DNA barcoding in animal species: progress, potential and pitfalls

John Waugh

Summary

Despite 250 years of work in systematics, the majority of species remains to be identified. Rising extinction rates and the need for increased biological monitoring lend urgency to this task. DNA sequencing, with key sequences serving as a ''barcode'', has therefore been proposed as a technology that might expedite species identification. In particular, the mitochondrial cytochrome *c* oxidase subunit 1 gene has been employed as a possible DNA marker for species and a number of studies in a variety of taxa have accordingly been carried out to examine its efficacy. In general, these studies demonstrate that DNA barcoding resolves most species, although some taxa have proved intractable. In some studies, barcoding provided a means of highlighting potential cryptic, synonymous or extinct species as well as matching adults with immature specimens. Higher taxa, however, have not been resolved as accurately as species. Nonetheless, DNA barcoding appears to offer a means of identifying species and may become a standard tool. *BioEssays* 29:188–197, 2007.
© 2007 Wiley Periodicals, Inc.

Introduction

The identification and characterisation of living things is fundamental to biological science. Modern taxonomy, with its origins in the mid 18th century, has described about 1.7 million species.[1] In addition, using morphological and behavioural techniques, much has been learnt about the relationships of living things with each other.

Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Private Bag 102 904, North Shore Mail Centre, Auckland, New Zealand.
E-mail: w.j.waugh@massey.ac.nz
DOI 10.1002/bies.20529
Published online in Wiley InterScience (www.interscience.wiley.com).

Abbreviations: BOLD, Barcode of Life Data Systems; CBOL, Consortium for the Barcode of Life; GBIF, Global Biodiversity Information Facility; K2P, Kimura 2 parameter; mtDNA, Mitochondrial DNA; NCBI, National Center for Biotechnology Information; PCR, Polymerase Chain Reaction; ToL, Tree of Life.

Unsurprisingly, larger animals have generally been the first to be described, while many smaller organisms remain unknown to science.[2] For example, it is thought that fewer than 10% of the vertebrates are yet to be identified but within the nematodes a bafflingly large number of species may exist, the vast majority of which has not been identified.[3] However, even among larger animals, there are doubts about species identification. The African elephant, long thought to be a single species, is now the subject of a debate based on nuclear and mitochondrial genomes over whether it incorporates two separate species.[4–6]

The earth's biota may contain between 10 and 100 million eukaryotic species.[7,8] The identification of numbers in this range represents an insurmountable workload for taxonomists using current methods. Even allowing for improvements in communications and the impact of the internet, the task is overwhelming. Moreover, phenotypic plasticity and genotypic variation in the features used for identification easily lead to identification errors and cryptic species or differing life stages can add to the confusion.[9] To compound matters, the task of cataloguing extant species is lent urgency by currently observed mass extinctions that are widely believed to be anthropogenic in origin.[10,11]

Field biologists, faced with the reality of species diversity, recognise the inadequacy of their own ability to access what *is* known about the biota, let alone what is not. These problems also impact upon people working in other areas such as combating the trade in endangered species, monitoring fisheries, identifying and controlling the spread of pest species or disease, identifying extinct lineages and regulating the movement of biological material around the world.[12–15]

Clearly there is a need to accelerate and simplify the processes of identification involved and, because of the scale of the problem, new methods will have to be employed. In addition, as more species are described, accessing the enlarged pool of taxonomic knowledge will become even more problematic.

Recent developments in DNA-sequencing technology have introduced the possibility of using variations in short sequences of DNA as labels for species in a process that has become known as DNA barcoding. The concept has already gained considerable acceptance among those working with species refractory to morphological identification such as viruses,[16] bacteria,[17,18] protists[19] and Rhodophyta.[20]

However, it is apparent that, since morphological techniques are difficult to access and apply without considerable training, some more rapid system of species identification is required for all taxa. This has led to the formation of the Consortium for the Barcode of Life (CBOL),[21] which aims to provide such a DNA barcode for every species on the planet (Box 1).

This review examines the progress, potential and pitfalls of DNA barcoding in animal species.

The DNA-barcoding process

A DNA barcode is a short sequence of nucleotides taken from an appropriate part of an organism's genome that is used to identify it at species level. Intraspecific variation in this sequence is an order of magnitude less than that observed interspecifically and this provides the means by which species are differentiated. It is not part of a DNA taxonomy nor is it a tool for phylogenetic reconstruction. It simply provides a means of linking sample specimens directly to existing voucher specimens and taxonomical information.

Central to the efficacy of DNA barcoding is the selection of a suitable segment of DNA. Its mutation rate must be slow enough so that intraspecific variation is minimised but sufficiently rapid to highlight interspecific variation.[9] It must be relatively easy to collect and should have as few insertions or deletions as possible to facilitate sequence alignment.

Mitochondrial DNA (mtDNA) offers several advantages over nuclear DNA. According to Drake's observation,[22] the rate of DNA mutation is inversely related to the size of the genome. Hence, nuclear DNA undergoes relatively slow mutation compared with mtDNA and, for this reason, would require a much longer nucleotide sequence than is necessary with mtDNA in order to provide a barcode capable of differentiating species.

In animals, mtDNA occurs as a single double-helical circular molecule containing 13 protein-coding genes, 2 ribosomal genes, a non-protein-coding control region, and several tRNAs (Fig. 1).[23] Each mitochondrion contains several such circular molecules and, therefore, several complete sets of mitochondrial genes. Furthermore, each cell has several mitochondria. Thus, when sample tissue is limited, the mitochondrion offers a relatively abundant source of DNA.

Cytochrome *c* oxidase is a large transmembrane protein found in the mitochondrion, which is highly conserved across

Box 1

The Consortium for the Barcode of Life (CBOL) is an international organisation devoted to developing DNA barcoding as a global standard in taxonomy. It comprises more than 120 member organisations from 45 countries and includes museums, herbaria, zoos, research organisations, governmental and intergovernmental agencies as well as other organisations involved in taxonomic research and biodiversity issues. Members agree to submit their DNA barcode sequences and voucher specimen data to a public database. CBOL was launched in May 2004 and is overseen by an executive committee that reports to the member organisations. It has five working groups to develop particular aspects of DNA barcoding, a Scientific Advisory Board and a small Secretariat Office to conduct its business. Within the auspices of CBOL a number of initiatives have been established including the All Birds Barcode Initiative, the Fish Barcode Initiative, the All Leps (Lepidoptera) Barcode Initiative and the International Network for Barcoding Invasive and Pest Species. Another group is exploring the barcoding of endangered vertebrates. Each of these initiatives aims to obtain DNA barcodes for every species within its group. The Canadian Centre for DNA Barcoding (a member of CBOL) oversees a website for Barcode of Life Data Systems (BOLD) that permits the uploading of sequences from the 5′ region of the COI gene and returns a species-level identification when one is possible. At present the site has more than 165,000 sequences from almost 20,000 species and these numbers are increasing steadily. The site also permits a variety of forms of data analysis for submitted sequences. CBOL works in cooperation with a number of other organisations including the Global Biodiversity Information Facility (GBIF), National Centre for Biotechnology Information (NCBI) and many taxanomic communities and web based projects.

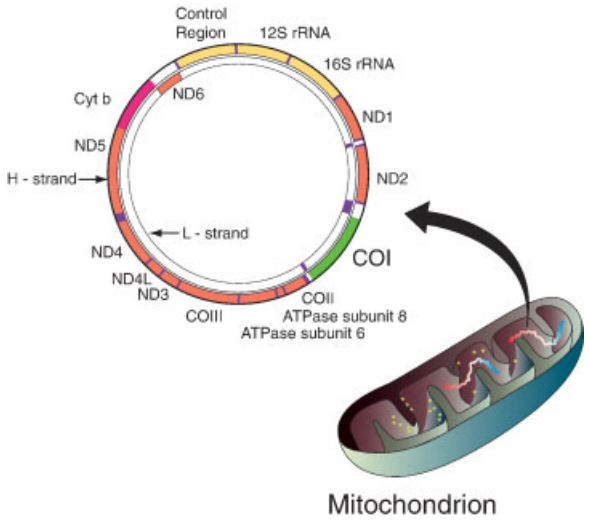

Figure 1. A diagrammatic representation of a mitochondrial genome based on complete mitochondrial DNA sequences from a variety of bird species.

species that employ oxidative phosphorylation for metabolism. It functions as the terminal electron acceptor in the respiratory chain, catalysing the reduction of oxygen to water and pumping protons across the membranes of the cristae.[23,24] The protein comprises several subunits of nuclear origin and three subunits synthesised in the mitochondrion. The mitochondrial subunits are known as subunits I, II and III. Cytochrome *c* oxidase subunit I (COI), the catalytic subunit of the enzyme, is predominantly imbedded in the membrane of the mitochondrial crista (Fig. 2). This structure would indicate a significant level of structural and functional constraint. However, the nucleotides of the gene that codes for it show sufficient variation to differentiate between species. Conversely, intraspecific variation in this gene is generally $<$10% of that observed between species. Moreover, insertions and deletions are rare.[3]

Recent studies associated with CBOL have generally selected a 648 bp segment of the COI gene, starting from the 5′ end, to generate a suitable barcode (Fig. 2).[21]

Having selected an appropriate segment of DNA for analysis, it must first be extracted from the sample specimen and amplified using the polymerase chain reaction (PCR). The amplified segment of the COI gene is sequenced and this sequence, the ''barcode'', is then matched with existing barcodes or material from voucher specimens. The Kimura 2-parameter (K2P) genetic distance correction is used to quantify sequence divergences among individuals because it is the most effective model when distances are low, as is the case with COI barcoding.[9]

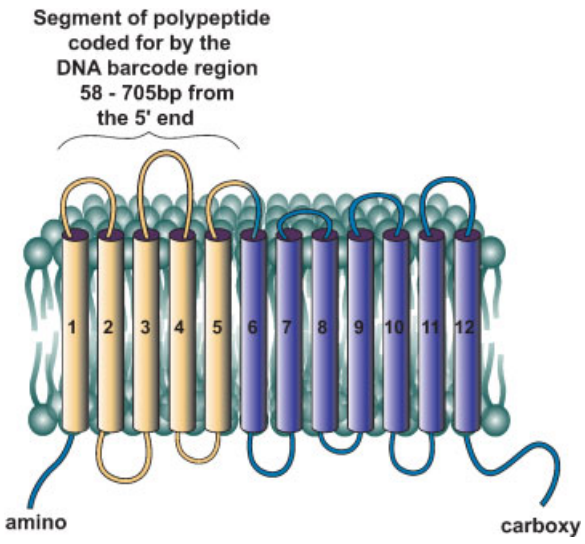

Figure 2. The predicted transmembrane structure for cytochrome *c* oxidase subunit 1 (COI). The area highlighted in yellow that includes five of the twelve transmembrane regions is coded for by the sequence designated by CBOL as the DNA barcode region.

The efficacy of DNA barcoding

To date, the literature contains a considerable number of fully published studies of animals in which DNA barcoding, using part of the COI gene, has been assessed[9,25–36] or used to help resolve taxonomic ambiguity[15,32–42] or used as part of a taxonomic description[43] (Table 1). DNA barcoding was employed to resolve species within narrow taxonomic groupings in some of the studies[15,25–27,31–36,40–43] or to identify higher taxa from wider assemblages of animals in others.[9,25,30] In addition, DNA-sequencing technology has been used for identifying organisms from other Kingdoms including plants,[44,45] bacteria,[17,18] protists[19] and viruses;[16] however, these are not reviewed here.

Although minor variations in protocols occurred between studies, broadly similar methods were used in each. DNA extracted from tissue samples was prepared using a variety of standard protocols. The sequence blocks of DNA used ranged in length from 350 bp to ~1000 bp. COI profiles were then generated using automated sequencers and these profiles were compared both within and between species as well as between higher taxa in some studies.[9,25,30]

Species

DNA barcoding provided a high degree of taxonomic resolution ($>$95%) for most species examined in the studies reviewed.[9,25–28,30,32–36,40–42] Each species had a unique nucleotide sequence at COI with only slight intraspecific K2P divergence. For example, in one study of 13,320 congeneric species pairs,[28] intraspecific variation was usually less than 1% and rarely more than 2%, while mean interspecific divergence was 11.3%. This difference between intra- and interspecific divergences at COI was also observed in arachnid,[25,32,40] lepidopteran,[9,33,36] (Fig. 3), dipteran,[34] avian,[26] fish[31] and Collembola[27] species (Table 1).

DNA barcoding was generally successful when used for identifying immature specimens,[40] extinct species[15] and individual species at differing stages in their life cycles.[32,36,40] Furthermore, possible cryptic species were identified in several studies.[25–27,31,34,36–39,41]

Identification difficulties caused by morphological differences between instars in cave-dwelling spiders (*Cicurina* spp.) were resolved using this technique, which also aided identification of populations where adult specimens are extremely rare.[40] Potential species identified in this study correspond closely to *a priori* species hypotheses except in the case of *C. caliga* and *C. hoodensis*, which contain COI sequences so similar that it is hypothesised they are synonymous. Furthermore, COI barcoding of two other species, *C. madla* and *C. vespera* (a species known only from one female specimen), suggested possible synonymy and indicated a need for further evaluation. This utility was also observed when 10 distinct caterpillars were linked to their morphologically similar adults.[36] While in another study, COI

Table 1. Design details and major results from selected studies that utilised part of the cytochrome *c* oxidase subunit 1 gene to provide a DNA barcode for identifying animal taxa

Reference	Sequence length	No. of taxa	% identified	K2P % intra vs interspecific COI variation	Comment
Barrett et al. 2005[23]	600bp	203 arachnid spp.	100	1.4 vs 16.4	Mean intra- and interspecific nucleotide divergences did not overlap except in the case of probable cryptic species.
Brown et al. 2003[41]	648bp	1 Lepidopteran sp. *Gnetom gnemon*	100		COI was used as a taxonomic descriptor for a new species.
Greenstone et al. 2005[30]	439bp	32 Carabidae spp. and 39 Araneae spp.			COI used to identify species and link different life stages.
Hajibabaei et al. 2006[31]	311–612bp	521 Lepidopteran spp.	97.9%	0.17–0.46 vs 4–6	Morphologically distinct sympatric species from three families identified.
Hebert et al. 2003a[7]	658bp	7 animal phyla	96.4		The efficacy of COI at identifying species, orders and phyla assessed.
		8 insect orders	100		
		200 Lepidopteran spp.	100	Lepidoptera 0.25 vs 6.84	
Hebert et al. 2003b[26]	>400bp	2238 Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Platyhelminthes, and minor phyla	>98	Overall, usually <2 vs 11.3	The efficacy of COI at identifying species from eight major and several minor phyla plus a variety of arthropod classes was assessed. Cnidarians showed less COI variation between species than all other taxonomic groups, 94.1% vs 1.9% showing <2% K2P between spp. ($p < 0.0001$).
Hebert et al. 2004a[34]	648bp	10 Lepidopteran spp.	100		Ten new taxa identified. Different life stages matched for species.
Hebert et al. 2004b[24]	648bp	260 Avian spp.	100	0.43 vs 7.93	Four possible cryptic species identified.
Hogg et al. 2004[25]	710bp	19 Collembola spp.	100	0.78 vs 19.0	Produces high resolution in Collembola species. Possible cryptic species identified.
Hu et al. 2002[35]	≤530bp	7 Hookworm spp.	—		Three of the seven species appeared to be possible species complexes based on intraspecific COI variation.
Hu et al. 2005[36]	≈450bp	3 *Progamotaenia* spp. (Platyhelminthes)			Variation at COI suggests that all three species are species complexes.
Lambert et al. 2005[13]	596bp	10 Moa spp.	100		COI used to identify extinct species. Possible species synonymy highlighted.
Lorenz et al. 2005[27]	727bp	56 primates		0.011 vs -	Problems with taxon specific patterns of 'universal primer' failure. Taxon specific primers developed.
Ngarmamonpirat et al. 2005[37]	450bp	1 *Gnathastoma spinigerum* (hookworm)	—		COI barcode variation did not match the morphological variation observed in 3rd larval stage of this hookworm.
32 Paquin et al. 2004[38]	≈1000bp	23 *Cicurina* spp.(Arachnida)	≈100	1.09 vs 7.12	Immature specimens identified to species level. Possible species synonymy identified.
Penton et al. 2004[39]	709bp	2 *Daphnia* spp.(Crustacea)	100		Identification of morphologically cryptic species with overlapping distribution.
Remigio et al. 2003[28]	672bp	70 Gastropod spp.	—		COI used to identify species and higher taxonomic relationships. Insertion or deletions more common in COI in this taxonomic group.
Smith et al. 2006[32]	658bp	32 Dipteran spp.	100	0.17 vs 5.78	Fifteen cryptic species found using COI.
Vences et al. 2005[33]	550–650bp	9 Mantellid frog spp. 4 *Aneides* spp. (salamanders)		5.4 (mantellid frogs) and 4.3 (salamanders) vs 20.7 and 13.5	Found high intraspecific variability (7–18%). The use of mitochondrial 16S rRNA gene to supplement COI suggested.
Ward et al. 2005[29]	655bp	207 fish	100	0.39 vs 9.93	Efficacy of COI at identifying species and higher taxonomic relationships assessed. Possible cryptic species identified.
Whiteman et al. 2004[40]	379bp	2 Lice spp. (Insecta)	100		Study of bird parasites with similar morphology that can be identified using COI.

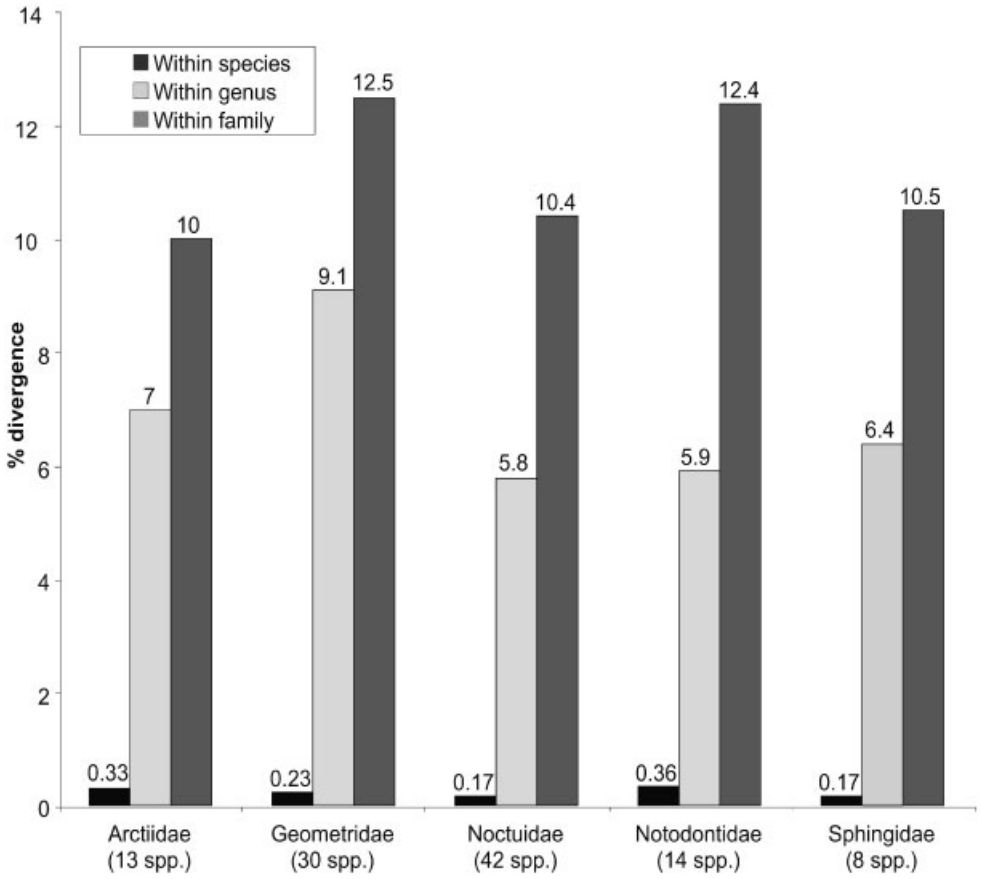

Figure 3. Nucleotide divergence in a 617 bp segment of the COI gene in five lepidopteron families at species, genus and family level. Data from Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a Proc R Soc Lond B 270:313–322

barcoding helped to identify arachnids and carabids of different life stages including eggs, larvae or nymphs and pupae.[32]

In extinct taxa, where full taxonomic descriptions are difficult due to the lack of soft tissue, identification of species can be particularly problematical. However, 6 species of an extinct ratite bird, the New Zealand moa, were identified using DNA barcoding when 2.7% sequence divergence at COI was used as the intraspecific threshold. This increased to 10 species when a threshold of 1.24% was used.[15] The species identified were generally confirmed and supported by results from a larger study of moa mtDNA, in which 125 specimens had mitochondrial control region sequences analysed (Fig. 4).[46]

Potential cryptic species were identified using DNA barcoding among butterflies,[36] flies,[34] birds,[26] arachnids,[25] springtails,[27] within the species *Daphnia obtusa*[41] and in three groups of parasitic worms.[37–39]

Larval caterpillars with distinct colour patterning and food plants were linked with adults that are phenotypically very similar to each other.[36] Divergence at COI was considerable (mean K2P 2.76%; range 0.0–7.95%) and when caterpillar/adult morphology and food plants were mapped onto a neighbour joining tree of the COI divergence, 10 probable new species were revealed that showed covariance between morphological, molecular and ecological characteristics.

Another study found fifteen cryptic species of parasitoid flies that show high host-specificity within a group of what had been thought were three generalist species.[34]

COI-identified cryptic species were not limited to relatively obscure taxa. For example, among 260 North American bird species, K2P distances were 18-fold higher between species than within them; however, in four species (*Tringa solitaia, Sturnella magna, Cisthorus palustris* and *Vireo gilvus*), high intraspecific K2P distances suggested the presence of cryptic species.[26]

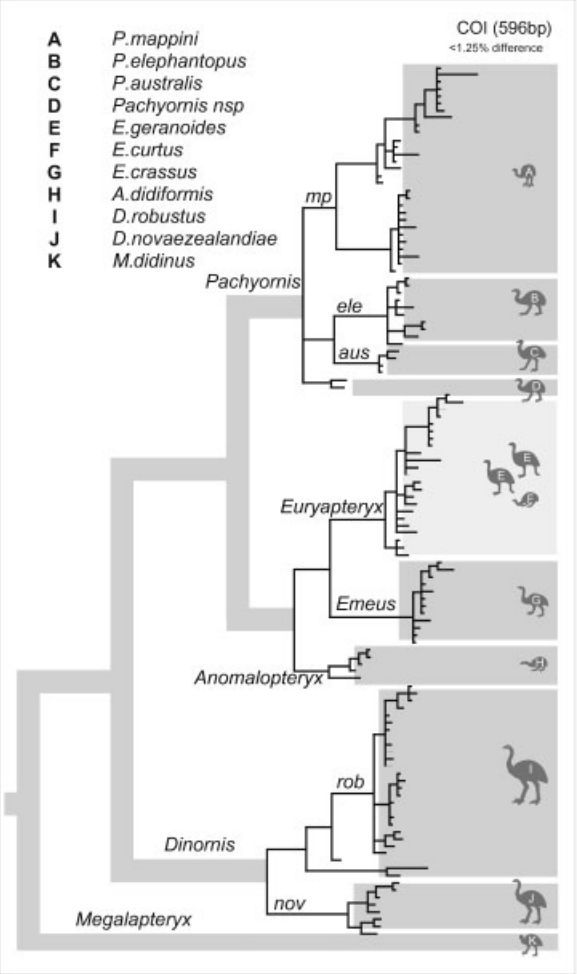

Figure 4. A comparison of the efficacy of mitochondrial control region (656 bp) with COI sequences (596 bp) at determining moa lineages.[13,44] The tree was constructed using control region mtDNA sequences. Dark grey boxes represent COI sequences that differ by less than 1.25%. Euryapteryx species (light grey box) appear to represent a complex of overlapping subgroups. The three lineages suggested by control region sequences for *D. robustus* and the two for *P. mappini* are reduced to two lineages using COI data.

Similarly, intraspecific K2P divergence averaged 1.4% among 37 arachnid species but in one species, *Latrodectus hesperus*, it was 3.6%.[25] This divergence was between northern and southern populations and suggested that they are probably separate species; a conclusion supported by the results of another study[47] that examined breeding and pheromones in these populations.

The presence of undescribed species was revealed in a study of springtails from the Canadian Arctic.[27] One species showed up to 13% intraspecific COI divergence compared with the <1% generally observed in species from this group. Likewise, wide sequence variation at COI was observed among members of the species *Daphnia obtusa* collected from 33 North American sites,[41] indicating that this may in fact be two species; one confined to the east and the other more broadly distributed.

The efficacy of DNA barcoding at revealing cryptic species was further demonstrated in three studies of parasitic worms.[37–39] Analysis of COI fragments in each study found high intraspecific COI variation among tapeworm and hookworm species. Two of the three studies[37,38] suggested that this might be due to the presence of cryptic species. In the third,[39] the authors conclude that there was little COI and morphological covariance; however, the presence of cryptic species was not precluded.

Problematic taxa

Some taxonomic groups were not readily resolved to species level. These included benthic Cnidarians,[28] two groups of amphibians[35] and some Gastropod species.[30]

There was little COI divergence between species of benthic Cnidarians, with 94.1% of species pairs showing <2% divergence versus 1.9% of species pairs from all other phyla in this large study.[28] According to Shearer et al,[48] Cnidarians show particularly slow mutation rates in their mitochondrial DNA when compared with other taxa and this may impair the resolving power of COI in this group.

Conversely, high intraspecific variation at COI of up to 18% was observed in two amphibian groups (mantellid frogs and salamanders).[35] This overlapped the interspecific variation, making species delineation difficult. Furthermore, variability within the mitochondrial genome of these taxa meant that a mix of primers was required to isolate the required segment. Nonetheless, COI sequences were able to correctly identify species including disparate geographic variants.

Gastropods also proved refractory to COI identification in a study of 70 species aimed at establishing phylogenetic relationships as well as species identification.[30] Insertions and deletions found in the COI gene of two subclasses, Heterobranchia and Patellogastropoda, complicated alignment.

These results with gastropods conformed to the general observation that where DNA barcoding was used to resolve higher taxa, results proved more equivocal. In one study,[25] 87% of genera that contained several species and 67% of families that contained several genera formed cohesive COI groups. Another study correctly assigned 96.4% of 55 taxa to phyla and 100% of 50 taxa to ordinal level (Table 1).[9]

Discussion

There is currently an acknowledged biodiversity crisis of anthropogenic origin.[10,11] It is the result of the destruction of habitats and unsustainable harvesting of natural products. Many species are becoming extinct without our ever having registered their existence. The fact that the majority of

Eukaryotes remain unknown to science has focussed attention on the overwhelming task that taxonomists face in trying to identify them. However, moments of crisis can precipitate novel and creative solutions. To quote Quentin Wheeler[49] "within the field of taxonomy, there is presently a conjunction of new theories, technological advances and urgent needs." The first two elements of this conjunction represent opportunities, the third, according to Plato, "is the mother of invention." Into this environment comes DNA barcoding; not, perhaps, a solution to the crisis but a tool that may help in its resolution.

DNA barcoding claims neither to replace taxonomy nor to reconstruct phylogenies. It does not absolve governments and funding bodies of the need to inject new life into the science of taxonomy. However, it may prove a useful tool for taxonomists and the many other agencies and individuals interested in species identification.

In general, DNA barcoding was effective at resolving species in the studies reviewed. However, resolution of higher taxonomic groups was less effective. Indeed, to paraphrase Greenstone et al., COI barcoding is a diagnostic tool for identifying animal species and cannot be expected to serve double duty as a character for deeper phylogenetic reconstruction.[32]

DNA barcoding showed utility at identifying potential candidates for taxonomic description by highlighting possible cryptic species among those already identified. It drew attention to a number of species, both extant and extinct, that may be synonymous and provided a means of identifying species regardless of life stage or maturity. Furthermore, data from a study using other mtDNA sequences in extinct bird species was confirmed using DNA barcoding.

Any system that is designed to identify species needs to tackle the issue of what a species is. There are many concepts including typological, phylogenetic, morphological, biological, isolation and mate recognition, to name a few. Each has its own perspective, none is universally accepted. However, it is beyond the scope of this paper to debate the relative merits of these species concepts. Moreover, DNA barcoding does not answer philosophical or ontological questions about the nature of species or higher taxa, it simply identifies highly correlated sequences, which, if it works, are derived from highly correlated (i.e. conspecific) individuals. Furthermore, once species limits have been defined, DNA barcoding may raise unexpected questions about what those limits are. The results presented in this review suggest that, empirically, DNA barcoding accurately identified species in $>$95% of cases.

Some questions relating to this technique remain to be answered. For example, what is a DNA barcode? Although CBOL states that "only cytochrome *c* oxidase 1 is approved as a barcode region, defined relative to the mouse mitochondrial genome as the 648 bp region that starts at position 58 and stops at position 705",[21] very few of the sequences lodged with GenBank are of this precise segment of the gene. A large number are considerably shorter and of those longer than 648 bp, many do not fully overlap the specified segment. This variability complicates and reduces the power of large-scale analyses of these data.

Another question arises: is it necessary to use many barcodes from differing genes in order to identify the broadest range of species? COI does not work for other Kingdoms and Nielsen et al.[50] suggest that the weakest aspect of DNA barcoding is that no single gene will always be invariant within species but different between species. The results of some of the problematic taxa bear this out. They further suggest that there is a need for statistical protocols to assess whether a sample barcode is sufficiently similar to a known barcode to justify species assignment.[50] Thus, issues of standardisation need to be addressed if barcoding is to achieve the rigour required of an enduring contribution to science.

In addition, taxa that are undergoing rapid speciation show little interspecific COI divergence, thus compromising the resolving power of COI barcoding. For example, New Zealand moas are thought to have undergone rapid speciation prior to extinction,[46] which may account for the relatively low interspecific K2P distances observed in this group.[15] The same may be true of the cichlid fish in Africa, a group known for its rapid speciation.[51] The converse may be true for species that have not undergone recent speciation events. Moreover, some groups, such as the Cnidarians, show slow mitochondrial DNA-sequence evolution resulting in negligible interspecific variation.

Critics of DNA barcoding suggest that it is unscientific because it does not set out to test hypotheses, that it generates information not knowledge.[52,53] However, arguably, any experiment generates information that requires interpretation. Moreover, barcoding tests the hypothesis that species can be identified using this technique and in future may be a source of data that will generate other hypotheses. Furthermore, it is possible to make similar comments about the invention of the microscope, perhaps the most-important scientific invention of the past millennium.

A number of people remain sceptical of the utility and efficacy of DNA barcoding. There are those who fear that the promoters of barcoding are seeking to replace conventional taxonomy.[52–55] These fears have been fuelled by an enthusiasm for DNA taxonomy in some quarters outside CBOL.[56] Some take exception to the use of the term barcode on the basis that it suggests that "each species has a fixed and invariant characteristic like the barcode on a supermarket product".[54] They also express reservations that sufficient numbers of congeneric species have been sampled or that those samples come from a wide-enough distribution to make generalisations about the efficacy of COI barcoding.[54] In addition, there are concerns that any attempt at producing a universal system for identifying species entailing a centralised database may be seen by third world countries as an attempt

by wealthier nations to monopolise taxonomic information.[54] It is also thought that any such system may be more authoritarian and will lack the flexibility inherent in the committee style consensus of existing botanical and zoological codes.

Proponents of barcoding respond that COI barcoding is not a substitute for taxonomy.[57,58] That it cannot be, since it is only by linking barcodes to fully described voucher specimens that the full power of the technique can be realised and that just as supermarket barcodes would be meaningless without the database of product details to which they are linked, so DNA barcoding requires a database of taxonomic information to which it links. They acknowledge that the barcode analogy is not an exact one but maintain that, with the current level of accuracy observed, COI barcoding has proved sufficiently discriminatory in trials to demonstrate considerable utility as a tool for differentiating species and, therefore, merits further investigation. With regard to the comments about the numbers of species tested using this method so far, it is only as the database of barcodes builds that the substance of these reservations will either be confirmed or otherwise.

Unquestionably, any progress that is to be made in accelerating species identification will be dependent on the use of new technologies and will employ accessible, easily searchable repositories of taxonomic information. Whether or not this is perceived as a threat by countries around the world will depend largely upon the sensitivity with which the process is managed. CBOL favours the open access approach of the Global Biodiversity Information Facility (GBIF)[59] and works in coordination with them.

Those who express concern that this may be an attempt to divest third world countries of information relating to their biotas, are also probably underestimating the power of modern technology to disseminate information. If the current open access approach is maintained, taxonomic information, much of which resides in relatively inaccessible Northern Hemisphere museums and collections, will become more accessible rather than the reverse. Thus, regions may be able to access and reclaim information relating to their indigenous biotas.

The methods by which species are named and described, however, will not be affected by DNA barcoding. This technique is not "a pretender to the taxonomic throne",[60] its principal utility is as a searchable label rather than as a contributing taxonomic feature. Thus, barcoding may serve to help inform the debate that generally surrounds species identification but is unlikely to undermine the flexibility of existing codes.

Museums around the world maintain large collections of plant and animal specimens and are, therefore, excellent sources of material for DNA barcoding. Their accumulated experience at curating these collections combined with a need for efficient cataloguing systems suggest the potential for a mutually beneficial relationship with CBOL. Furthermore, these institutions play a central role in the collection and description of new species and may, therefore, be primary beneficiaries of the development of new technologies for identifying known and unknown species. To date, several major museums and research institutions around the world have lent support to the consortium; however, some remain sceptical about the utility of a barcode system.

Evidence suggests that DNA barcoding can serve as a means of accessing taxonomic information and help in the identification of species. However, even if a complete barcode resource of the world's biota is produced, it cannot achieve its full potential unless the processes involved in obtaining barcodes from specimens and accessing the taxonomic information relating to them are simplified and streamlined so that they can be quickly carried out by relatively unskilled workers in a variety of locations. Those involved with CBOL envisage the development of a hand-held device to facilitate this. Although such a device does not represent an insuperable developmental problem, it requires both capital investment and the determination to ensure that it does reach production. However, taxonomy is not a science that tends to attract this level of funding.

The proliferation of unrelated web-based initiatives represents a possible impediment to the accretion of this investment. There are a large number of projects, unrelated to CBOL, that are attempting to collate taxonomic information either as phylogenetic trees or as catalogues of species. Some of these initiatives show clear areas of overlap, others are markedly dissimilar and have much to offer each other. An example of these is the Tree of Life (ToL) project[61] that provides a framework in which to electronically publish taxonomic information in a searchable database. It is a collaborative effort and currently consists of approximately 4000 web pages. Each page contains information about a particular taxonomic group and is linked hierarchically to other pages in the form of a phylogeny of life. The CBOL and ToL initiatives share links with the GBIF and are complimentary, with CBOL focused on species and ToL on higher classification. DNA barcodes may supply a key to rapidly linking specimens to ToL information and might expedite navigation around this large and growing database. ToL may provide taxonomic information to which DNA barcodes link. Symbiosis between these projects can facilitate the process of species identification and may also lead to as yet unenvisaged use of ToL information, thus, increasing the value of both initiatives.

Conclusion and outlook

DNA barcoding shows considerable potential as a system for identifying species that may allow users to link specimens to databases of taxonomic information as well as highlighting those species for which no data are yet available. It is not a thorough taxonomic description nor is it a tool for phylogenetic reconstruction. However, it may help speed the work of

taxonomists and others interested in species identification. To date, evidence from a number of studies largely confirms the feasibility of such a system.

The success of DNA barcoding will depend on the rationalisation and coordination of the many online species identification programmes, so that sufficient resources can be concentrated to develop rapid in situ sequencing. There is also a need for the problems associated with species refractory to COI identification to be resolved. Furthermore, the support of taxonomists, from around the world, is a vital prerequisite that may prove the most intractable obstacle faced by CBOL. However, with sufficient impetus and dedication, none of these problems are insoluble.

In the 270 years since Carl Linnaeus first published his *Systema Natura*, approximately 1.7 million species have been described. There may be as many as 100 million Eukaryotes that remain unknown to science. If that is true, at the present rate, it will take about 600 generations of scientists or 16,000 years before the job is complete. Clearly, systems for accelerating these processes are required. Use of electronic media can speed progress and help to reduce duplication or the loss of information that occurs with the death of each taxonomist but new methods are necessary if a breakthrough is to be made. Regardless of whether DNA barcoding is this breakthrough, it may prove to be a very useful taxonomic tool. Given the enormity of the task of identifying the world's biota and the many other potential tasks for which it might be employed, it would seem imprudent to ignore the promise of DNA barcoding.

Acknowledgments

I am grateful to David Lambert, Craig Millar, Jennifer Hay, Philip Smith, Gimme Walter and Andrew Dodd for their useful comments and for proof-reading this manuscript.

References

1. Stoeckle M. 2003. Taxonomy, DNA, and the Bar Code of Life. 2003. BioScience 53:2.
2. Blaxter ML. 2003. Counting angels with DNA. Nature 421:122–124.
3. Blaxter ML. 2004. The promise of a DNA taxonomy. Phil Trans R Soc Lond B 359:669–679.
4. Comstock KE, Georgiadis N, Pecon-Slattery J, Roca AL, Ostrander EA, et al. 2002. Patterns of molecular genetic variation among African elephant populations. Molecular Ecology 11:2489–2498.
5. Roca AL, Geordiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. Nature genetics 37:96–100.
6. Debruyne R. 2004. Contribution of molecular phylogeny and morphometrics to the systematics of African elephants. J Soc Biol 198:335–342.
7. May RM. 1988. How many species are there on Earth?. Science 241: 1441–1449.
8. Whitfield J. 2003. DNA barcodes catalogue animals. Nature. News Service. Available on URL http://www.nature.com/news/2003/030512/full/0305127.html. Accessed 11 December 2005.
9. Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a. Biological identifications through DNA barcodes. Proc R Soc Lond B 270:313–322.
10. American Museum of Natural History (AMNH). 1998. National survey reveals biodiversity crisis. Press release. available on URL http://www.amnh.org/museum/press/feature/biohall.html. Accessed 24 May 2006.
11. Ananthaswamy A. 2004. Scene set for next mass extinction. New Scientist Mar 27.
12. Baker CS, Palumbi SR. 1994. Which whales are hunted? A molecular approach to monitoring whaling. Science 265:1538–1539.
13. Baker CS, Dalebout ML, Lavery S, Ross HA. 2003. www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. Trends in Ecol and Evol 18:271–272.
14. 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.
15. Lambert DM, Baker A, Huynen L, Haddrath O, Hebert PDN,, et al. 2005. Is a large-scale DNA-based inventory of ancient life possible? J Hered 96:1–6.
16. Mayo MA, Horzinek MC. 1998. A revised version of the International Code of Virus Classification and Nomenclature. Arch Virol 143:1645–1654.
17. Woese CR. 1987. Bacterial evolution. Microbiol Rev 51:221–271.
18. Unwin R, Maiden MCJ. 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol 11:479–487.
19. Cavalier-Smith T. 2004. Only six kingdoms of life. Proc R Soc Lond *B* 271:1251–1262.
20. Saunders GW. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Phil Trans R Soc Lond 360:1879–1888.
21. Barcode of Life website. http://www.barcodinglife.org/ . Accessed 3 May 2006.
22. Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.
23. Research Collaboratory for Structural Bioinformatics (RCSB) Protein Databank (PDB). available on URL http://www.rcsb.org/pdb/cgi/explore.cgi?pdbId=2OCC . Accessed 19 May 2006.
24. Cytochrome c oxidase homepage. available on URL http://www.bioc.rice.edu/~graham/CcO.html . Accessed 19 May 2006.
25. Barrett RDH, Hebert PDN. 2005. Identifying spiders through DNA barcoding. Can J Zool 83:481–489.
26. Hebert PDN, Stoeckle MY, Zemiak TS, Francis CM. 2004b. Identification of birds through DNA barcodes. PLoS Biol 2004b 2:1657–1663.
27. Hogg ID, Hebert PDN. 2004. Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zool 82:749–754.
28. Hebert PDN, Ratnasingham S, deWaard JR. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc R Soc Lond B 270:1–4.
29. Lorenz JG, Jackson WE, Beck JC, Hanner R. 2005. The problems and promise of DNA barcodes for special diagnosis of primate biomaterials. Phil Trans R Soc Lond 360:1869–1877.
30. Remigio EA, Hebert PDN. 2003. Testing the utility of partial COI sequences for phylogenetic estimates of gastropod relationships. Mol Phylogenet Evol 29:641–647.
31. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. Phil Trans R Soc Lond 360:1847–1857.
32. Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS, et al. 2005. Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Molecular Ecology 14:3247–3266.
33. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. 2006, DNA barcodes distinguish species of tropical Lepidoptera. PNAS 103: 968–971.
34. Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). PNAS 103:3657–3662.
35. Vences M, Thomas M, Bonnett RM, Vieites DR. 2005. Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philosophical Transactions of the Royal Society B 360:1859–1868.
36. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. PNAS 101:14812–14817.

37. Hu M, Chilton NB, Zhu X, Gasser RB. 2002. Singe-strand conformation polymorphism-based analysis of mitochondrial cytochrome *c* oxidase subunit 1 reveals significant substructuring in hookworm populations. Electrophoresis 23:27–34.
38. Hu M, Gasser RB, Chilton NB, Beveridge I. 2005. Genetic variation in the mitochondrial cytochrome *c* oxidase subunit 1 within three species of *Progamotaenial* (Cestoda: Anoplocephalidae) from macropodid marsupials. Parasitology 130:117–129.
39. Ngarmamonpirat C, Waikagul J, Petmitr S, Dekumyoy P, Rojekittikhun W, et al. 2005. Analysis of sequence variations in *Gnathostoma spinigerum* mitochondrial DNA by single-strand conformation polymorphism analysis and DNA sequence. Parasitology International 4:65–68.
40. Paquin P, Hedin M. 2004. The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Circurina* (Araneae: Dictynidae) from Texas caves. Molecular Ecol 13:3239–3255.
41. Penton EH, Hebert PDN, Crease TJ. 2004. Mitochodrial DNA variation in North American populations of *Daphnia obtuse*: continentalism or cryptic endemism? Molecular Ecology 13:97–107.
42. Whiteman NK, Santiago-Alarcon D, Johnson KP, Parker PG. 2004. Differences in straggling rates between two genera of dove lice (Insect: Phthiraptera) reinforce population genetic and cophylogenetic patterns. Int J Parasitol 34:1113–1119.
43. Brown JW, Miller SE, Horak M. 2003. Studies on New Guinea moths. 2. Description of a new species of Xenothictis Meyrick (Lepidoptera: Tortricidae: Archipini). Proc Entomol Soc Wash 105:1043–1050.
44. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. PNAS 102:8369–8374.
45. Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, et al. 2005 The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142–165.
46. Baker AJ, Huynen LJ, Haddrath O, Millar CD, Lambert DM. 2005. Reconstructing the tempo and mode of evolution in an extinct clade of birds with ancient DNA: the giant moas of New Zealand. PNAS 102: 8257–8262.
47. Kasumovic MM, Andrade MCB. 2004. Discrimination of airborne pheromones by mate searching male western black widow spiders (*Latrodectus Hesperus*): species- and population-specific responses. Can J Zool 82:1027–1034.
48. Shearer TL, Van Oppen MJH, Romano SL, Wörheide G. 2002. Slow mitochondrial DNA sequence evolution in the Anthazoa (Cnidaria). Molecular Ecology 11:2475–2487.
49. Wheeler QD. 2004. Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society B 360:571–583.
50. Nielsen R. Matz M. 2006 Statistical approaches for DNA barcoding. Systematic Biology 55:162–169.
51. Won YJ, Sivasundar A, Wang Y, Hey J. 2005. On the origin of Lake Malawi cichlid species: A population genetic analysis of divergence. PNAS 102:6581–6586.
52. Lipscomb D, Platnick N, Wheeler Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. Trends in Ecology and Evolution 18:65–66.
53. Ebach MC, Holdrege C. 2005. DNA barcoding is no substitute for taxonomy. Nature 434:697.
54. Moritz C, Cicero C. 2004. DNA Barcoding: promise and pitfalls. PLoS Biology 2:1529–1531.
55. Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, et al. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. Trends in Ecology and Evolution 18:63–65.
56. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP. 2003. A plea for DNA taxonomy. Trends in Ecology and Evolution 18:70–74.
57. Hebert PDN, Gregory TR. 2005. The promise of DNA barcoding for taxonomy. Systematic Biology 54:852–859.
58. Schindel DE, Miller SE. 2005. DNA barcoding a useful tool for taxonomists. Nature 435:17.
59. The Global Biodiversity Information Facility website. http://www.gbif.org/ Accessed 1 November 2006.
60. Smith V. 2005. DNA barcoding: perspectives from a ''partnerships for Enhancing Expertise in Taxonomy'' (PEET) debate. Systematic Biology 54:841–844.
61. Tree of Life website. http://tolweb.org/tree/phylogeny.html Accessed 6 May 2006.