Journal of Fish Biology (2009) **74,** 329–356
doi:10.1111/j.1095-8649.2008.02080.x, available online at http://www.blackwell-synergy.com

REVIEW PAPER

The campaign to DNA barcode all fishes, FISH-BOL

R. D. Ward*†, R. Hanner‡ and P. D. N. Hebert‡

**CSIRO Wealth from Oceans Flagship, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia and ‡Biodiversity Institute of Ontario and Department of Integrative Biology, University of Guelph, Ontario, Canada*

(Received 20 February 2008, Accepted 2 September 2008)

FISH-BOL, the Fish Barcode of Life campaign, is an international research collaboration that is assembling a standardized reference DNA sequence library for all fishes. Analysis is targeting a 648 base pair region of the mitochondrial cytochrome *c* oxidase I (COI) gene. More than 5000 species have already been DNA barcoded, with an average of five specimens per species, typically vouchers with authoritative identifications. The barcode sequence from any fish, fillet, fin, egg or larva can be matched against these reference sequences using BOLD; the Barcode of Life Data System (http://www.barcodinglife.org). The benefits of barcoding fishes include facilitating species identification, highlighting cases of range expansion for known species, flagging previously overlooked species and enabling identifications where traditional methods cannot be applied. Results thus far indicate that barcodes separate *c.* 98 and 93% of already described marine and freshwater fish species, respectively. Several specimens with divergent barcode sequences have been confirmed by integrative taxonomic analysis as new species. Past concerns in relation to the use of fish barcoding for species discrimination are discussed. These include hybridization, recent radiations, regional differentiation in barcode sequences and nuclear copies of the barcode region. However, current results indicate these issues are of little concern for the great majority of specimens.

Key words: biodiversity; COI; cox1; cytochrome *c* oxidase; DNA barcoding; species identification.

INTRODUCTION

Historical methods of identifying, naming and classifying fishes are largely based on visible morphology. Although modern taxonomic work regularly employs many other traits, including internal anatomy, physiology, behaviour, genes, isozymes and geography, morphological characters remain the cornerstone of taxonomic treatments. However, there are difficulties in relying

†Author to whom correspondence should be addressed. Tel.: +61 3 6232 5370; fax: +61 3 6232 5000; email: bob.ward@csiro.au

primarily on morphology when attempting to identify fishes during various stages of their development not considered in original treatments or when examining fragmentary, partial or processed remains. Even when intact adult specimens are available, the morphological characters used to discern species can be so subtle that identification is difficult, even for trained taxonomists.

Large-scale fishery surveys add complexity – many taxonomic experts may be required to identify specimens from just a single collection. It is time-consuming and expensive to assemble and deploy such teams and to distribute specimens for identification. Moreover, accessing the historical literature and assessing the validity of species with a controversial taxonomic history are challenging tasks, even for experts. For the non-specialist faced with an assemblage of suboptimal specimens, identification can be extremely difficult if not impossible. This taxonomic impediment regularly hinders the assessment, conservation and management of global fish biodiversity.

Species identification by molecular analysis has been utilized for many years. Initially, allozyme differences were used (Avise, 1975), followed by mtDNA examination (Avise, 1994). DNA-based methods have several advantages over their protein-based counterparts because DNA is less sensitive to degradation (Hanner *et al.*, 2005) and can be accessed in all stages from egg to adult. Furthermore, synonymous mutations can be recognized in sequencing approaches, and polymerase chain reaction (PCR) amplification protocols make it possible to analyse minute amounts of tissue. Perhaps most importantly, DNA sequence data are easier to replicate and interpret across laboratories. Bartlett & Davidson (1991) were among the first to use mtDNA sequencing for fish identification, showing that cytochrome *b* sequences could discriminate four species of tuna (*Thunnus* spp.). They subsequently proposed forensically important nucleotide sequences (FINS; Bartlett & Davidson, 1992) as a means of identifying fishes.

Early studies using DNA sequences to discriminate species examined a variety of genes and different parts of these genes. This fluidity in analytical targets compromised the use of DNA sequencing for species identification because some knowledge of species identity was required to select the appropriate primers for the gene region that characterized a particular taxon. DNA barcoding differs from these earlier approaches as it proposed (Hebert *et al.*, 2003) that the sequence of a single gene region could be used as the basis of a global bio-identification system for animals. The availability of broad-range primers for the amplification of a 655 base pair (bp) fragment of cytochrome *c* oxidase subunit I (COI) from diverse phyla (Folmer *et al.*, 1994) established the 5′ end of this mitochondrial gene as a particularly promising target for species identification. COI encodes part of the terminal enzyme of the respiratory chain of mitochondria. Within-species variation for this gene is low compared with between-species variation. As a consequence, species are regularly delineated by a particular sequence or by a tight cluster of very similar sequences.

The effectiveness of this gene region for species-level identifications has now been validated for many groups of animals (Waugh, 2007), both invertebrate and vertebrate (*e.g.* springtails, Hogg & Hebert, 2004; butterflies, Hebert *et al.*, 2004*a*; Hajibabaei *et al.*, 2006*a*, but see Elias *et al.*, 2007; crustaceans, Costa *et al.*, 2007; Lefebure *et al.*, 2006; birds, Hebert *et al.*, 2004*b*; Yoo *et al.*, 2006; bats, Clare *et al.*, 2007). Barcoding has also been employed to

validate the identity of animal cell lines (Lorenz *et al.*, 2005; Cooper *et al.*, 2007) and is a recommended characterization step for materials in biodiversity repositories (Hanner & Gregory, 2007). Interestingly, the same gene region of COI has also been shown to be effective for species identification in red macroalgae (Saunders, 2005), in single celled protists (*Tetrahymena*; Chantangsi *et al.*, 2007) and for some fungi (Seifert *et al.*, 2007). Its power to discriminate closely related species is largely attributable to the abundance of synonymous nucleotide changes (Ward & Holmes, 2007).

However, barcoding has not been without controversy. Concerns have been expressed in relation to the use of a mitochondrial gene for species identification, its ability to discriminate recently radiated species and those species with extensive spatial differentiation and its ability to uncover new species (Moritz & Cicero, 2004; Hickerson *et al.*, 2006; Rubinoff *et al.*, 2006). There are certainly problems with some specific groups. For example, COI evolves very slowly in certain groups of benthic Cnidaria (France & Hoover, 2002; Shearer *et al.*, 2002) and hence does not deliver species-level resolution (Hebert *et al.*, 2003). Some have thus questioned the choice of target region (Erpenbeck *et al.*, 2006), while others have suggested sequencing a larger portion of the gene (Roe & Sperling, 2007). However, choice of a different marker would be difficult to justify for barcoding (reviewed in Neigel *et al.*, 2007). Meier *et al.* (2006) found that COI sequences generated only 70% success in species-level identification for Diptera using GenBank sequences, but many records in this repository are known to derive from misidentified specimens (Harris, 2003; Mitchell, 2008) and, or lack reference to the underlying specimens from which sequences were generated (Ruedas *et al.*, 2000; Pleijel *et al.*, 2008). Direct tests have now shown that COI delivers species-level resolution in various dipteran groups including tachinids (Smith *et al.*, 2006), mosquitoes (Cywinska *et al.*, 2006; Kumar *et al.*, 2007) and chironomid midges (Carew *et al.*, 2007; Ekrem *et al.*, 2007), although apparent introgressive hybridization associated with Wolbachia infection may make it an unreliable species indicator in blowflies of the genus *Protocalliphora* (Whitworth *et al.*, 2007).

The need for comprehensive and reliable species identification tools combined with early barcoding success with fishes (Savolainen *et al.*, 2005; Ward *et al.*, 2005) provoked the formation of the the Fish Barcode of Life campaign (FISH-BOL) initiative (http://www.fishbol.org). This campaign has the primary goal of gathering DNA barcode records for all the world's fishes, some 30 000 species. Given the estimated $US200 billion annual value of fisheries worldwide, FISH-BOL is addressing socially relevant questions concerning market substitution and quota management of commercial fisheries. Furthermore, the database is assisting the reconciliation of divergences in scientific, market and common names across nations. For ichthyologists, FISH-BOL promises a powerful tool for extending understanding of the natural history and ecological interactions of fish species. Although mutational saturation limits the value of COI barcode sequences for the independent resolution of deep-level phylogenetic relationships, barcode data are currently being incorporated into several 'tree of life' projects. Indeed, high-throughput barcoding is complementary to phylogenetic studies because it sheds light on divergent lineages for subsequent inclusion in such analyses (Hajibabaei *et al.*, 2007). Therefore, the data

generated from FISH-BOL will contribute to an ongoing synthesis of the evolutionary history of the most diverse group of vertebrates on earth.

BARCODING PROTOCOLS AND TOOLS

Large-scale digitization projects are providing enhanced access to the taxonomic literature needed by the global biodiversity community (*e.g.* the Biodiversity Heritage Library Project, http://www.biodiversitylibrary.org). Web-based repositories are compiling lists of valid taxonomic names and their synonymies (*e.g.* the Online Registry for Zoological Nomenclature, http://www.zoobank.org). When combined with online keys and high-resolution digital images, these advances are making taxonomic knowledge more accessible than ever before. However, these developments do not allow for species-level identifications of larval, juvenile, cryptic or fragmentary specimens. Standard reference DNA sequences collected from expertly identified morphological voucher specimens can be used to better characterize and broadly identify species. When properly referenced, the ensuing catalogue of unique genetic sequences or 'DNA barcodes' can conceptually unite diverse assemblages of specimens, collections and associated species information under a common registry of sequence accessions (Walters & Hanner, 2006). Such a reference database is essential for performing DNA-based identifications on unknown samples.

FISH-BOL is creating a valuable public resource in the form of an electronic database that contains DNA barcodes, images and geospatial co-ordinates for the analysed specimens. This information is initially organized and analysed using the BOLD (Barcode of Life Data System, http://www.barcodinglife.org; Ratnasingham & Hebert, 2007). The information is then delivered *via* a data feed to the FISH-BOL website, which uses a taxonomic authority file derived from FishBase (http://www.fishbase.org), the Catalog of Fishes (CoF; Eschmeyer, 2003) and the Integrated Taxonomic Information System (ITIS, see http://www.itis.gov) to monitor progress in barcode species coverage. In this respect, FISH-BOL complements and enhances existing genomics and fisheries databases.

GenBank includes thousands of fish DNA sequences. An unknown percentage of these derive from vouchered specimens, but nearly all past records lack any explicit connection to vouchers. As a result, cases of sequence disorder among supposedly conspecific individuals, which may reflect specimen misidentifications, cannot be easily resolved (Ruedas *et al.*, 2000; Pleijel *et al.*, 2008). Similarly, the raw sequence data from which GenBank submissions derive are rarely archived. Typically only the sequence record itself is deposited, preventing any critical evaluation of the reported nucleotide base calls. Thus, existing GenBank data are of limited utility for molecular diagnostic applications. The proposed adoption of 5′ COI for DNA barcoding was followed by calls for more stringent data quality standards by the barcode community (Lorenz *et al.*, 2005). The Consortium for the Barcode of Life (CBOL, see http://www.barcoding.si.edu) Database Working Group was convened to address this challenge. Together with GenBank and the other members of the International Nucleotide Sequence Database Collaboration (INSDC), a reserved keyword was established for sequences that meet an emerging community data standard (Walters & Hanner, 2006).

The INSDC will archive DNA sequences from the FISH-BOL campaign and will annotate each sequence with the keyword 'BARCODE' when it meets the barcode data standard. Such sequences must be associated with a valid species name and include a bidirectionally sequenced record of at least 500 bp from the 5′ end of COI. Details concerning the voucher specimen are also required, including co-ordinates for the collection locality, collection date, collector and identifier. Wherever possible, voucher specimen should be stored in a publicly accessible collection (Pleijel *et al.*, 2008). Sometimes, for a variety of reasons, including the expense of storing large fish or the unavailability of whole specimens of endangered species, retention of traditional morphological vouchers may not be achievable, in such situations, one or more photographic images of the specimen can be retained as e-vouchers (Monk & Baker, 2001) along with other provenance data to further document the collection event from which tissues were obtained for barcoding.

In addition, BARCODE annotated entries require reference to the PCR primers used to generate the reference sequences and archival of the underlying electropherogram 'trace files' in a publicly accessible repository such as the National Center for Biotechnology Information (NCBI) Trace Archive (Walters & Hanner, 2006). Together, these elements establish 'the fitness for use' of barcode data in molecular diagnostic applications. The existence of standard symbolic codes for institutional resource collections in ichthyology (Leviton *et al.*, 1985; http://www.asih.org/codons.pdf), combined with the taxonomic treatments of CoF, FishBase and ITIS, provides an excellent organizational framework for conducting the FISH-BOL campaign. CBOL, GBIF (the Global Biodiversity Information Facility, http://www.gbif.org) and GenBank are using this information to compile an online registry of biological repositories (http://www.biorepositories.org) to establish a structured reference to barcoded voucher specimens held in existing reference collections. These subtle but important distinctions differentiate barcoding from historical approaches such as FINS (Bartlett & Davidson, 1992).

Historically, there were far more cytochrome *b* than COI sequence records for vertebrates in GenBank as a consequence of the wide use of varied primer sets for the former gene (Kocher *et al.*, 1989). However, this imbalance has now been redressed. Cytochrome *b* data are available for 15 000 specimens representing 3000 fish species, and these sequences are not taken from a standard region of the gene. By comparison, the fish barcode data set on BOLD contains 26 000 sequences from 5000 species, all derived from a standard segment of COI. Moreover, because the records on BOLD generally adhere to the BARCODE data standard, they provide a reliable basis for the identification of unknown samples. For these reasons, community support is building rapidly for fish COI barcoding. Indeed, rapid growth and active curation of the FISH-BOL data set on BOLD mean that it is already a more effective tool for fish identification than GenBank (Wong & Hanner, 2008).

Barcoding typically uses inexpensive high-throughput extraction (Ivanova *et al.*, 2006), and sequencing (Hajibabaei *et al.*, 2005) protocols coupled with primers and primer cocktails developed for fishes (Ward *et al.*, 2005; Ivanova *et al.*, 2007). Tissues destined for barcoding can be fresh, frozen or stored using a suitable preservative (*e.g.* RNALater®, FTA® cards, EDTA lysis buffer, 95%

alcohol). However, cooked or processed fish can also be barcoded (see later). DNA extraction from formalin-preserved tissue remains difficult, despite several attempts to resolve this problem (Klanten *et al.*, 2003; Chakraborty *et al.*, 2006). Development of a reliable and inexpensive protocol to retrieve DNA from formalin-preserved specimens will be invaluable to tap into historically vouchered specimens in museums throughout the world.

FISH-BOL has adopted the BOLD (Ratnasingham & Hebert, 2007) as an online workbench, and it is used to relate a given barcode record to both a voucher specimen and to other barcode sequences. Various options exist in BOLD for generating genetic distance estimates and neighbour-joining phenograms from specimen barcode sequences. The system can also plot specimen collection localities on a distribution map with a resolution of 1 km pixel^{-1} and facilitates morphological comparison of voucher specimens when appropriate digital images are uploaded. Even when these images are not necessarily diagnostic to species, such images can be extremely useful (*e.g.* e-vouchers, *sensu* Monk & Baker, 2001). They do provide a check on identification accuracy, providing that species or generic diagnostic features are visible. They can act as surrogates when specimens cannot be deposited in a museum (*e.g.* when they are too large for deposition or remain alive) and are widely available to a broad group of users.

BOLD is a collaborative workbench containing both public data and private projects being prepared for publication. Most sequence records in BOLD are password protected, reflecting their 'ownership' by the participants in each specific project. While the sequences and collateral data generally remain unavailable to the public until published, all sequence records are made available for identifying unknown specimens using the BOLD identification engine. All barcode data will eventually become publicly accessible, although the taxonomic annotation of specimens requires careful data validation prior to publication.

THE FISH-BOL CAMPAIGN

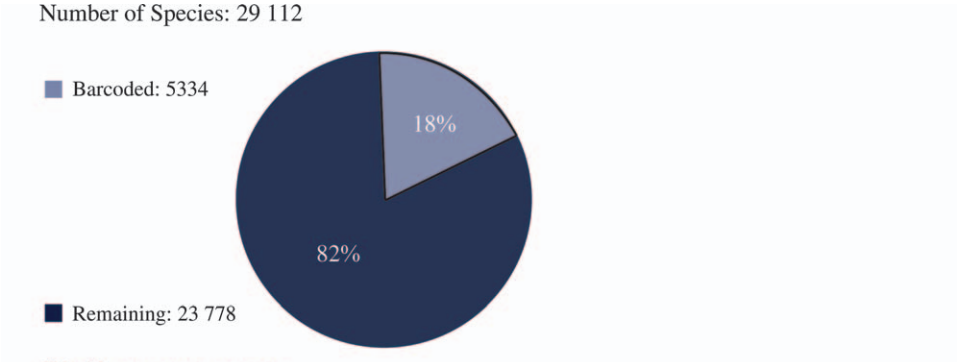
The FISH-BOL campaign (http://www.fishbol.org) was conceived in 2004, and in June 2005, >50 fish geneticists, taxonomists and interested parties met at the University of Guelph to consider its activation. FISH-BOL has two co-chairs, a campaign co-ordinator and 11 regional working groups – Africa, Australia, Europe, India, Meso-America, Middle East, North America, North East Asia, Oceania–Antarctica, South America and South East Asia. The chair, deputy chair and members of each working group are listed (with affiliations and email addresses) on the FISH-BOL website. About 160 researchers are now involved in this campaign, as geneticists, taxonomists or collectors. New participants are always welcome and should contact the appropriate regional working group chair to learn more about the needs and ongoing organizational activities taking place in their region.

The overarching goal of FISH-BOL is to barcode all the world's fishes. Initially, it aims at obtaining barcode records from a minimum of five specimens per species, increasing to at least five specimens per FAO area. It may be necessary to analyse more specimens for freshwater species because of their greater degree of spatial genetic differentiation (Ward *et al.*, 1994). However, because

collections depend on the interests and locations of FISH-BOL participants, some species will be widely barcoded, while others will receive little attention. Unsurprisingly, much of the early work has involved species of commercial importance, largely marine fishes. It is estimated that at least 200 000 specimens will need analysis to generate a comprehensive barcode library for marine fishes alone.

The FISH-BOL website details the vision and goals of FISH-BOL, the participants and current progress (overall or by regional working group). It also provides a search engine that returns barcoding progress for any particular taxon or species, a map of collecting points for any barcoded species, a checklist for fishes of the world (derived from FishBase) and various short news items and links to other relevant databases.

As of late-July 2008, a total of 29 112 specimens representing 5334 fish species had been barcoded (Fig. 1) representing a mean of 5·46 individuals per species. However, while the median is 3, the mode is just 1, indicating that the FISH-BOL campaign still has a long way to go to meet its sample size targets. New participants are therefore welcomed to this campaign. In actuality, more species have been barcoded: the species total does not include taxa with provisional names (*e.g. Apristurus* sp. *A* and *Dasyatis cf. annotata*) or specimens lacking an FAO area designation (because this is the basis for tracking progress among regional working groups). Over the past 18 months, the number of specimens analysed has risen rapidly, but growth in species coverage has slowed as multiple barcodes are collected for common species with wide geographical ranges and commercial significance. Cartilaginous fish have been



Working group summary

Class	Barcoded	Species	Progress (%)	Class	Barcoded	Species	Progress (%)
Actinopterygii	4942	27 984	18	Cephalaspidomorphi	14	42	18
Elasmobranchii	353	968	36	Holecephali	15	37	36
Myxini	8	70	11	Sarcopterygii	2	11	11

FIG. 1. Global progress in FISH-BOL, by class (as of July 28, 2008). Numbers of species barcoded and total numbers of species are given. This figure is from the FISH-BOL website, which allows barcoding progress at any taxonomic rank or of any regional working group to be assessed.

particularly well studied (Table I), reflecting the interests of some participants and the commercial or conservation significance of certain shark groups.

RESULTS FROM FISH BARCODING

EXTENT OF COI DIVERGENCE AMONG DIFFERENT TAXONOMIC LEVELS AND SPECIES' UNIQUENESS

To illustrate progress and to highlight the species discriminatory power of fish barcoding, sequences from four published studies on Australasian fishes were combined (Ward *et al.*, 2005; Ward & Holmes, 2007; Ward *et al.*, 2007; Ward *et al.*, 2008*b*). These records include data from 546 species represented by 1677 specimens. Most are marine, but some freshwater and estuarine species are included.

Average intraspecific differentiation or distance (*D*, Kimura's two-parameter model; Kimura, 1980), for the 294 species with multiple specimens (an average of 4·85 specimens per species), is 0·35%, while the average congeneric differentiation for the 103 genera with multiple species (average of 3·67 species per genus) is 8·11% (Table II). These results suggest a strong correlation between morphological taxonomy and molecular divergence among congeners, a result typical for barcoding studies. This has led some (Hebert *et al.*, 2004*b*; Lefebure *et al.*, 2006) to suggest a molecular threshold to aid species delimitation. If just a single specimen is taken per species for the congeneric estimates, the average congeneric distance increases to 9·64%. Thus, at present, the well-sampled genera (in terms of individuals per species) have lower rates of divergence than the less intensively studied genera. This likely is only a sampling artefact, and this difference will disappear as more genera are sampled.

Nearly 62% of intraspecific divergences are $<0{\cdot}2\%$ *D* and 93% are $<1{\cdot}0\%$ *D* (Fig. 2). A small secondary mode *c.* 3% *D* is composed almost entirely of values from two rays, *Dasyatis kuhlii* (Müller & Henle) and *Dasyatis leylandi* Last. By contrast, the corresponding values for congeneric comparisons are just

TABLE I. Current global progress for elasmobranchs, by order (as of July 28, 2008)

	Numbers of species		
	Barcoded	Total	Progress
Carcharhiniformes	106	230	46%
Heterodontiformes	4	9	44%
Hexanchiformes	5	5	100%
Lamniformes	14	16	88%
Orectolobiformes	14	33	42%
Pristiformes	3	7	43%
Pristiophoriformes	3	5	60%
Rajiformes	145	483	30%
Squaliformes	40	103	39%
Squatiniformes	6	17	35%
Torpediniformes	13	60	22%

TABLE II. Summary of genetic divergences (% K2P distance) within various taxonomic levels. Data are from 1677 sequences from 546 species (294 represented by multiple specimens[a]) and 273 genera (103 represented by multiple species)

			Distance	
Comparisons within	Number of comparisons	Mean	Minimum	Maximum
Species	3851	0·35 ± 0·01	0	10·91[b]
Genera	20 371	8·11 ± 0·04	0	24·18
Family	21 723	16·19 ± 0·04	0·63	35·79
Order	185 393	21·74 ± 0·01	8·19	38·03
Class	443 606	24·46 ± 0·01	14·27	37·47

K2P, Kimura's two-parameter model.

[a]252 species are represented here by single representatives. Most of these 252 species exist as multiple specimens in BOLD but their barcodes are yet unpublished – one of the four published data sets contributing to this table (see text) deliberately restricted itself to the analysis of single specimens per species (Ward & Holmes, 2007).

[b]A single shortnose sawshark *Pristiophorus nudipinnis* with a barcode typical of *Pristiophorus cirratus* – most likely a misidentification as there appear to be no records of hybridization among shark species.

0·32% for the 0–0·2% *D* range and only 3·40% for the 0–1·0% *D* range. Thus, while some 93% of intraspecies distance values are <1% *D*, only *c.* 3% of congeneric comparisons fall in this range. The overall distribution of congeneric distances shows some bimodality, which is almost absent from the within-family range (Fig. 3). This bimodality is also likely an artefact of the current sampling, with more sampling of recently diverged genera than of older genera.

The BOLD database not only provides these hierarchical analyses of distance but also offers a nearest-neighbour approach for highlighting apparent conflicts. For the combined data set, this analysis shows that for species represented by multiple specimens, the nearest neighbour of any sequence is a conspecific, except for three instances. The exceptions are *Urolophus cruciatus* (Lacépède) and *Urolophus sufflavus* Whitley, *Carcharhinus cautus* (Whitley) and *Carcharhinus melanopterus* (Quoy & Gaimard), and *Pristiophorus nudipinnis* Günther and *Pristiophorus cirratus* (Latham). These exceptions are discussed below.

Just 28 of 20 371 pair-wise sequence comparisons from the barcode records of congeners show sequence identity ($D = 0$), and no cases of sequence sharing are apparent for more distantly related taxa. Most of these instances of sequence identity involve two species of sting ray, *U. cruciatus* and *U. sufflavus*, which cannot be separated by their COI barcodes (Ward & Holmes, 2007; Ward *et al.*, 2008*b*). The remainder derive from one of five specimens of the shark *C. cautus*, which cannot be separated from *C. melanopterus*, and from one of nine specimens of the sawshark *P. nudipinnis*, which cannot be separated (Ward *et al.*, 2008*b*) from *P. cirratus*. These latter instances might be attributable to hybridization, but because hybridization is unknown among chondrichthyans (Gardner, 1997), misidentification is the most likely cause. Unfortunately, these two specimens were not retained and cannot be

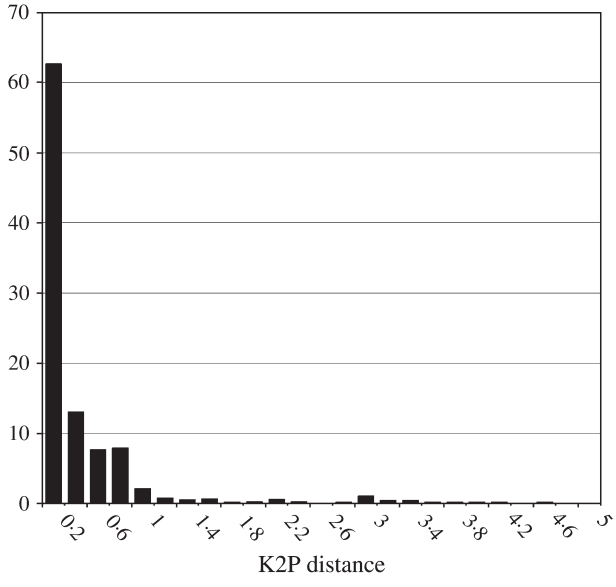


FIG. 2. Distribution of within species Kimura's two-parameter model (K2P) distances (%) from 294 species with multiple specimens per species.

morphologically re-examined, underscoring the need for voucher retention wherever possible. Physical voucher specimens are also critical in cases involving cryptic species where photographs cannot substitute for a specimen.

In summary, the two sting rays are the only two (0·37%) of the 546 species considered in this study that cannot be distinguished using the COI barcode, or, more conservatively, two (0·68%) of the 294 species with multiple specimens per species. Overall, *c.* 99% of fish species in this analysis can be distinguished by COI barcode analysis. A few other instances of barcodes failing to discriminate among marine fishes are on record. Spies *et al.* (2006) analysed the COI sequences of the 15 species of skates and found two (*Bathyraja lindbergi* Ishiyama & Ishihara and *Bathyraja maculata* Ishiyama & Ishihara) that could not be separated. A study of Scotia Sea fishes revealed that barcoding enabled species discrimination in all cases except for *Bathydraco* and *Artedidraco*, where introgressive hybridization or misidentification may be involved (Rock *et al.*, 2008).

Data from freshwater fishes are more limited, but shared haplotypes were observed in 7–8% of the 190 Canadian species (Hubert *et al.*, 2008). For these samples, identifications were only possible to species groups. It was conjectured that the species that could not be unequivocally separated reflected recent radiation, introgressive hybridization or erroneous taxonomy (some species pairs with overlapping haplotypes perhaps being single species).

SPATIAL DIFFERENTIATION WITHIN SPECIES

Extensive intraspecific geographical differentiation in barcode frequencies, particularly if barcode sequences vary from region to region, could pose

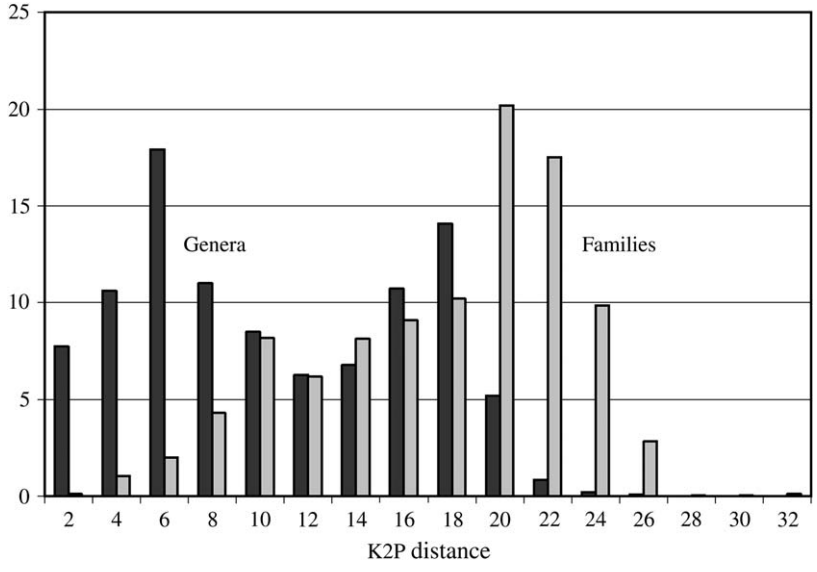


FIG. 3. Distribution of within genera (black) and within family (grey) K2P distances (%) from 546 species generated using a single individual per species.

problems for species identification (Moritz & Cicero, 2004). Is this likely to be an issue for fishes?

While most species within the FISH-BOL campaign have relatively few barcode records (mean = 5·34), sample sizes are growing. A few species already have large numbers of records from many different localities. At present, the most intensely barcoded fish is the tiger shark *Galeocerdo cuvier* (Péron & Lesueur), with 171 specimens from the Atlantic, Pacific and Indian Oceans; they show a mean divergence of just 0·09% and a maximum divergence of only 0·47%. This pattern is typical of most marine fish species, even those with very wide geographical ranges. For example, the piked dogfish *Squalus acanthias* L. from the North Atlantic to the South Pacific (Fig. 4) shows essentially no spatial differentiation (Fig. 5, clade A, divergence 0·26%). Some specimens from the U.K., Eastern U.S.A., Uruguay, Argentina, Chile and Tasmania have identical haplotypes. There is a separate mtDNA clade in the North Pacific, but it too shows very little intra-clade divergence (Fig. 5, clade B, divergence 0·10%). Interestingly, the North Pacific population was once (Jensen, 1966) recognized as a separate species, *Squalus suckleyi* (Girard), and perhaps, its taxonomy needs to be revisited. Minimal geographical differentiation for COI barcodes is also seen for many teleosts. Intraspecies divergence of the Mediterranean slimehead *Hoplostethus mediterraneus* (Cuvier) barcoded from Portugese, South African and Tasmanian waters is 0·26% and intraspecies divergence of orange roughy *Hoplostethus atlanticus* (Collett) from the mid-Atlantic and Tasmania is 0·51%. Neither species shows any clustering of COI haplotypes by location (Fig. 6; Ward *et al.*, 2008*c*).

At a smaller spatial scale, chondrichthyans in Australasian waters collected thousands of kilometres apart tend to show either minimal or undetectable

FIG. 4. Map from FISH-BOL for *Squalus acanthias*, showing the origin of the barcoded samples with GPS locations. Pointers can be clicked for additional information.

COI divergence among specimens or extensive divergence >3% (*e.g. D. kuhlii* and *D. leylandi*). The latter is more on a par with levels expected of species differentiation, likely flagging cryptic species (Ward *et al.*, 2008*b*).

Barcoding data are currently more limited for freshwater fishes, which show greater spatial genetic differentiation than marine species (Ward *et al.*, 1994). However, the first study of DNA barcoding in freshwater fishes did not reveal deeper intraspecific or interspecific divergences than marine fishes, concluding that barcoding will be a powerful tool for freshwater fish discrimination (Hubert *et al.*, 2008).

In summary, geographical differentiation does not appear to pose a significant issue for barcode-mediated identification of marine or freshwater fishes, as the barcode array for one species remains distinct from those found in other species.

DEEP DIVERGENCES WITHIN SPECIES

Deep divergence between COI haplotypes within nominal species may flag overlooked fish species. One such instance in the Australian study is that of the barramundi, *Lates calcarifer* (Bloch). There is no barcode divergence among specimens of this teleost from Australia or among Myanmar specimens imported into Australia but a deep divergence (*c.* 9·5%) between them (Ward *et al.*, 2008*a*). Other possible examples concern the ray genus *Dasyatis* where, as already mentioned, there is high 'intraspecific' divergence (*c.* 3–4%) within both *D. kuhlii* and *D. leylandi*, and the divergent lineages in each species are allopatric with little intra-lineage diversity. While this could represent regional variation within a species, the level and pattern of divergence is characteristic of that found between species. There are other examples of deep divergences

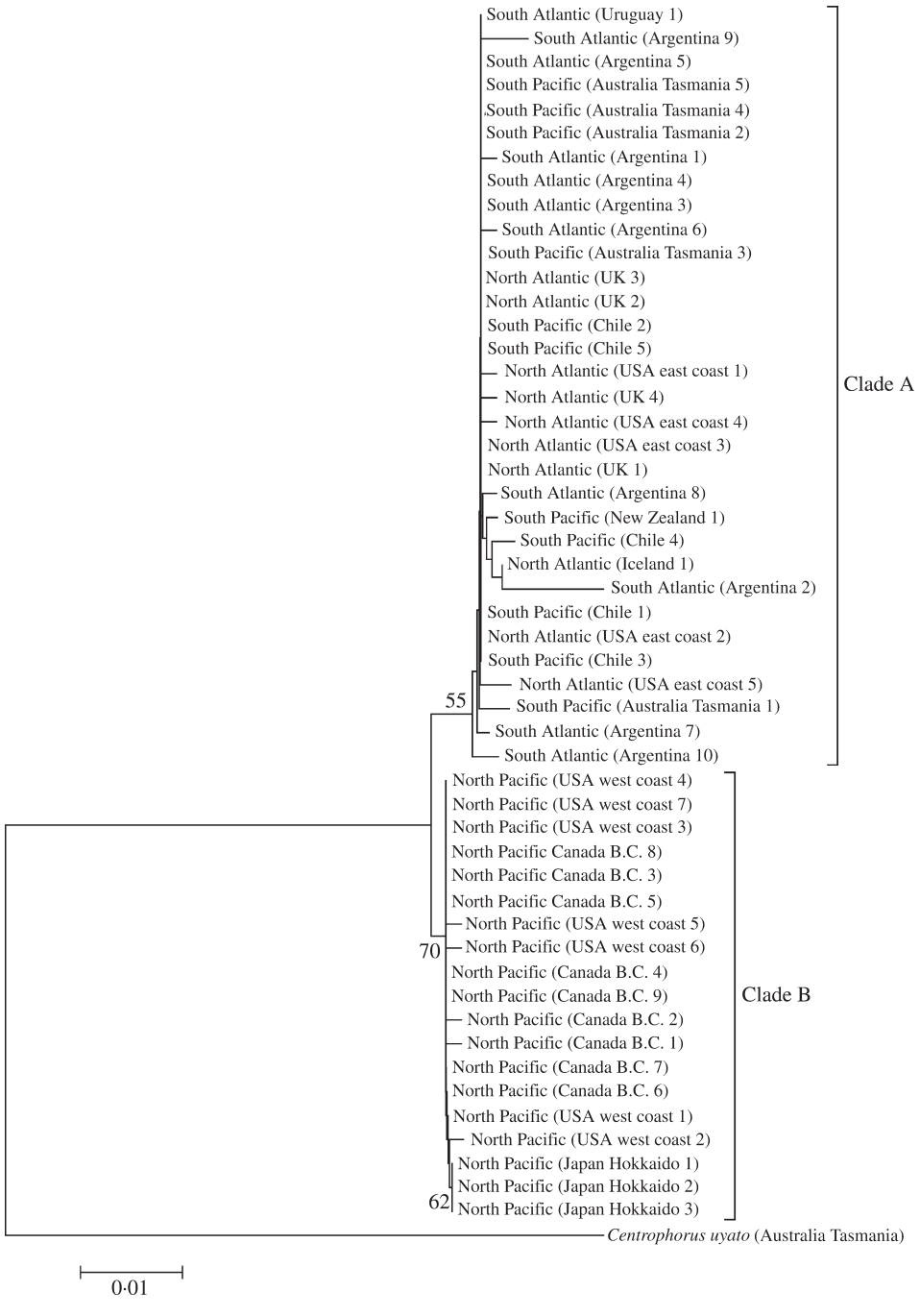


FIG. 5. The two COI clades of the piked dogfish, *Squalus acanthias*, clade A from the Atlantic and South Pacific ($n = 32$), clade B from the North Pacific ($n = 19$). 1000 bootstraps, values >50% given, as is a K2P distance scale bar. Outgroup is the little gulper shark, *Centrophorus uyato*.

within chondrichthyans (Ward *et al.*, 2008*b*), illustrating how barcoding can reveal hidden diversity and highlight taxa of potential revisionary interest.

Hebert *et al.* (2004*b*) proposed that specimens showing >10 times the average intraspecies distance for the group under study should be flagged as provisional new species. Based on the average intraspecific sequence divergence, this would suggest application of a threshold of 3·5% for the fish species considered in this study. This threshold includes those species with high intraspecies divergences such as *D. kuhlii* and *D. leylandi*, which might themselves contain cryptic species. Exclusion of intraspecies divergences of 3% or greater reduces the species level threshold flag from 3·5 to 2·8%. Under either criterion, the *Lates* case would certainly qualify for provisional species status, with the *Dasyatis* cases borderline. While this thresholding approach can usefully flag situations worthy of more detailed taxonomic study, it will necessarily fail to recognize recently radiated or closely related species (Meyer & Paulay, 2005; Hickerson *et al.*, 2006). In fact, *c.* 15% of recognized fish species in the survey group have congeneric distances <2·8%, and 3·4% have distances <1·0%. Clearly, there is no absolute distance value that can be employed as a hard criterion so that values above indicate interspecific divergence, while those below are intraspecific.

SPECIES DISCOVERY AND VALIDATION

Fish specimens exhibiting deep intraspecies divergence for COI require detailed analysis by taxonomic experts before they can be confirmed (or rejected) as new species. Genetic analysis has been used for >40 years to flag the existence of cryptic species, so DNA barcoding is far from unique in this respect. An early example is the discovery of sibling species of sea cucumbers by allozyme electrophoresis (Manwell & Baker, 1963), and more recently, DNA analysis supported the discovery of a new large mammal species from Vietnam (Dung *et al.*, 1993). DNA barcoding primarily differs from past applications of genetics to identification and taxonomy by the scale of the enterprise, at least as regards numbers of target species, and by the consistent use of a single DNA region. Once flagged as possible new species, taxonomic validation requires detailed morphological and meristic description and diagnosis of voucher specimens, together with consideration of geographical distribution and, perhaps, ecological factors. DNA barcoding can and should be 'fully integrated into taxonomic practise' (Padial & de la Riva, 2007). Inclusion of sequence data from other mitochondrial or nuclear markers might also be warranted. It is emphasized in this study that while the inclusion of the COI barcode sequence provides a powerful baseline for flagging taxa worthy of further study, it does not imply acceptance of a 'mitochondrial DNA species concept', rather it recognizes the role of molecular analysis in contemporary systematic and taxonomic studies.

The first Australian fish barcode study (Ward *et al.*, 2005) examined a number of taxa that had only gained provisional recognition from morphological examination. In every case, these were separable by their barcodes, corroborating their likely species status. Some have subsequently been named; for example, species of the dogfish genus *Squalus* initially barcoded and referred to as *Squalus* sp. B, C, D, E and F have now gained formal scientific names (Last

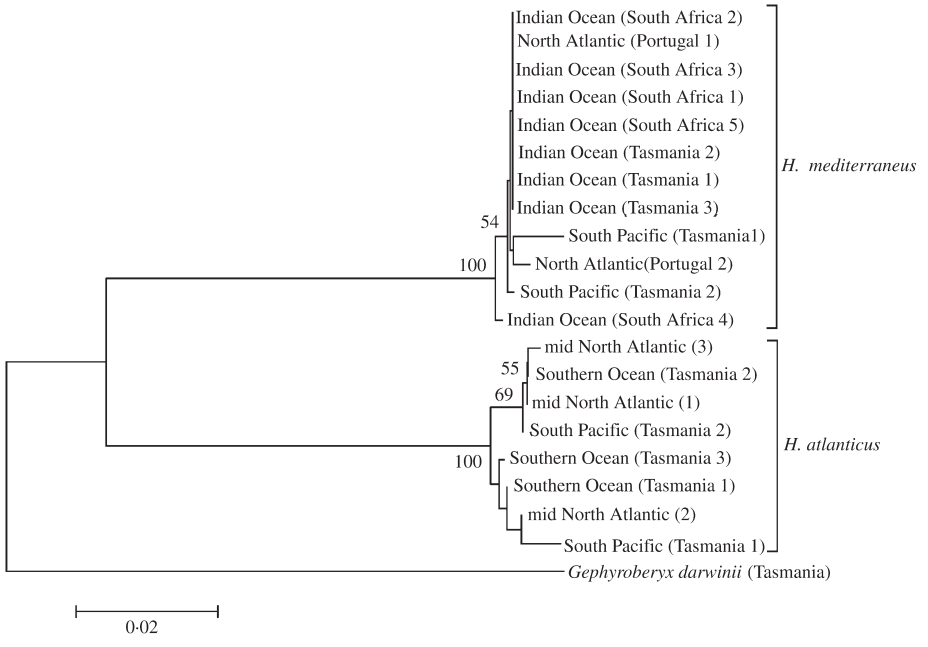
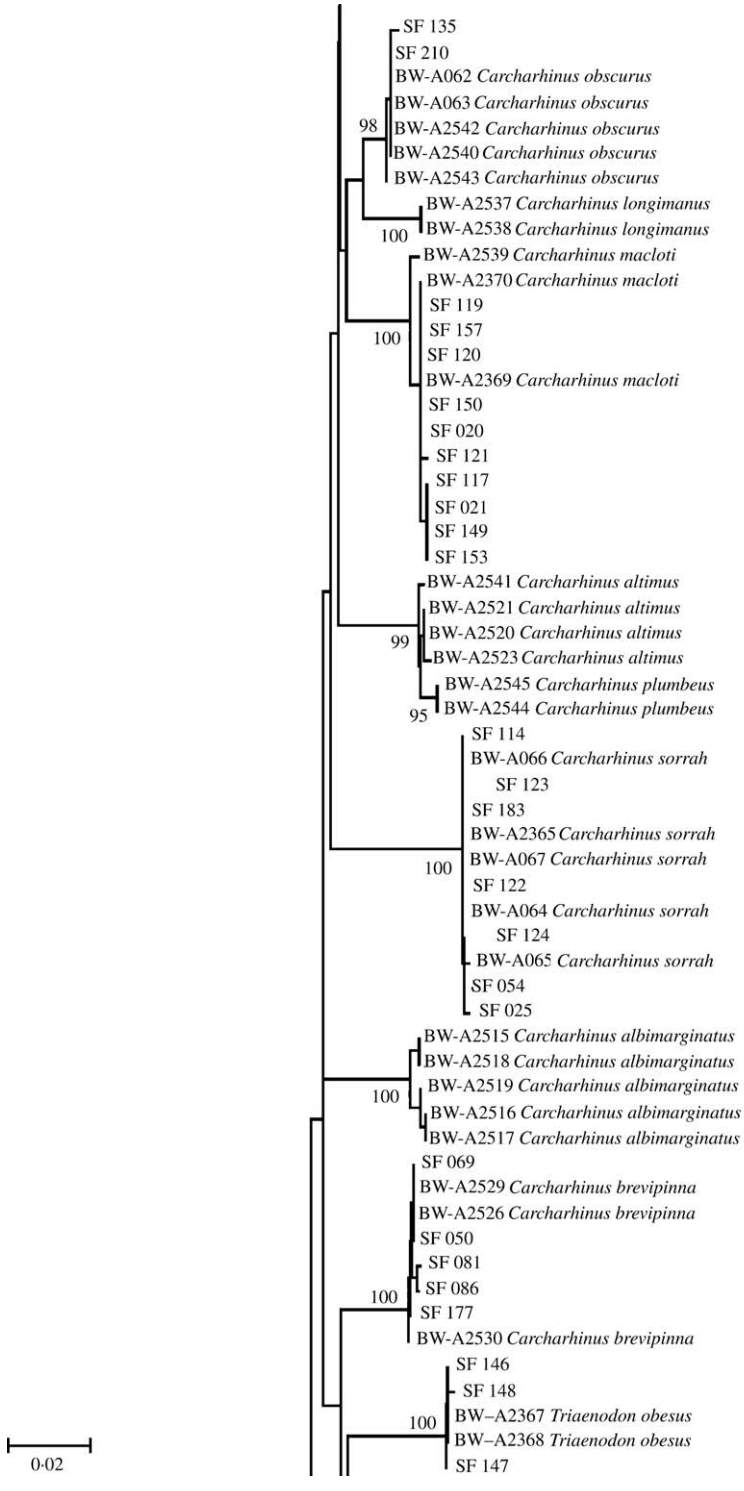


FIG. 6. K2P distance tree of two teleost species, Mediterranean slimehead, *Hoplostethus mediterraneus* (n = 12), and orange roughy, *Hoplostethus atlanticus* (n = 8), from widely divergent locations. 1000 bootstraps, values >50% given, as is a K2P distance scale bar. Outgroup is Darwin's slimehead, *Gephyroberyx darwinii*.

et al., 2007*a*; Ward *et al.*, 2007). The first vertebrate to include a COI barcode as part of its species description was the goby *Coryphopterus kuna* Victor (Victor, 2007). COI barcoding has helped validate a second new species of goby (Victor, 2008), a new sting ray (*Urolophus kapalensis*, Yearsley & Last), a handfish *Brachionichthys australis* (Last, Gledhill & Holmes), (Last *et al.*, 2007*b*) and five new species of damselfish (*Chromis*, Pyle *et al.*, 2008). Ichthyologists are beginning to recognize the value of including a barcode in the description of new species, a trend that further distinguishes barcoding from earlier molecular efforts at species discrimination.

APPLICATIONS OF FISH BARCODING TO SPECIMEN IDENTIFICATION

The simplest way of attempting the identification of an unknown specimen is to paste its COI sequence into the BOLD identification engine. This then compares that sequence with all public and private BOLD sequences and all available GenBank sequences of the COI barcode region. GenBank records are regularly downloaded to BOLD. Although some GenBank records include a larger mtDNA region, the COI 5′ region alone is used by BOLD for matching. Thus, sequences are available from species that have not been examined through FISH-BOL (such as the coelacanths *Latimeria chalumnae* Smith and *Latimeria menadoensis* Pouyaud, Wirjoatmodjo, Rachmatika, Tjakrawidjaja,


SF 135
SF 210
BW-A062 Carcharhinus obscurus
BW-A063 Carcharhinus obscurus
98
BW-A2542 Carcharhinus obscurus
BW-A2540 Carcharhinus obscurus
BW-A2543 Carcharhinus obscurus
BW-A2537 Carcharhinus longimanus
100
BW-A2538 Carcharhinus longimanus
BW-A2539 Carcharhinus macloti
BW-A2370 Carcharhinus macloti
SF 119
100
SF 157
SF 120
BW-A2369 Carcharhinus macloti
SF 150
SF 020
SF 121
SF 117
SF 021
SF 149
SF 153
BW-A2541 Carcharhinus altimus
BW-A2521 Carcharhinus altimus
99
BW-A2520 Carcharhinus altimus
BW-A2523 Carcharhinus altimus
BW-A2545 Carcharhinus plumbeus
95
BW-A2544 Carcharhinus plumbeus
SF 114
BW-A066 Carcharhinus sorrah
SF 123
SF 183
BW-A2365 Carcharhinus sorrah
100
BW-A067 Carcharhinus sorrah
SF 122
BW-A064 Carcharhinus sorrah
SF 124
BW-A065 Carcharhinus sorrah
SF 054
SF 025
BW-A2515 Carcharhinus albimarginatus
BW-A2518 Carcharhinus albimarginatus
100
BW-A2519 Carcharhinus albimarginatus
BW-A2516 Carcharhinus albimarginatus
BW-A2517 Carcharhinus albimarginatus
SF 069
BW-A2529 Carcharhinus brevipinna
BW-A2526 Carcharhinus brevipinna
SF 050
SF 081
100
SF 086
SF 177
BW-A2530 Carcharhinus brevipinna
SF 146
SF 148
100
BW–A2367 Triaenodon obesus
BW–A2368 Triaenodon obesus
SF 147
0·02

Hadiaty & Hadie). BOLD then displays the top 20 matching records (default value or 10, 50 or 99 records) with similarity values, and a neighbour-joining phenogram or identification 'tree' of the 100 nearest-neighbours is available if desired. Ratnasingham & Hebert (2007) provide further details about the BOLD search engine. If there is no sequence in the reference data set that shows <1% divergence from the query sequence, the search engine lists the closest-matching sequences but does not assign a species-level identification. Clearly, wherever possible, the reference data set should be densely sampled to improve identification accuracy – this is a major goal of FISH-BOL.

Ross *et al.* (2008) have used simulations to examine the performance of distance and tree-based methods for species identification under a variety of scenarios. Where the correct species is absent from the reference data set, they recommend a strict tree-based approach coupled with a distance threshold to protect against false positives. Errors in species diagnosis might also be made in other circumstances, including species paraphyly or hybridization, or, of course, when reference specimens have been misidentified. If reference sequences are available, there are alternative, non-distance-based, identification procedures for unknown sequences. These include character-based approaches (Kelly *et al.*, 2007; Rach *et al.*, 2008) or new methods such as back-propagation neural networks (Zhang *et al.*, 2008). However, these methods are currently too computationally intensive to offer identifications in real-time, and must therefore be implemented on a restricted set of taxa, on a case-by-case basis. It has been suggested that initially the entire database be searched with similarity methods to isolate the likely genus or family, then more computationally intensive methods be used to determine the precise species (Frézal & Leblois, 2008). While this is a very important and evolving area of debate, all identification algorithms eventually rely on having a comprehensive reference database (Ekrem *et al.*, 2007).

DNA barcoding can be used to identify whole fish, fillets, fins, fragments, juveniles, larvae or eggs. Fillets or fins often require molecular analysis because the morphological traits used for species identification are generally absent. Identification of eggs and larvae is similarly difficult because discriminating characteristics are often either absent or undescribed. The ability of barcoding to provide unequivocal species assignments from whole or part specimens has important implications for fish retailing (accidental or intentional species substitution, consumer protection and regulation of trade in ornamental species), fish management (quota and by-catch monitoring, sustainable fisheries), fish conservation (identification of threatened, endangered and protected species or parts thereof), fish invasions (recognition of range changes associated with climate change and introductions) and fisheries and aquatic research generally (Costa & Carvalho, 2007). It is important to note that only 65·9% of fishery captures reported to the FAO for 1996 were at a species level, ranging from *c.* 90% in temperate areas to sometimes <40% for tropical regions (Caddy &

FIG. 7. Part of a much larger neighbour-joining tree based on genetic distances that illustrates the identification of shark fins (SF) by COI barcoding. Known reference samples are given along with their BOLD numbers. Distance scale (K2P) given. Bootstrap values ≥95% given.

Garibaldi, 2000). Surveys into the accuracies of species identifications apparently have not been reported, but many 'identifications' may still be erroneous.

COI barcoding has been validated as a diagnostic marker for genetic species identification (Dawnay *et al.*, 2007) and is under consideration by the Food and Drug Administration in the U.S.A. as a replacement for the technique of protein isoelectric focusing for fish and fish product identification (Yancy *et al.*, 2007). The need is great. Previous cytochrome *b* sequencing showed that between 60 and 94% of fish labelled as red snapper *Lutjanus campechanus* (Poey) in the U.S.A. were mislabelled (Marko *et al.*, 2004), and a recent COI barcoding study of fish samples from markets and restaurants in North America revealed that some 25% of analysed specimens were mislabelled (Wong & Hanner, 2008).

The forensic identification of specimens from mtDNA sequences is not new (Murray *et al.*, 1995; Palumbi & Cipriano, 1998), but FISH-BOL will build a comprehensive reference database for fishes, based on the COI barcode sequence. Already, most commercial food-fish species are represented in the BOLD reference library and have distinguishable COI barcodes. Thus, the question can now shift from, for example, 'Is this really a barramundi?', with the response 'Yes' or 'No', to 'What species is this fillet?', with the response 'It is species X'.

DNA barcoding has been used to gain species identifications for shark fins confiscated from illegal fishers in northern Australian waters (B. H. Holmes & R. D. Ward, unpubl. data). Barcodes from 145 fins (left pectoral fins, so that each was known to have come from a different individual) were matched to reference sequences on BOLD and were identified as deriving from 26 species of sharks and rays from 12 genera. Matching involved both the placement on neighbour-joining trees (Fig. 7) and the sequence similarity estimations. Although assignments appeared robust, the shark genus *Carcharhinus* includes some 30 species, 27 currently with barcodes, and some of these, such as *Carcharhinus altimus* (Springer) and *Carcharhinus plumbeus* (Nardo), are very closely related (*C. plumbeus* may be paraphyletic; Fig. 7). In such instances, COI may not enable rigorous species discrimination and an additional more rapidly evolving marker might be required. While there are some species-specific DNA probes available for sharks (Shivji *et al.*, 2002), these would only have identified a minority of the species present and are unavailable for the most abundant species in this collection – the small shark *Carcharhinus dussumieri* (Müller & Henle).

DNA barcoding can also be used for cooked or processed fish. Smith *et al.* (2008) obtained species-level identifications for smoked fillets of the genera *Anguilla*, *Polyprion*, *Seriolella*, *Thunnus* and *Trachurus*. In this instance, samples were correctly labelled with respect to common name, although smoked product labelled as the mackerel *Trachurus novaezelandiae* Richardson was actually derived from the morphologically very similar species *Trachurus declivis* (Jenyns). Grilled and deep-fried fillets can also be identified by barcoding (R. Hanner & R. D. Ward, pers. obs.)

COI barcoding has also been used to identify fish larvae taken from the Great Barrier Reef (Pegg *et al.*, 2006) and the Caribbean (Victor, 2007) and fish eggs from temperate Australian waters (F. Neira, pers. comm.). It worked well, although not all sequenced specimens could be identified to species because the

required reference library was incomplete. As more species are barcoded, the identification power of BOLD will increase.

DNA barcoding may also allow a deeper understanding of food webs by enabling the identification of fish species in gut contents. Prey items in stomachs or highly processed samples such as canned fish are likely to have partially degraded DNA that may not allow amplification of full-length barcodes. Genetic screening of gut contents has begun (Deagle *et al.*, 2005; Parsons *et al.*, 2005; Sigler *et al.*, 2006). A 250 bp COI fragment has been used to screen billfish stomach contents (Paine *et al.*, 2007), but this sequence had little overlap with the 655 bp barcode region and could not utilize the BOLD reference library. Given the large amount of sequence divergence separating most species, internal barcode regions as short as 100 or 200 bp retain much discriminatory ability (Hajibabaei *et al.*, 2006*b*). Such primers are now being developed (M. Hajibabaei, in prep.) and their deployment will be enlightening.

COMPLICATING ISSUES FOR FISH BARCODING

Species hybridization and introgression can represent a problem for barcoding. Because the barcode gene, COI, is mitochondrial, it typically derives from the maternal parent of any barcoded specimen. Thus, any hybrid fish will be inevitably and wrongly diagnosed as its maternal species. If the identified species is one known to hybridize [see lists provided by Gardner (1997) and Scribner *et al.* (2000)], and an accurate diagnosis is vital, then DNA barcoding should be accompanied by the analysis of known species-specific nuclear DNA alleles to elucidate its status. If the specimen is shown to be a hybrid, the COI barcoding result will enable the female and, by elimination, the male parent to be identified.

Instances of hybridization have been recorded among a wide variety of teleosts, both freshwater (139 species pairs; Scribner *et al.*, 2000) and marine (58 species pairs; Gardner, 1997), but not among chondrichthyans (Gardner, 1997). This latter observation likely reflects the fact that all chondrichthyans employ internal fertilization rather than the external fertilization typical of most teleosts. However, even among teleosts, only a small percentage of species (*c.* 1%) is known to hybridize. Results to date indicate that hybridization is so uncommon that it rarely compromises the ability of barcoding to generate an accurate identification. As mentioned earlier (Ward *et al.*, 2005), extremely few cases of the haplotype sharing among Australasian marine fishes have been found that would result from hybridization and introgression. Three instances of haplotype sharing (excluding the two urolophids that could not separate by barcoding) were encountered, but two of these cases involved sharks and probably represent erroneous identification. The third example involved a reef teleost, the coral trout [*Plectropomus leopardus* (Lacépède)], whose barcode fell within a cluster of *Plectropomus maculatus* (Bloch) barcodes rather than with the only other *P. leopardus* sequence (Ward *et al.*, 2005). Hybridization has been detected in some coral reef fish families (Yaakub *et al.*, 2006), but unfortunately, this misplaced specimen was not vouchered: nuclear DNA analysis would reveal if it was a hybrid or a misidentification. Barcoding results indicated possible introgressive hybridization between two Canadian freshwater

fishes, the darters *Etheostoma nigrum* Rafinesque and *Etheostoma olmstedi* Storer (Hubert *et al.*, 2008).

Hybridization and introgression are certainly a concern for a small number of closely allied teleosts and will lead to misidentifications unless barcoding is accompanied by nuclear DNA analysis, but it is not common enough to have much effect on the ability of barcoding to deliver accurate species diagnoses in the vast majority of cases. Indeed, accumulating FISH-BOL data suggest that initial specimen misidentification appears to be of considerably more concern than complications caused by hybridization.

The inability of barcoding to discriminate very recently radiated species also merits consideration. Genetic divergence of reproductively isolated lineages increases with time, although at a progressively slower rate as genes becomes saturated with mutations. Species of the genus *Thunnus* show little overall divergence (mean congeneric distance of eight species of the 1·11%; Ward *et al.*, 2005), and Pacific bluefin tuna *Thunnus orientalis* (Temminck & Schlegel) and albacore tuna *Thunnus alalunga* (Bonnaterre) only diverge by 0·15% (a single base pair). Interestingly, despite the low divergences between these two species, this single diagnostic base pair permits robust species separation. However, some recently radiated species may simply lack diagnostic mutations in the COI barcode region.

There are undoubtedly species pairs that cannot be separated by barcoding. The sting rays *U. sufflavus* and *U. cruciatus* are one such pair (Ward & Holmes, 2007), and there are some pairs of freshwater species which share barcode haplotypes (Hubert *et al.*, 2008). However, such cases of incomplete resolution are uncommon and invariably involve sister taxa that show limited morphological divergence. Possibly some reflect inadequate taxonomy. The COI sequence from an unknown specimen of a sibling species pair with barcode identity or barcode sharing will return a match with both species when queried using the BOLD search engine. When such cases require accurate resolution, sequence analysis of the more rapidly evolving mtDNA control region or examination of other markers such as microsatellites may offer a solution, assuming the taxa are in fact 'good' species.

While most marine fish species have been reproductively isolated for a long period of time, and now possess diagnostic barcodes, the situation may be different in other settings. For example, the cichlid fishes in the Lake Victoria catchment are thought to have diversified to a species flock of 500 species within the last 400 000 years. Although COI sequences can separate cichlid species on a broader scale (Sparks, 2004; Chakrabarty, 2006), they are not effective in separating all African Great Lake cichlids (analyses of GenBank data not presented). A faster evolving gene will likely be required in such situations, but such cases appear to be rare.

It has been suggested that NUMTs (inserts of mtDNA sequences into the nuclear genome, where they are released from selection and undergo mutational changes as pseudogenes) could provoke misidentifications (Thalmann *et al.*, 2004; Sword *et al.*, 2007). Most NUMTs are <200 bp in length (Richly & Leister, 2004), meaning that they would be unlikely to amplify with the regular barcode primers that target a 650 bp region. Furthermore, the relaxation of selection in NUMTS means that they can often be recognized by the mutations

producing premature stop codons or radical amino acid changes. BOLD automatically flags any barcode sequence that contains a stop codon and the electropherogram trace files of these specimens can then be re-examined. The few instances of such flagged sequences in fish reflected errors in transcribing the electropherograms to a finished nucleotide sequence 'contig assembly' and have been corrected. No NUMTs have been detected in >4000 barcodes from *c.* 900 species of Australasian fishes, which is consistent with their known rarity in fishes (Bensasson *et al.*, 2001). Earlier reports (Richly & Leister, 2004; Antunes & Ramos, 2005) of NUMTs in the fugu *Takifugu rubripes* (Temminck & Schlegel) have been discredited, and those for two other teleosts are also regarded as questionable (Venkatesh *et al.*, 2006). Therefore, NUMTs are not a significant issue for fish barcoding.

Perhaps, the major issue that barcoding, as a taxonomic tool, needs to address concerns the ongoing development of the database. In particular, reference specimens, those that form the 'look-up table' of the database, have to be accurately identified. Historic problems with the accuracy of taxonomic identifications continue to provoke 'error cascades' in the biological sciences (Bortolus, 2008), and these problems have resulted in the call for a 'taxonomic affidavit' to accompany species identification in published research (Dov Por, 2007). Both BOLD and GenBank include structured data fields for recording the identity of the individual responsible for making specimen identifications. However, inclusion of barcodes from misidentified specimens during an ongoing research programme is, unfortunately, inevitable. The BOLD search engine, in its search of 'All Available Barcode Records', searches through all barcode records, but many of these belong to projects for which validation of identifications is in progress. The 'Reference Barcode Database' in BOLD seeks to eliminate dubious records by taking a conservative approach, including only those species represented by three or more individuals showing <2% divergence. While this eliminates many errors, it also eliminates otherwise good records that are contaminated by one or more erroneously identified specimens together with those species that may truly show extensive intraspecies divergence. It also excludes species represented by only one or two barcodes, even though most of these records are valid. Investigators are encouraged to sequester their own suspect sequences from the BOLD search engine pending verification or correction. Where vouchers are held, suspect identifications can be re-examined by a taxonomic expert and errors corrected. Unfortunately, in some instances, whole fish vouchers are not available for re-inspection, and photographs may not be adequate. In these cases, accumulated evidence from multiple specimens might lead to a tentative resolution. Generating a critical mass of BARCODE compliant specimen records and the development of an error-free searchable database remain critically important issues for FISH-BOL to tackle. Ongoing community support for data curation is vital and recruitment of expert ichthyologists is one of the primary objectives of this review.

In conclusion, the organization and optimization of global efforts to generate and share taxonomic knowledge is an urgent issue (Lleonart *et al.*, 2006). DNA barcoding is revealing itself as an immensely valuable tool in such efforts. For fishes, it promises the unambiguous identification of the vast majority of

species including eggs, larvae and processed material. It employs a technology that is both inexpensive and suited to high-throughput procedures. Furthermore, sequencing technologies are advancing rapidly, delivering results more cheaply and more easily. About 5000 fish species have already been barcoded through FISH-BOL, but its goal of barcoding all 30 000 fish species will be achieved only by building broad community participation. While the completion of this mission promises an effective identification system for fishes, it will not be perfect. There will be some taxonomic groups for which COI barcoding cannot provide species-level resolution. However, existing results suggest that these cases will be rare and they are flagged in the search engine. A few instances of apparent barcode sharing between species have been detected; these might be attributable to recent radiation, hybridization or errors in identification or labelling. The barcode records for fishes in BOLD contain some identification errors, but cleansing of such errors is in progress and will be achieved through ongoing curation of the database by members of the taxonomic community. This is a major challenge for FISH-BOL participants, requiring a mechanism for adjudicating discrepancies in the application of names among collections. Because the development of a DNA barcoding system for species identification lies upon a foundation of accurate taxonomic identification of the reference specimens, the success of FISH-BOL will demonstrate the ongoing need for investments in collections and in the broader taxonomic support system.

We thank N. Maclean for inviting submission of this manuscript and to J. Patil and three reviewers for their comments on the original manuscript. This work involved collaboration between the CSIRO Wealth from Oceans Flagship and the Canadian Barcode of Life Network. The latter was supported by Genome Canada (through the Ontario Genomics Institute), NSERC, and other sponsors listed at www.BOLNET.ca.

References

Antunes, A. & Ramos, M. J. (2005). Discovery of a large number of previously unrecognized mitochondrial pseudogenes in fish genomes. *Genomics* **86,** 708–717.

Avise, J. C. (1975). Systematic value of electrophoretic data. *Systematic Zoology* **23,** 465–481.

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. New York, NY: Chapman & Hall.

Bartlett, S. E. & Davidson, W. S. (1991). Identification of *Thunnus* tuna species by the polymerase chain reaction and direct sequence analysis of their mitochondrial cytochrome *b* genes. *Canadian Journal of Fisheries and Aquatic Science* **48,** 309–317.

Bartlett, S. E. & Davidson, W. S. (1992). FINS (forensically important nucleotide sequences): a procedure for identifying the animal origin of biological specimens. *Biotechniques* **12,** 408–411.

Bensasson, D., Zhang, D.-X., Hartl, D. L. & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16,** 314–321.

Bortolus, A. (2008). Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio* **37,** 114–118.

Caddy, J. & Garibaldi, L. (2000). Apparent changes in the trophic composition of wild marine harvest: the perspective from FAO capture database. *Ocean and Coastal Management* **43,** 615–655.

Carew, M. E., Pettigrove, V., Cox, R. L. & Hoffmann, A. A. (2007). DNA identification of urban Tanytarsini chironomids (Diptera: Chironomidae). *Journal of the North American Benthological Society* **26,** 587–600.
Chakrabarty, P. (2006). Systematics and historical biogeography of Greater Antillean Cichlidae. *Molecular Phylogenetics and Evolution* **39,** 619–627.
Chakraborty, A., Sakai, M. & Iwatsuki, Y. (2006). Museum fish specimens and molecular taxonomy: a comparative study on DNA extraction protocols and preservation techniques. *Journal of Applied Ichthyology* **22,** 160–166.
Chantangsi, C., Lynn, D. H., Brandl, M. T., Cole, J. C., Hetrick, N. & Ikonomi, P. (2007). Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. *International Journal of Systematic and Evolutionary Microbiology* **57,** 2412–2425.
Clare, E. L., Lim, B. K., Engstrom, M. D., Eger, J. L. & Hebert, P. D. N. (2007). DNA barcoding of neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes* **7,** 184–190.
Cooper, J. K., Sykes, G., King, S., Cottrill, K., Ivanova, N. V., Hanner, R. & Ikonomi, P. (2007). Species identification in cell culture: a two-pronged molecular approach. *In Vitro Cellular & Developmental Biology. Animal* **43,** 344–351.
Costa, F. O. & Carvalho, G. R. (2007). The barcode of life initiative: synopsis and prospective societal impacts of DNA barcoding of fish. *Genomics, Society and Policy* **3,** 29–40.
Costa, F. O., deWaard, J. R., Boutillier, J., Ratnasingham, S., Dooh, R. T., Hajibabaei, M. & Hebert, P. D. N. (2007). Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences* **64,** 272–295.
Cywinska, A., Hunter, F. F. & Hebert, P. D. N. (2006). Identifying Canadian mosquito species through DNA barcodes. *Medical and Veterinary Entomology* **20,** 413–424.
Dawnay, N., Ogden, R., McEwing, R., Carvalho, G. R. & Thorpe, R. S. (2007). Validation of the barcoding gene CO1 for use in forensic genetic species identification. *Forensic Science International* **173,** 1–6.
Deagle, B. E., Jarman, S. N., Pemberton, D. & Gales, N. J. (2005). Genetic screening for prey in the gut contents from a giant squid (*Architeuthis* sp.). *Journal of Heredity* **96,** 417–423.
Dov Por, F. (2007). A ''taxonomic affidavit'': why it is needed? *Integrative Zoology* **2,** 57–59.
Dung, V. V., Giao, P. M., Chinh, N. N., Tuoc, D., Arctander, P. & MacKinnon, J. (1993). A new species of living bovid from Vietnam. *Nature* **363,** 443–445.
Ekrem, T., Willassen, E. & Stur, E. (2007). A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution* **43,** 530–542.
Elias, M., Hill, R. I., Willmott, K. R., Dasmahapatra, K. K., Brower, A. V. Z., Mallet, J. & Jiggins, C. D. (2007). Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B* **274,** 2881–2889.
Erpenbeck, D., Hooper, J. N. A. & Worheide, G. (2006). COI phylogenies in diploblasts and the ''Barcoding of Life' – are we sequencing a suboptimal partition? *Molecular Ecology Notes* **6,** 550–553.
Eschmeyer, W. N. (2003). *Introduction to the Series Annotated Checklists of Fishes*. San Francisco, CA: California Academy of Sciences.
Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **5,** 294–299.
France, S. C. & Hoover, I. L. (2002). DNA sequences of the mitochondrial COI gene have low levels of divergence among deep-sea octocorals (Cnidaria: Anthozoa). *Hydrobiology* **471,** 149–155.
Frézal, L. & Leblois, R. (2008). Four years of DNA barcoding: current advances and prospects. *Infection, Genetics and Evolution* **5,** 727–736.

Gardner, J. P. A. (1997). Hybridization in the sea. In *Advances in Marine Biology*, Vol. 31 (Blaxter, J. H. S. & Southward, A. J., eds), pp. 2–78. New York, NY: Academic Press.

Hajibabaei, M., deWaard, J. R., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M. & Hebert, P. D. N. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B* **360,** 1959–1967.

Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. & Hebert, P. D. N. (2006*a*). DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Science of the United States of America* **103,** 968–971.

Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B. & Hebert, P. D. N. (2006*b*). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* **6,** 959–964.

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

Hanner, R. & Gregory, T. R. (2007). Genomic diversity research and the role of biorepositories. *Cell Preservation Technology* **5,** 93–103.

Hanner, R., Corthals, A. & Dessauer, H. C. (2005). Salvage of genetically valuable tissues following a freezer failure. *Molecular Phylogenetics and Evolution* **34,** 452–455.

Harris, J. D. (2003). Can you bank on GenBank? *Trends in Ecology and Evolution* **18,** 317–319.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & de Waard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B* **270,** 313–322.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. (2004*a*). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Science of the United States of America* **101,** 14812–14817.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. (2004*b*). Identification of birds through DNA barcodes. *PloS Biology* **2,** 1657–1663.

Hickerson, M. J., Meyer, C. P. & Moritz, C. (2006). DNA barcoding will often fail to discover new animal species over broad parameter space. *Systematic Biology* **55,** 729–739.

Hogg, I. D. & Hebert, P. D. N. (2004). Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Canadian Journal of Zoology* **82,** 749–754.

Hubert, N., Hanner, R., Holm, E., Mandrak, N. E., Taylor, E., Burridge, M., Watkinson, D., Dumont, P., Curry, A., Bentzen, P., Zhang, J., April, J. & Bernatchez, L. (2008). Identifying Canadian freshwater fishes through DNA barcodes. *Plos One* **3,** e2490. doi: 10.1371/journal.pone.0002490

Ivanova, N. V., deWaard, J. R. & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* **6,** 998–1002.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H. & Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* **7,** 544–548.

Jensen, A. C. (1966). Life history of the spiny dogfish. *Fishery Bulletin* **65,** 527–554.

Kelly, R. P., Sarkar, I. N., Eernisse, D. J. & DeSalle, R. (2007). DNA barcoding using chitons (genus *Mopalia*). *Molecular Ecology Notes* **7,** 177–183.

Kimura, M. (1980). A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16,** 111–120.

Klanten, S. O., van Herwerden, L. & Choat, J. H. (2003). Acquiring reef fish DNA sequences from formalin-fixed museum specimens. *Bulletin of Marine Science* **73,** 771–776.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X. & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in animals:

amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 6196–6200.
Kumar, N. P., Rajavel, A. R., Natarajan, R. & Jambulingam, P. (2007). DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology* **44,** 1–7.
Last, P. R., White, W. T. & Pogonoski, J. J. (2007*a*). *Descriptions of New Dogfishes of the Genus Squalus (Squaloidea: Squalidae). Marine and Atmospheric Research Paper* **014**. Hobart: CSIRO.
Last, P. R., Gledhill, D. C. & Holmes, B. H. (2007*b*). A new handfish, *Brachionichthys australis* sp. nov, (Lophiiformes: Brachionichthyidae), with a redescription of the critically endangered spotted handfish, *B. hirsutus* (Lacépède). *Zootaxa* **1666,** 53–68.
Lefebure, T., Douady, C. J., Gouy, M. & Gibiert, J. (2006). Relationship between morphological taxonomy and molecular divergence within Crustacea: proposal of a molecular threshold to help species delimitation. *Molecular Phylogenetics and Evolution* **40,** 435–447.
Leviton, A. E., Gibbs, R. H., Heale, E. & Dawson, C. E. (1985). Standards in herpetology and ichthyology.1. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. *Copeia* **1985,** 802–832.
Lleonart, J., Taconet, M. & Lamboeuf, M. (2006). Integrating information on marine species identification for fishery purposes. *Marine Ecology Progress Series* **316,** 231–238.
Lorenz, J. G., Jackson, W. E., Beck, J. C. & Hanner, R. (2005). The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society B* **360,** 1869–1877.
Manwell, C. & Baker, C. M. A. (1963). A sibling species of sea-cucumber discovered by starch-gel electrophoresis. *Comparative Biochemistry and Physiology* **10,** 39–53.
Marko, P. B., Lee, S. C., Rice, A. M., Gramling, J. M., Fitzhenry, T. M., McAlister, J. S., Harper, G. R. & Moran, A. L. (2004). Mislabelling of a depleted reef fish. *Nature* **430,** 309–310.
Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology* **55,** 715–728.
Meyer, C. P. & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLosS Biology* **3,** 2229–2238.
Mitchell, A. (2008). DNA barcoding demystified. *Australian Journal of Entomology* **47,** 169–173.
Monk, R. R. & Baker, R. J. (2001). e-Vouchers and the use of digital imagery in natural history collections. *Museology, Museum of Texas Tech University* **10,** 1–8.
Moritz, C. & Cicero, C. (2004). DNA barcoding: promise and pitfalls. *PLoS Biology* **2,** 1529–1531.
Murray, B. W., McClymont, R. A. & Strobeck, C. (1995). Forensic identification of ungulate species using restriction digests of PCR-amplified mitochondrial DNA. *Journal of Forensic Sciences* **40,** 943–951.
Neigel, J., Domingo, A. & Stake, J. (2007). DNA barcoding as a tool for coral reef conservation. *Coral Reefs* **26,** 487–499.
Padial, J. M. & de la Riva, I. (2007). Integrative taxonomists should use and produce DNA barcodes. *Zootaxa* **1586,** 67–68.
Paine, M. A., McDowell, J. R. & Graves, J. E. (2007). Specific identification of western Atlantic Ocean scombrids using mitochondrial DNA cytochrome *c* oxidase subunit I (COI) gene region sequences. *Bulletin of Marine Science* **80,** 353–367.
Palumbi, S. R. & Cipriano, F. (1998). Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. *Journal of Heredity* **89,** 459–464.
Parsons, K. M., Piertney, S. B., Middlemas, S. J., Hammond, P. S. & Armstrong, J. D. (2005). DNA-based identification of salmonid prey species in seal faeces. *Journal of Zoology* **266,** 275–281.

Pegg, G. G., Sinclair, B., Briskey, L. & Aspden, W. J. (2006). MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Scientia Marina* **70** (Suppl. 2), 7–12.

Pleijel, F., Jondelius, U., Norlinder, E., Nygren, A., Oxelman, B., Schander, C., Sundberg, P. & Thollesson, M. (2008). Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Molecular Phylogenetics and Evolution* **48**, 369–371. doi: 10.1016/j.ympev.2008.03.024

Pyle, R. L., Earle, J. L. & Greene, B. D. (2008). Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific. *Zootaxa* **1671,** 3–31.

Rach, J., DeSalle, R., Sarkar, I. N., Schierwater, B. & Hadrys, H. (2008). Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. *Proceedings of the Royal Society B* **275,** 237–247.

Ratnasingham, S. & Hebert, P. D. N. (2007). BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* **7,** 355–364.

Richly, E. & Leister, D. (2004). NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21,** 1081–1084.

Rock, J., Costa, F. O., Walker, D. I., North, A. W., Hutchinson, W. F. & Carvalho, G. R. (2008). DNA barcodes of fish of the Antarctic Scotia Sea indicate priority groups for taxonomic and systematics focus. *Antarctic Science* **20,** 253–262.

Roe, A. D. & Sperling, F. A. H. (2007). Patterns of evolution of mitochondrial cytochrome *c* oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution* **44,** 325–345.

Ross, H. A., Murugan, S. & Si, W. L. S. (2008). Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology* **57,** 216–230.

Rubinoff, D., Cameron, S. & Will, K. (2006). A genomic perspective on the shortcomings of mitochondrial DNA for ''barcoding'' identification. *Journal of Heredity* **97,** 581–594.

Ruedas, L. A., Salazar-Bravo, J., Dragoo, J. W. & Yates, T. L. (2000). The importance of being earnest: what, if anything, constitutes a ''Specimen Examined?'' *Molecular Phylogenetics and Evolution* **17,** 129–132.

Saunders, G. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society B* **360,** 1879–1888.

Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K. & Lan, R. (2005). Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B* **360,** 1805–1811.

Scribner, K. T., Page, K. S. & Bartron, M. L. (2000). Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries* **10,** 293–323.

Seifert, K. A., Samson, R. A., deWaard, J. R., Houbraken, J., Levesque, C. A., Moncalvo, J. M., Louis-Seize, G. & Hebert, P. D. N. (2007). Prospects for fungus identification using CO1 DNA barcodes, with *Penicillium* as a test case. *Proceedings of the National Academy of Science of the United States of America* **104,** 3901–3906.

Shearer, T. L., Van Oppen, M. J. H., Romano, S. L. & Worheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology* **11,** 2475–2487.

Shivji, M., Clarke, S., Pank, M., Natanson, L., Kohler, N. & Stanhope, M. (2002). Genetic identification of pelagic shark body parts for conservation and trade monitoring. *Conservation Biology* **16,** 1036–1047.

Sigler, M. F., Hulbert, L. B., Lunsford, C. R., Thompson, N. H., Burek, K., O'Corry-Crowe, G. & Hirons, A. C. (2006). Diet of Pacific sleeper shark, a potential Steller sea lion predator, in the north-east Pacific Ocean. *Journal of Fish Biology* **69,** 392–405.

Smith, M. A., Woodely, N. E., Janzen, D. H., Hallwachs, W. & Hebert, P. D. N. (2006). DNA barcoding reveal cryptic host-specificity within the presumed polyphagous

members of a genus of parasitoid flies (Diptera, Tachinidae). *Proceedings of the National Academy of Science of the United States of America* **103,** 3657–3662.
Smith, P. J., McVeagh, S. M. & Steinke, D. (2008). DNA barcoding for the identification of smoked fish products. *Journal of Fish Biology* **72,** 464–471.
Sparks, J. S. (2004). Molecular phylogeny and biogeography of the Malagasy and South Asian cichlids (Teleostei: Perciformes: Cichlidae). *Molecular Phylogenetics and Evolution* **30,** 599–614.
Spies, I. B., Gaichas, S., Stevenson, D. E., Orr, J. W. & Canino, M. F. (2006). DNA-based identification of Alaska skates (*Amblyraja*, *Bathyraja* and *Raja*: Rajidae) using cytochrome c oxidase subunit I (COI) variation. *Journal of Fish Biology* **69** (Suppl. B), 283–292.
Sword, G. A., Senior, L. B., Gaskin, J. F. & Joern, A. (2007). Double trouble for grasshopper molecular systematics: intra-individual heterogeneity of both mitochondrial 12S-valine-16S and nuclear internal transcribed spacer ribosomal DNA sequences in *Hesperotettix viridis* (Orthoptera; Acrididae). *Systematic Entomology* **32,** 420–428.
Thalmann, O., Hebler, J., Poinar, N., Pääbo, S. & Vigilant, L. (2004). Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Molecular Ecology* **13,** 321–335.
Venkatesh, B., Dandona, N. & Brenner, S. (2006). Fugu genome does not contain mitochondrial pseudogenes. *Genomics* **87,** 307–310.
Victor, B. C. (2007). *Coryphopterus kuna*, a new goby (Perciformes: Gobiidae: Gobinae) from the western Caribbean, with the identification of the late larval stage and an estimate of the pelagic larval duration. *Zootaxa* **1526,** 51–61.
Victor, B. C. (2008). Redescription of *Coryphopterus tortugae* (Jordan) and a new allied species *Coryphopterus bol* (Perciformes: Gobiidae: Gobiinae) from the tropical western Atlantic Ocean. *Journal of the Ocean Science Foundation* **1,** 1–19.
Walters, C. & Hanner, R. (2006). Platforms for DNA Banking. In *DNA Banks – Providing Novel Options for Gene Banks? Topical Reviews in Agricultural Biodiversity* (De Vicente, M. C. & Andersson, M. S., eds), pp. 25–35. Rome: International Plant Genetic Resources Institute.
Ward, R. D. & Holmes, B. H. (2007). An analysis of nucleotide and amino acid variability in the barcode region of cytochrome *c* oxidase I (*cox1*) in fishes. *Molecular Ecology Notes* **7,** 899–907.
Ward, R. D., Woodwark, M. & Skibinski, D. O. F. (1994). A comparison of genetic diversity levels in marine, freshwater and anadromous fish. *Journal of Fish Biology* **44,** 213–232.
Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. & Hebert, P. D. N. (2005). Barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B* **360,** 1847–1857.
Ward, R. D., Holmes, B. H., Zemlak, T. S. & Smith, P. J. (2007). DNA barcoding discriminates spurdogs of the genus *Squalus*. In *Descriptions of New Dogfishes of the Genus Squalus (Squaloidea: Squalidae)* (Last, P. R., White, W. T. & Pogonoski, J. J., eds), pp. 117–130. *Marine and Atmospheric Research Paper* **014**. Hobart: CSIRO.
Ward, R. D., Holmes, B. H. & Yearsley, G. K. (2008*a*). DNA barcoding reveals a likely second species of Asian seabass (barramundi) (*Lates calcarifer*). *Journal of Fish Biology* **72,** 458–463.
Ward, R. D., Holmes, B. H., White, W. T. & Last, P. R. (2008*b*). DNA barcoding Australian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research* **59,** 57–71.
Ward, R. D., Costa, F. O., Holmes, B. H. & Steinke, D. (2008*c*). DNA barcoding shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa but a likely two species for both *Zeus faber* (John dory) and *Lepidopus caudatus* (silver scabbardfish). *Aquatic Biology* **3,** 71–78.
Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays* **29,** 188–197.

Whitworth, T. L., Dawson, R. D., Magalon, H. & Baudry, E. (2007). DNA barcoding cannot reliably identify species of the blowfly genus *Protocalliphora* (Diptera: Calliphoridae). *Proceedings of the Royal Society B* **274,** 1731–1739.
Wong, E. & Hanner, R. H. (2008). DNA barcoding detects market substitution in North American seafood. *Food Research International* **41,** 828–837.
Yaakub, S. M., Bellwood, D. R., van Herwerden, L. & Walsh, F. M. (2006). Hybridization in coral reef fishes: introgression and bi-directional gene exchange in *Thalassoma* (family Labridae). *Molecular Phylogenetics and Evolution* **40,** 84–100.
Yancy, H. F., Zemlak, T. S., Mason, J. A., Washington, J. D., Tenge, B. J., Nguyen, N.-L.N. T., Barnett, J. D., Savary, W. E., Hill, W. E., Moore, M. M., Fry, F. S., Randolph, S. C., Rogers, P. L. & Hebert, P. D. N. (2007). The potential use of DNA barcodes in regulatory science: applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection* **70,** 210–217.
Yearsley, G. K. & Last, P. R. (2006). *Urolophus kapalensis* sp nov., a new stingaree (Myliobatiformes: Urolophidae) off eastern Australia. *Zootaxa* **1176,** 41–52.
Yoo, H. S., Eah, J. Y., Kim, J. S., Kim, Y. J., Min, M. S., Paek, W. K., Lee, H. & Kim, C. B. (2006). DNA barcoding Korean birds. *Molecules and Cells* **22,** 323–327.
Zhang, A. B., Sikes, D. S., Muster, C. & Li, S. Q. (2008). Inferring species membership using DNA sequences with back-propagation neural networks. *Systematic Biology* **57,** 202–215.