species in the group were unique. However, the inclusion of sequence data from museum specimens proved critical to an accurate assessment of the species status. In the literature there had been a description of a species (*M. rooseveltorum* Osgood 1932) based on a single individual collected in Laos ~70 years ago and never collected again. The DNA from the Type specimen of *M. rooseveltorum* was shown to share diagnostic sites with all the newly collected specimens of the putative new species (figure 3). This led to the conclusion that the newly collected specimens represented a rediscovery of *M. rooseveltorum*. The use of DNA alone, without the inclusion of all species in the group (and the type specimen in this case), would have led to the incorrect conclusion that the *M. rooseveltorum* specimens represented a new species. This highlights the importance of complete taxonomic sampling, literature review and corroboration with a second line of evidence. It also highlights the need to obtain voucher specimens that would provide not only DNA but also provide the means to examine a second line of evidence such as morphology.

(b) *Leech barcoding*

For a variety of reasons, leeches provide a unique framework for examination of the utility of DNA barcoding methods. The group (Hirudinida) is well circumscribed with approximately 750 known species. There is, nonetheless, a diversity of habitat preferences and life-history strategies represented across the clade including some extremes of parental care, various trophic modes ranging from blood feeding to predation as well as life in marine, freshwater or terrestrial environments. Furthermore, leeches already are well characterized for the *cox1* locus (e.g. Siddall & Burreson 1998; Siddall *et al.* 2001; Borda & Siddall 2004) that is typically advocated for barcoding studies (Hebert *et al.* 2003*a*,*b*, 2004*a*,*b*). However, the use of *cox1* alone for barcoding leech diversity may warrant some caution in that this locus has a highly biased base composition. Among the New World medicinal leeches, for example, adenosine and thymidine represent up to 72% of the nucleotide composition in *cox1* (and up to 96% at third positions; Siddall & Burreson 1998). As a result, approximately 24% of the variable sites in *cox1* are rendered binary among leeches as opposed to having all four nucleotides available.

(i) *Glossiphoniidae example*
One of the best-represented families of leeches in freshwater environments is Glossiphoniidae. The clade comprises taxa that are dorsoventrally flattened freshwater species normally found feeding on anuran or chelonian hosts, though a few are fish parasites and one (*Placobdelloides jaegerskioeldi*) is even specific to the rectal tissues of hippos (*Hippopotamus amphibius*). However, many species in the family, in particular those in the unrelated genera *Glossiphonia* and *Helobdella* have abandoned sanquivory in favour of a predatory lifestyle on mollusks and oligochaetes, respectively. Species in the genus *Helobdella* have their greatest diversity in South America and less so in North America. Since being described by Linnaeus, *Helobdella stagnalis* was the only species known from Europe, and none were known to occur either in Africa or Australia. As such, recent descriptions of new species like *Helobdella europaea* from an urban pond in Berlin (Kutschera 1985, 1987) and *Helobdella papillornata* from streams in Australia (Govedich & Davies 1998) were entirely unexpected, as was the surreptitious discovery of undescribed representatives of *Helobdella* in each of South Africa, Hawaii and New Zealand, all in a span of three years. Siddall and Budinoff (2005) employed DNA barcoding to assess this distribution of leeches, which at first presented a historical biogeographic conundrum. Their results (reproduced here in figure 4) clarified the fact that in each case the leeches were genetically indistinguishable both at the *cox1* and ND1 mitochondrial loci and represented a single species of *Helobdella* nestled in a South American

```
                          1 1 1 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 5
                        1 5 5 9 3 4 5 5 5 6 6 7 7 8 8 9 0 2 2 2 4 4 5 5 5 6 7 9 0 4 0
                        5 2 7 9 8 0 4 9 1 3 6 7 9 3 9 0 9 2 5 9 3 7 1 6 7 1 9 3 6 6 1
M.rooseveltorum Type    T A C T G A C A T T C G T T G T T T A T T A T C C A T A C A T
M.rooseveltorum 1       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.rooseveltorum 2       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.rooseveltorum 3       . . . . A . . . . . . . . . . . . . . . . . . . . C . . . . .
M.rooseveltorum 4       . . . . . . . . . . . . . . . . . . . . . . . . . C . . . . .
M.rooseveltorum 5       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.rooseveltorum 6       : : : . : . . . . . . . . . . . . . . . . . . . . C . . . . .
M.rooseveltorum 7       . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.rooseveltorum 8       . . . . . . . . . . . . . . . . . . . . . . . . . C . . . . .
M.rooseveltorum 9       : . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.rooseveltorum 10      . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
M.putaoensis(N=18)      : : . A . . . . C . . A . . . . . . . W . . C . . . . C . . .
M.truongsonensis(N=3)   : : . . R . . . . . . A . . . . . . . . . . C . . G C . . . .
M.vuquangensis(N=2)     : : Y . . G . . . . T A C . A C . . . . . C . . T . . C . . .
M.crinifrons(N=21)      : : . . . . . . . . T A . . A C Y . G . . . . . . . . C . . G .
M.feae(N=2)             : G . . . . . . . . T A . . . C . . G . . . . . . . . C . T G .
M.reevesi(N=3)          C . . . . . . . . . T A . C . C . C . . . . . . . . . C T . . C
M.muntjak(N=49)         : : T . R . . . Y . T A . Y . . Y . . . . R . . . Y . C . . . .
M.atherodes(N=1)        : : : : : : : : : : : : : : : : : : : : : : : T . . C T . . N
Elaphodus cephalophus(N=2) : . . . . . T T Y A . A C . A C . . . . . . . . Y . C . . . C
Cervus eldi(N=2)        C . . . . . . C . A T A C . . C . C . C G A . . . . T . . . . C
```


$$G]G]C]GGT$$
276   390   306   57   110   183

Figure 3. A DNA barcoding example for barking deer (genus *Muntiacus*). The table at the top of the figure shows variable nucleotide positions including several diagnostic sites in the 16s mt rDNA of multiple individuals of muntjac species. DNA from the type specimen of *Muntiacus rooseveltorum* was compared to recently collected putative *M. rooseveltorum* specimens to clarify their nomenclature (Amato *et al.* 1999). The word Type after the binomial indicates the sequence obtained from the type specimen of *M. rooseveltorum*. Shaded area indicates nucleotide position diagnostic for *M. rooseveltorum*. Dots (.) indicate sequence identity to the reference sequence on the first line. Colons (:) indicate missing data. Position 1 in the region sequenced corresponds to position 2305 in the *Bos taurus* mitochondrial DNA, complete genome (GenBank Accession Number: AB074962). Photograph of the skull of the Type specimen of *M. rooseveltorum* courtesy of the Field Museum of Natural History (Field Museum negative number Z82184: *Muntiacus rooseveltorum* Zoology specimen 31783), is used with permission. The graphic in the centre shows the multiple gene region barcode for *M. rooseveltorum* separated by right brackets; reading from left to right, it shows the diagnostic nucleotides and position numbers found in the mitochondrial gene regions: 16s, cytochrome *b*, 12s and D loop.

```
Helobdella europaea
    Helobdella "papillornata" type locality      TGG G
    Helobdella "papillornata" Madill Creek       TGG G
    Helobdella NZ                                 TGG G
    Helobdella ZA                                 TGG G
    Helobdella europaea                           TGG G

Helobdella paranensis                            AAA A
Helobdella fusca                                 AAA A
Helobdella lineata                               AAA A
Helobdella modesta                               AAA T
Helobdella stagnalis                             AAA T
Helobdella transversa                            AAA A
Helobdella cordobensis                           AAA A
Helobdella elongata                              AAA A
Helobdella papillata                             AAA A
Helobdella nunununununojensis                    AAA A
Helobdella sorojchi                              AAA A
Helobdella ringueleti                            CAA C
Helobdella bolivianita                           AAA T
Helobdella triserialis                           AAA A
Helobdella michaelseni                           AAC A

Haementeria gracilis                             AAA T
Haementeria lutzi                                AAA A
Haementeria ghilianii                            AAA T
Haementeria tuberculifera                        AAA T
Haementeria molesta                              AAA T
```

Figure 4. A DNA barcoding example from the leech species *Helobdella europaea* (Glossiphoniidae). The table shows diagnostic sites for *Helobdella europaea* (figured at top). Diagnostic characters, reading left to right, are position numbers 273, 471 and 501 in *cox1*, and 1160 in ND1. DNA barcoding of a broad geographic sampling along with morphological dissections and comparison to described species allowed Siddall & Budinoff (2005) to clarify the nomenclature of this species and shed light on possible explanations for an unexpected geographic distribution.

group known as the *triserialis* complex (Ringuelet 1943). Accidental introductions to each locality could easily have been coincident with introductions of common aquatic invasive plant species like *Pistia stratiotes* and *Salvinia molesta*, known to have happened in each of Germany, Australia, New Zealand and Hawaii. The genetic determination alone, however, was not accomplished in isolation from taxonomic considerations. Rather, through dissections and comparison to described taxa, Siddall & Budinoff (2005) asserted that this leech species corresponds exactly to Ringuelet's (1943) *Helobdella triserialis* var *lineata*; an unfortunate result since Verrill's (1872) North American *Helobdella lineata* preoccupied that appropriate specific epithet. The globally invasive leech species, thus, is now known as *Helobdella europaea* notwith-

standing its suspected South American origin. Significant to the successful barcoding result in the foregoing was a broad geographic coverage of the species of *Helobdella* from the known range of the genus. Without that global coverage, and by, for example, only focusing collections on a restricted geographic area in which the suspect leech was found (e.g. Hebert *et al*. 2004*a*) probably would have abrogated discovery of its true identity and ultimate origin.

(ii) *Hirudo example*
The European medicinal leech remains a valuable tool available to the biomedical sciences notwithstanding its having historically been used for some rather dubious purposes like the treatment of obesity and Stalin's fatal strokes. In fact, just this past year the US Food and

Drug Administration formally approved the European medicinal leech (*Hirudo medicinalis*) as a 'medical device' and several companies like LeechesUSA, BioPharm and Ricarimpex specialize in the global distribution of leeches for use in microsurgery and related procedures. One would think that a species of annelid that is so broadly used in medicine, neurobiology, developmental biology and genomics, and for which various genomic libraries are being developed would have been better characterized in terms of its species limits. The first phylogenetic analyses to incorporate the European medicinal leech as a taxon using *cox1* were Black *et al.* (1997) and Siddall & Burreson (1998), though neither of those analyses considered multiple representatives of the species. More recently, Trontelj & Utevsky (2005) demonstrated several unusual findings on the basis of *cox1*: specifically, that so-called European medicinal leeches group into four distinct lineages and that what Black *et al.* (1997) and Siddall & Burreson (1998) each sequenced bears little resemblance to the *cox1* gene found in wild-caught European medicinal leech populations. Revisiting their analysis here, we have reanalysed the available data with some wild-caught and commercially available material in a broader taxonomic scope for the Hirudinidae (figure 5). Notably, the results corroborate the findings of Trontelj & Utevsky (2005) in that the European medicinal leech species complex seems to include at least four species, three of which previously had been synonymized with *Hirudo medicinalis*. If DNA barcoding results are accepted as is, then we would need to resurrect each of *Hirudo verbana* Carena 1820, *Hirudo troctina* Johnson 1816, and establish a new species for the Persian medicinal leech denoted '*Hirudo* sp.' in figure 5. Conveniently, each of these species may not be as 'cryptic' as previously (Sawyer 1986) thought insofar as they appear to be readily distinguishable on the basis of external colour patterns. More alarmingly, though, is the status of those leeches that previously have been called '*Hirudo medicinalis*'. For example, a leech obtained from Ward's Biological for this study and shipped under the name of '*Hirudo medicinalis*' unequivocally groups with *Hirudo verbana* (figure 5). Also, both Black *et al.* (1997) and Siddall & Burreson (1998) obtained their representatives of *Hirudo medicinalis* from Carolina Biological supply, and both of those sequences group with the Asian *Hirudinaria manillensis* notwithstanding the fact that the specimen used by Siddall & Burreson (1998) is morphologically indistinguishable from *H. verbana*. The latter suggests a remarkable ability for introgression that is as yet not well understood for these leeches, and yet which should cause some concern for the overall utility of DNA barcoding methods based on a single locus.

(c) Sturgeon barcoding

There are three species of fish in the family Acipenseridae—*Husu huso*, *Acipenser gueldenstadtii* and *A. stellatus*, all of them listed as endangered by CITES—that are the source for the grand majority of the world's commercial caviar trade. One of these species, *A. gueldenstadtii*, has been an enigma with respect to the surveillance of imported caviar since DNA sequence methods were introduced to monitor importation of caviar from the three highly endangered fish (DeSalle & Birstein 1996; Birstein *et al.* 2000). One of the major problems with the diagnosis of *A. gueldenstadti* caviar has been the occurrence at high frequency of caviar purportedly from *A. baerii* (a close relative to *A. gueldenstadti*), or the Siberian sturgeon (Birstein *et al.* 2000).

More detailed examination of the problem using larger numbers of individuals from both the *A. gueldenstadti* and *A. baerii* clades of sturgeons now indicates the presence of a cryptic species identical to *A. gueldenstadti* in morphology, but also similar (but not identical to) *A. baerii*. In fact, several DNA sequence changes exist that diagnose this aggregate of fish that are morphologically identical to *A. gueldenstadti*, as distinct from *A. baerii*. In this case the DNA diagnostics indicate that this second confused form of *A. gueldenstadti* is a separate entity (Birstein *et al.* 2005). This case is an excellent example of cryptic species and how DNA sequence information can reveal the crypticism. More importantly, this case exemplifies two important technical aspects of DNA barcoding. First, the need for large sample sizes and continual revision using larger sample sizes is highlighted by this example. When small sample sizes are used, the second cryptic *A. gueldenstadti*-like species is improperly diagnosed as *A. baerii*. Second, the case emphasizes the importance of precision in species delimitation in the practical application of any DNA barcoding system. Since animal forensics is a major positive outcome of DNA barcoding, this example reinforces the notion that large sample sizes and comprehensive databases coupled with classical techniques (as in this case meristics) be incorporated to implement the barcoding approach.

4. CONCLUSIONS

We conclude for the following reasons that the non-tree based approaches are more appropriate for the construction of a barcode reader. First, tree based approaches will produce phylogenies based on a single poorly chosen (for phylogenetics) molecule. While the trees will often times make sense, the support for hypotheses from such trees is almost always low, limiting the robustness of any phylogenetic hypothesis from such trees. Related to this issue is the well known widely accepted approach in phylogenetics that uses concatenated data matrices to produce phylogenies (Kluge 1989; Gatesy *et al.* 2003; Rokas *et al.* 2003). To base any inferences of relationship of species on a phylogenetic tree generated from a single molecular marker would be in conflict with the current approaches to modern systematics. Second, current taxonomic approaches use diagnostics discovery independent of trees to establish taxonomic systems. Using DNA characters in a diagnostic context would be entirely compatible with the process of current taxonomic research. Third, our proposed framework, requiring corroboration from more than one line of evidence, is also consistent with current taxonomic practices, would serve as a bridge between morphological and molecular approaches and provides sufficient rigor for species identification and discovery. We readily admit that certain barcoding problems such as environmental microbial species
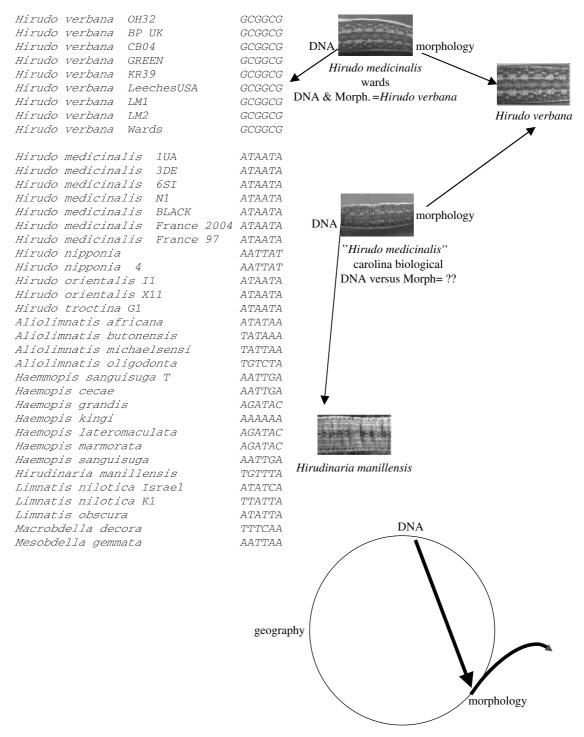
Hirudo verbana	OH32	GCGGCG
Hirudo verbana	BP UK	GCGGCG
Hirudo verbana	CB04	GCGGCG
Hirudo verbana	GREEN	GCGGCG
Hirudo verbana	KR39	GCGGCG
Hirudo verbana	LeechesUSA	GCGGCG
Hirudo verbana	LM1	GCGGCG
Hirudo verbana	LM2	GCGGCG
Hirudo verbana	Wards	GCGGCG
Hirudo medicinalis	1UA	ATAATA
Hirudo medicinalis	3DE	ATAATA
Hirudo medicinalis	6SI	ATAATA
Hirudo medicinalis	N1	ATAATA
Hirudo medicinalis	BLACK	ATAATA
Hirudo medicinalis	France 2004	ATAATA
Hirudo medicinalis	France 97	ATAATA
Hirudo nipponia		AATTAT
Hirudo nipponia	4	AATTAT
Hirudo orientalis	I1	ATAATA
Hirudo orientalis	X11	ATAATA
Hirudo troctina	G1	ATAATA
Aliolimnatis africana		ATATAA
Aliolimnatis butonensis		TATAAA
Aliolimnatis michaelsensi		TATTAA
Aliolimnatis oligodonta		TGTCTA
Haemmopis sanguisuga T		AATTGA
Haemopis cecae		AATTGA
Haemopis grandis		AGATAC
Haemopis kingi		AAAAAA
Haemopis lateromaculata		AGATAC
Haemopis marmorata		AGATAC
Haemopis sanguisuga		AATTGA
Hirudinaria manillensis		TGTTTA
Limnatis nilotica Israel		ATATCA
Limnatis nilotica K1		TTATTA
Limnatis obscura		ATATTA
Macrobdella decora		TTTCAA
Mesobdella gemmata		AATTAA

Figure 5. Diagnostic sites in *cox1* for *Hirudo verbana* (position numbers: 267, 360, 507, 513, 543, and 579). A leech obtained from Ward's Biological for this study and shipped under the name of '*Hirudo medicinalis*' unequivocally groups with *Hirudo verbana*. Also, both Black *et al.* (1997) and Siddall & Burreson (1998) obtained their representatives of *Hirudo medicinalis* from Carolina Biological supply, and both of those sequences group with the Asian *Hirudinaria manillensis* notwithstanding the fact that the specimen used by Siddall & Burreson (1998) is morphologically indistinguishable from *H. verbana*.

identification will be problematic due to the lack of geographical and morphological information for corroboration. However, we suggest that in these problematic cases additional gene regions and ecological information might also be used to support or refute hypotheses of species cohesion.

Finally, when thinking about the possible formats for an actual field usable DNA barcoder, having a diagnostic system would be most appropriate for a small device. The coding of the diagnostics can be included in the design of a microarray format or in a rapid single nucleotide polymorphism detection format. These highly technical molecular approaches utilize character based detection methods, and would bring the development of a small field usable DNA barcoder closer to reality.

REFERENCES

Agosti, D. 2003 Encyclopedia of life: should species description equal gene sequence? *Trends Ecol. Evol.* **18**, 273. (doi:10.1016/S0169-5347(03)00099-5.)

Altschul, S. F., Gish, W., Miller, W., Meyers, E. W. & Lipman, D. J. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1006/jmbi.1990.9999.)

Amato, G., Egan, M. G., Schaller, G. B., Baker, R. H., Rosenbaum, H. C., Robichaud, W. G. & DeSalle, R. 1999 The rediscovery of Roosevelt's barking deer (*Muntiacus rooseveltorum*). *J. Mamm.* **80**, 639–643.

Avise, J. C. 1989 A role for molecular genetic in the recognition and conservation of endangered species. *Trends Ecol. Evol.* **4**, 279–281. (doi:10.1016/0169-5347(89)90203-6.)

Avise, J. C. & Ball, R. M. 1990 In *Principles of genealogical concordance in species concepts and biological taxonomy* (ed. J. Antonovics & D. Futuyma) *Oxford Surveys in Evolutionary Biology*, vol. 7, pp. 45–67. London: Oxford University Press.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. 1987 Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* **18**, 489–522.

Baker, R. H. & DeSalle, R. 1997 Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. *Syst. Biol.* **46**, 654–673.

Baker, R. H., Yu, X. & DeSalle, R. 1998 Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. *Mol. Phylogenet. Evol.* **9**, 427–436. (doi:10.1006/mpev.1998.0519.)

Baker, C. S., Dalebout, M. L., Lavery, S. & Ross, H. A. 2003 www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends Ecol. Evol.* **18**, 271. (doi:10.1016/S0169-5347(03)00101-0.)

Baum, D. A. 1992 Phylogenetic species concepts. *Trends Ecol. Evol.* **7**, 2. (doi:10.1016/0169-5347(92)90187-G.)

Baum, D. A. & Donoghue, M. J. 1995 Choosing among alternative 'Phylogenetic' species concepts. *Syst. Bot.* **20**, 560–573.

Baum, D. A. & Shaw, K. L. 1995 Genealogical perspectives on the species problem. In *Experimental and molecular approaches to plant biosynthesis* (ed. P. C. Hoch & G. D. Stephenson), pp. 289–303. St. Louis Missouri Botanical Garden: St. Louis.

Birstein, V. J., Doukakis, P. & DeSalle, R. 2000 Polyphyly of mtDNA lineages in the Russian sturgeon, *Acipenser gueldenstaedtii*: forensic and evolutionary implications. *Cons. Genet.* **1**, 81–88. (doi:10.1023/A:1010141906100.)

Birstein, V. J., Ruban, G., Ludwig, A., Doukakis, P. & DeSalle, R. 2005 The enigmatic Caspian Sea Russian sturgeon: How many cryptic forms does it contain? *Systematics and Biodiversity* **3**, 203–218.

Black, M. B., Halanych, K. M., Maas, P. A. Y., Hoeh, W. R., Hashimoto, J., Desbruyeres, D., Lutz, R. A. & Vrijenhoek, R. C. 1997 Molecular systematics of vestimentiferan tubeworms from hydrothermal vents and cold-water seeps. *Mar. Biol.* **130**, 141–149. (doi:10.1007/s002270050233.)

Blaxter, M. & Floyd, R. 2003 Molecular taxonomics for biodiversity surveys: already a reality. *Trends Ecol. Evol.* **18**, 268. (doi:10.1016/S0169-5347(03)00102-2.)

Borda, E. & Siddall, M. E. 2004 Arhynchobdellida (Annelida: Oligochaeta: Hirudinida): phylogenetic relationships and evolution. *Mol. Phylogenet. Evol.* **30**, 213–225. (doi:10.1016/j.ympev.2003.09.002.)

Brower, A. V. Z. 1999 A critique of Davis and Nixon's population aggregation analysis. *Syst. Biol.* **48**, 199–213. (doi:10.1080/106351599260535.)

Cracraft, J. 1983 *Species concepts and speciation analysis. Current Ornithology*, vol. 1. New York: Plenum Press pp. 159–187

Cracraft, J. 1989 Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In *Speciation and its consequences* (ed. D. Otte & J. A. Endler), pp. 28–59. Sunderland, MA: Sinauer Associates, Inc.

Davis, J. I. & Nixon, K. C. 1992 Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* **41**, 421–435.

DeSalle, R. & Birstein, V. J. 1996 PCR identification of black caviar. *Nature* **381**, 197–198. (doi:10.1038/381197a0.)

Doukakis, P., Birstein, V. J., Ruban, G. I. & DeSalle, R. 1999 Molecular genetic analysis among subspecies of two Eurasian sturgeon species, *Acipenser baerii*, and *A. stellatus*. *Mol. Ecol.* **8**, 117–129. (doi:10.1046/j.1365-294X.1999.00816.x.)

Dunn, C. P. 2003 Keeping taxonomy based in morphology. *Trends Ecol. Evol.* **18**, 270. (doi:10.1016/S0169-5347(03)00094-6.)

Edmands, S. 2001 Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. *Mol. Ecol.* **10**, 1743–1750. (doi:10.1046/j.0962-1083.2001.01306.x.)

Frost, D. R. & Kluge, A. G. 1994 A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* **10**, 259–294. (doi:10.1111/j.1096-0031.1994.tb00178.x.)

Gatesy, J., O'Grady, P. & Baker, R. H. 1999 Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics* **15**, 271–313. (doi:10.1111/j.1096-0031.1999.tb00268.x.)

Gatesy, J., Matthee, C., DeSalle, R. & Hayashi, C. 2002 Resolution of a supertree/supermatrix paradox. *Syst. Biol.* **51**, 652–664. (doi:10.1080/10635150290102311.)

Gatesy, J., Amato, G., Norell, M., DeSalle, R. & Hayashi, C. 2003 Combined support for wholesale taxic atavism in gavialine crocodylians. *Syst. Biol.* **52**, 403–422.

Gatesy, J., Baker, R. H. & Hayashi, C. 2004 Inconsistencies in arguments for the supertree approach: supermatrices versus supertrees of Crocodylia. *Syst. Biol.* **53**, 342–355. (doi:10.1080/10635150490423971.)

Godfray, H. C. J. 2002 Challenges for taxonomy. *Nature* **417**, 17–19. (doi:10.1038/417017a.)

Goetze, E. 2003 Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proc. R. Soc. B* **270**, 2321–2331. (doi:10.1098/rspb.2003.2505.)

Goldstein, P. Z. & DeSalle, R. 2000 Phylogenetic species, nested hierarchies, and character fixation. *Cladistics* **16**, 364–384. (doi:10.1111/j.1096-0031.2000.tb00356.x.)

Goldstein, P. Z. & DeSalle, R. 2003 Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. *Mol. Ecol.* **12**, 1993–1998. (doi:10.1046/j.1365-294X.2003.01860.x.)

Goldstein, P. Z., DeSalle, R., Amato, G. & Vogler, A. P. 2000 Conservation genetics at the species boundary. *Conserv. Biol.* **14**, 120–131. (doi:10.1046/j.1523-1739.2000.98122.x.)

Govedich, F. R. & Davies, R. W. 1998 The first record of the genus *Helobdella* (Hirudinoidea: Glossiphoniidae) from Australia, with a description of a new species, *Helobdella papillornata*. *Hydrobiologia* **389**, 45–49. (doi:10.1023/A:1003543314841.)

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. 2003a Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–321. (doi:10.1098/rspb.2002.2218.)

Hebert, P. D. N., Ratnasingham, S. & deWaard, J. R. 2003b Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. R. Soc. B* **270**(Suppl.), S96–S99.

Hebert, P. D., 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. (doi:10.1073/pnas.0406166101.)

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. 2004b Identification of birds through DNA barcodes. *PLoS Biol.* **2**, 1657–1663. (doi:10.1371/journal.pbio.0020312.)

Kluge, A. G. 1989 A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38**, 7–25.

Koski, L. B. & Golding, G. B. 2001 The closest BLAST hit is often not the nearest neighbor. *J. Mol. Evol.* **52**, 540–542.

Kutschera, U. 1985 Beschreibung einer neuen Egelart, *Helobdella striata* nov. sp. (Hirudinea: Glossiphoniidae). *Zool. Jb. Syst.* **112**, 469–476.

Kutschera, U. 1987 Notes on the taxonomy and biology of leeches of the genus *Helobdella* Blanchard 1896 (Hirudinea: Glossiphoniidae). *Zool. Anz.* **219**, 321–323.

Lipscomb, D., Platnick, N. & Wheeler, Q. 2003 The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends Ecol. Evol.* **18**, 65–66. (doi:10.1016/S0169-5347(02)00060-5.)

Mallet, J. & Willmott, K. 2003 Taxonomy: renaissance or Tower of Babel? *Trends Ecol. Evol.* **18**, 57. (doi:10.1016/S0169-5347(02)00061-7.)

McKitrick, M. C. & Zink, R. M. 1988 Species concepts in ornithology. *Condor* **90**, 1–14.

Nixon, K. C. & Wheeler, Q. D. 1990 An amplification of the phylogenetic species concept. *Cladistics* **6**, 211–223.

Osgood, W. 1932 Mammals of the Kelley-Roosevelts and Delacour Asiatic expeditions. *Field Mus. Nat. Hist. Publ.* 312, *Zool. Ser.* **18**, 192–339.

Prendini, L. 2005 Comment on 'Identifying spiders through DNA barcoding'. *Can. J. Zool.* **83**, 498–504. (doi:10.1139/z05-025.)

Proudlove, G. & Wood, P. J. 2003 The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends Ecol. Evol.* **18**, 272. (doi:10.1016/S0169-5347(03)00095-8.)

Ringuelet, R. 1943 Sobre le morfología y variabilidad de *Helobdella triserialis* (Em. Bl.) (Hirudinea, Glossiphoniidae). *Notas del Museo de La Plata* **69**, 215–240.

Rokas, A., Williams, B. L., King, N. & Carroll, S. B. 2003 Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804. (doi:10.1038/nature02053.)

Ronquist, F. & Gardenfors, U. 2003 Taxonomy and biodiversity inventories: time to deliver. *Trends Ecol. Evol.* **18**, 269. (doi:10.1016/S0169-5347(03)00098-3.)

Saitou, N. & Nei, M. 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.

Sarkar, I. N., Thornton, J. W., Planet, P. J., Figurski, D. H., Schierwater, B. & DeSalle, R. 2002 An automated phylogenetic key for classifying homeoboxes. *Mol. Phylogenet. Evol.* **24**, 388–399. (doi:10.1016/S1055-7903(02)00259-2.)

Sawyer, R. T. 1986 *Leech biology and behaviour.* Oxford: Clarendon Press.

Seberg, O., Humphries, C. J., Knapp, S., Stevenson, D. W., Petersen, G., Scharff, N. & Andersen, N. M. 2003 Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends. Ecol. Evol.* **18**, 63–65. (doi:10.1016/S0169-5347(02)00059-9.)

Siddall, M. E. & Budinoff, R. B. 2005 DNA-barcoding evidence for widespread introductions of a leech from the South American *Helobdella triserialis* complex. *Conservation Genet.* **6**, 467–472.

Siddall, M. E. & Burreson, E. M. 1998 Phylogeny of leeches (Hirudinea) based on mitochondrial cytochrome c oxidase subunit I. *Mol. Phylogenet. Evol.* **9**, 156–162. (doi:10.1006/mpev.1997.0455.)

Siddall, M. E., Apakupakul, K., Burreson, E. M., Coates, K. A., Erséus, C., Gelder, S. R., Källersjö, M. & Trapido-Rosenthal, H. 2001 Validating Livanow's hypothesis: molecular data agree that leeches, Branchiobdellidans and *Acanthobdella peledina* form a monophyletic group of oligochaetes. *Mol. Phylogenet. Evol.* **21**, 346–351. (doi:10.1006/mpev.2001.1021.)

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. 2003 A plea for DNA taxonomy. *Trends Ecol. Evol.* **18**, 70–74. (doi:10.1016/S0169-5347(02)00041-1.)

Trontelj, P. & Utevsky, S. Y. 2005 Celebrity with a neglected taxonomy: molecular systematics of the medicinal leech (genus *Hirudo*). *Mol. Phylogenet. Evol.* **34**, 616–624. (doi:10.1016/j.ympev.2004.10.012.)

Verrill, A. E. 1872 Descriptions of North American freshwater leeches. *Amer. J. Sci.* **3**, 126–139.

Walsh, P. 2000 Sample size for the diagnosis of conservation units. *Conserv. Biol.* **14**, 1533–1535. (doi:10.1046/j.1523-1739.2000.98149.x.)

Wheeler, Q. D., Raven, P. H. & Wilson, E. O. 2004 Taxonomy: impediment or expedient? *Science* **303**, 285. (doi:10.1126/science.303.5656.285.)

Wiens, J. J. & Penkrot, T. A. 2002 Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). *Syst. Biol.* **51**, 69–91. (doi:10.1080/106351502753475880.)

Wilson, E. O. 2003 The encyclopedia of life. *Trends Ecol. Evol.* **18**, 77. (doi:10.1016/S0169-5347(02)00040-X.)