PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B
FirstCite® e-publishing

Phil. Trans. R. Soc. B
doi:10.1098/rstb.2005.1716
Published online

DNA barcoding Australia's fish species

Robert D. Ward[1,*], Tyler S. Zemlak[2], Bronwyn H. Innes[1], Peter R. Last[1] and Paul D. N. Hebert[2]

[1]*CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia*
[2]*Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2WI*

Two hundred and seven species of fish, mostly Australian marine fish, were sequenced (barcoded) for a 655 bp region of the mitochondrial cytochrome oxidase subunit I gene (*cox1*). Most species were represented by multiple specimens, and 754 sequences were generated. The GC content of the 143 species of teleosts was higher than the 61 species of sharks and rays (47.1% versus 42.2%), largely due to a higher GC content of codon position 3 in the former (41.1% versus 29.9%). Rays had higher GC than sharks (44.7% versus 41.0%), again largely due to higher GC in the 3rd codon position in the former (36.3% versus 26.8%). Average within-species, genus, family, order and class Kimura two parameter (K2P) distances were 0.39%, 9.93%, 15.46%, 22.18% and 23.27%, respectively. All species could be differentiated by their *cox1* sequence, although single individuals of each of two species had haplotypes characteristic of a congener. Although DNA barcoding aims to develop species identification systems, some phylogenetic signal was apparent in the data. In the neighbour-joining tree for all 754 sequences, four major clusters were apparent: chimaerids, rays, sharks and teleosts. Species within genera invariably clustered, and generally so did genera within families. Three taxonomic groups—dogfishes of the genus *Squalus*, flatheads of the family Platycephalidae, and tunas of the genus *Thunnus*—were examined more closely. The clades revealed after bootstrapping generally corresponded well with expectations. Individuals from operational taxonomic units designated as *Squalus* species B through F formed individual clades, supporting morphological evidence for each of these being separate species. We conclude that *cox1* sequencing, or 'barcoding', can be used to identify fish species.

Keywords: cytochrome oxidase subunit I; *CO1*; fish identification; sharks; rays; teleosts

1. INTRODUCTION

It has long been recognized that DNA sequence diversity, whether assessed directly or indirectly through protein analysis, can be used to discriminate species. More than 40 years ago, starch gel electrophoresis of proteins was first used to identify species (Manwell & Baker 1963). Nearly 30 years ago, single gene sequence analysis of ribosomal DNA was being used to investigate evolutionary relationships at a high level (Woese & Fox 1977), and mitochondrial DNA approaches dominated molecular systematics in the late 1970s and 1980s (Avise 1994).

Recently, Tautz *et al.* (2002, 2003) made the case for a DNA-based taxonomic system. DNA sequence analysis has been used for 30 years to assist species identifications, but different sequences have been used for different taxonomic groups and in different laboratories. Hebert *et al.* (2003) proposed that a single gene sequence would be sufficient to differentiate all, or at least the vast majority of, animal species, and proposed the use of the mitochondrial DNA gene cytochrome oxidase subunit I (*cox1*) as a global bioidentification system for animals. The sequence was likened to a barcode, with species being delineated by a particular sequence or by a tight cluster of very similar sequences.

Empirical support for the barcoding concept ranges from studies of invertebrates (e.g. springtails and butterflies) to birds (Hebert *et al.* 2004*a*,*b*; Hogg & Hebert 2004). However, the approach is not without controversy (e.g. Lipscomb *et al.* 2003; Moritz & Cicero 2004). For a barcoding approach to species identification to succeed, within-species DNA sequences need to be more similar to one another than to sequences in different species. Recent studies show that this is generally the case, but there are exceptions. Hybridization among species would create taxonomic uncertainty: mitochondrial DNA is maternally inherited and any hybrid or subsequent generation would have the maternal species DNA only.

Here we examine whether barcoding can be used to discriminate fish species. There are probably close to 30 000 fish species worldwide (the FishBase count of species on March 31 2005 was 28 800—see www.fishbase.org), constituting about 50% of all vertebrate species. They are systematically very diverse, ranging from ancient jawless species (Agnatha: hagfish and lampreys) through to cartilaginous fishes (Chondrichthyes: chimaeras, sharks and rays) and to old and modern bony fish (Osteichthyes: coelacanths, eels, carps, tunas, flatfishes, salmonids, seahorses, etc.). In 2000, fisheries provided more than 15% of total animal protein to the global food supply, employed about

* Author for correspondence (Bob.Ward@csiro.au).

One contribution of 18 to a Theme Issue 'DNA barcoding of life'.

35 million people, and had an estimated first sale value of about US$81 billion (FAO 2002): fish and fish products are important contributors to human food security.

Accurate and unambiguous identification of fish and fish products, from eggs to adults, is important in many areas. It would enable retail substitutions of species to be detected, assist in managing fisheries for long-term sustainability, and improve ecosystem research and conservation. Hitherto, a wide variety of protein- and DNA-based methods have been used for the genetic identification of fish species (see, for example, Ward & Grewe 1994; Pérez-Martin & Sotelo 2003). Here we examine *cox1* diversity within and among 207 fish species, most of which have been examined from multiple specimens, with the goal of determining whether DNA barcoding can achieve unambiguous species recognition in fish. Most of the species examined were Australian commercial species, but many more such species and bycatch species remain to be barcoded.

2. MATERIAL AND METHODS

Tissue subsamples were isolated from fragments of white muscle of (mostly) Australian fish species that had been stored at −80 °C for several years. The majority of these samples had been initially collected for the protein fingerprinting of Australia's commercial domestic and imported fish species (Yearsley *et al.* 1999, 2003), others were from species of special interest to scientists at the Commonwealth Scientific and Industrial Research Organization's division of Marine and Atmospheric Research. Some of the samples were from voucher specimens. Generally we aimed, where possible, to sample five individuals per species (four species were sampled at greater intensity) and achieved this target for 47.5% of the 207 species sequenced. Numbers per species ranged from one to 15 with a mean of 3.66.

DNA extracts were prepared from muscle tissue using Chelex dry release (Hajibabaei *et al.* 2005). Approximately 655 bp were amplified from the 5′ region of the *cox1* gene from mitochondrial DNA using different combinations of four newly designed primers:

FishF1-5′TCAACCAACCACAAAGACATTGGCAC3′,
FishF2-5′TCGACTAATCATAAAGATATCGGCAC3′,
FishR1-5′TAGACTTCTGGGTGGCCAAAGAATCA3′,
FishR2-5′ACTTCAGGGTGACCGAAGAATCAGAA3′.

In one case, the broadnose shark (*Notorhynchus cepedianus*), an internal forward primer (5′ATCTTTGGTGCATGAGCAGGAATAGT3′) was used in conjunction with FishR2 to yield a shorter fragment (616 bp). The 25 µl PCR reaction mixes included 18.75 µl of ultrapure water, 2.25 µl of 10× PCR buffer, 1.25 µl of $MgCl_2$ (50 mM), 0.25 µl of each primer (0.01 mM), 0.125 µl of each dNTP (0.05 mM), 0.625 U of *Taq* polymerase, and 0.5–2.0 µl of DNA template. Amplifications were performed using a Mastercycler® Eppendorf gradient thermal cycler (Brinkmann Instruments, Inc.). The thermal regime consisted of an initial step of 2 min at 95 °C followed by 35 cycles of 0.5 min at 94 °C, 0.5 min at 54 °C, and 1 min at 72 °C, followed in turn by 10 min at 72 °C and then held at 4 °C. PCR products were visualized on 1.2% agarose gels and the most intense products were selected for sequencing. Products were labelled using the BigDye® Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) and sequenced bidirectionally using an ABI 3730 capillary sequencer following manufacturer's instructions. Sequence data, electropherograms, and primer details for specimens are available within the completed project file 'Fishes of Australia Part 1' on the Barcode of Life database site at the University of Guelph (see http://www.barcodinglife.org). GenBank numbers are DQ107581 to DQ108334 and are matched against individual specimens in the 'Fishes of Australia Part 1' file.

Sequences were aligned using SeqScape v.2.5 software (Applied Biosystems, Inc.). Sequence divergences were calculated using the Kimura two parameter (K2P) distance model (Kimura 1980). Neighbour-joining (NJ) trees of K2P distances were created to provide a graphic representation of the patterning of divergence between species (Saitou & Nei 1987). In the three chosen subgroups of fish, bootstrapping was performed in MEGA3 (Kumar *et al.* 2004) with 1000 replications.

Nearly all the target species amplified using one or both primer sets. However, there was insufficient product for sequencing from five species: *N. cepedianus* ($n=5$), *Asymbolus rubiginosus* (orange spotted catshark, $n=2$), *Neoplatycephalus conatus* (deepwater flathead, $n=5$), *Cephalopholis igarashiensis* (goldbar grouper, $n=1$) and *Epinephelus undulatostriatus* (Maori rockcod, $n=2$). The Maori rockcod samples were in a degraded condition, probably explaining their PCR recalcitrance. For *N. cepedianus*, a newly designed internal forward primer was used as described above to obtain product that could be sequenced.

3. RESULTS

We present results for all 207 species followed by more detailed examinations of three subgroups of fish. These subgroups were flatheads (14 species from three genera), large tunas (genus *Thunnus*, eight species), and dogfish or spurdogs of the genus *Squalus* (which contains several Australian morphs that have not been formally described as species). Flatheads and tunas are teleosts; spurdogs are chondrichthyans.

(a) *All species*

A total of 207 species were analysed, giving (because of multiple specimens for most species) a total of 754 sequences. The full K2P/NJ tree has been lodged as an electronic supplementary material. It is presented here in summary form as figure 3. All 207 species can be differentiated by *coxI* barcoding. We did not choose species known to be readily differentiable; indeed, we included several sets of sibling species (for example, in the genera *Squalus* and *Thunnus*) and many species were congeneric.

Read lengths were all about 655 bp long, although in some instances some base calls were uncertain. No insertions, deletions or stop codons were observed in any sequence. The lack of stop codons is consistent with all amplified sequences being functional mitochondrial *cox1* sequences, and that, together with the fact that all amplified sequences were about 655 bp in length, suggests that NUMTs (nuclear DNA sequences originating from mitochondrial DNA sequences) were not sequenced (vertebrate NUMTS are typically smaller than 600 bp; Zhang & Hewitt 1996).

The average K2P distance of individuals within species was 0.39% compared with 9.93% for species within genera (table 1). Overall, therefore, there was *ca* 25× more variation among congeneric species than

Table 1. Summary of genetic divergences (K2P percent) within various taxonomic levels. Data are from 754 sequences from 207 species (173 represented by multiple specimens) and 122 genera (43 represented by multiple species).

comparisons within	number of comparisons	distance minimum	distance mean	distance maximum	s.e.
species	1315	0	0.39	14.08[b]	0.031
genera	4259	0[a]	9.93	20.63	0.096
families	9479	1.39	15.46	35.72	0.049
orders	68083	9.55	22.18	37.52	0.012
classes	83265	14.33	23.27	37.39	0.009

[a] One example in *Pristiophorus* and one in *Plectropomus*, where in each instance one sequence among multiple specimens appeared to be of a different, but congeneric, species (see §3*a*).
[b] *Hydrolagus novazaelandiae* (see §3*a*).

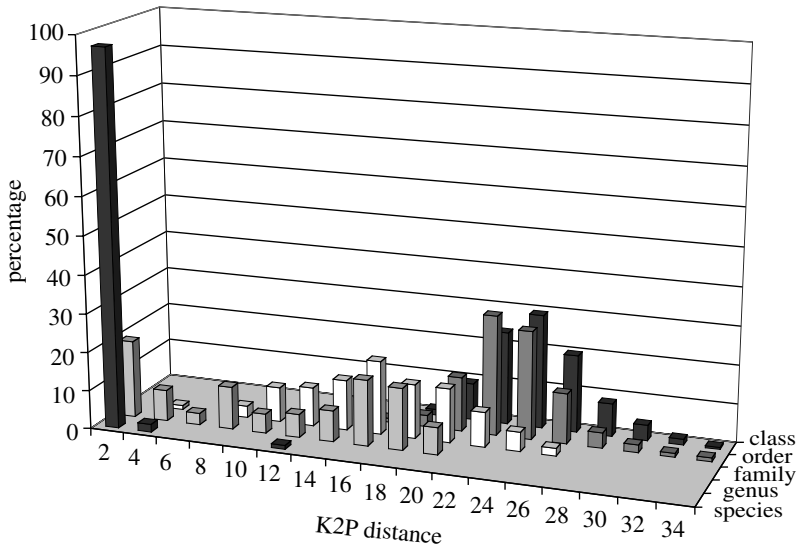

Figure 1. Distribution of K2P distances (percent) for *coxI* within different taxonomic categories. See also table 1. Note that cells with a frequency of less than 1% are not represented.

among conspecific individuals. Mean divergence among species within families increases to 15.5%, and among species within orders and classes it increases to 22.2% and 23.3%, respectively (table 1, figure 1). The rate of increase declines in the higher taxonomic categories due to substitutional saturation.

There was a higher overall GC content in the 143 species of bony fish compared with the 61 species of sharks and rays (47.1±0.2% versus 42.2±0.3%, see also table 2). This difference was attributable (table 3, figure 2) to the GC content of the 1st (57.1% versus 53.5%) and, especially, 3rd codon base (41.1% versus 29.9%). The variance of GC content among species of each of the three groups of fishes (chimaerids, sharks and rays, teleosts) was much higher for the 3rd base (GC_3) than the 1st base (table 3, figure 2); the 2nd base was nearly invariant. The mean GC content of the 20 barcoded ray species was higher than the 41 shark species (44.7±0.4% versus 41.0±0.3%). This was again mostly attributable to GC_3 (36.3±1.3% versus 26.8±0.9%). Ten of the 13 species with GC_3 >35% were rays, while all seven with GC_3 <21% were sharks. This disparity helps to explain the multimodal distribution in GC_3 content in sharks and rays (figure 2).

Originally we had thought that we had two species that were identical with respect to their *cox1* barcodes, *Centrophorus moluccensis* (Endeavour dogfish) and *Centrophorus uyato* (southern dogfish). However, on receipt of the barcode data we re-checked the identification history of these samples and discovered that the two specimens originally classified as *C. uyato* had in fact independently been re-identified by a shark taxonomist as *C. moluccensis*. This experience attests to the absolute necessity of correct species identification in the development of any barcode library. Further, while such a library is being developed, specimens should be retained until fully analysed to facilitate the reappraisal of any potentially spurious results. Voucher collection is a necessity.

Two species of *Hydrolagus*, *Hydrolagus lemures* (blackfin ghostshark) and *Hydrolagus ogilbyi* (Ogilbys ghostshark), were originally regarded as specifically distinct (Last & Stevens 1994) but have been provisionally amalgamated in a world revision of the group (D. Didier, personal communication). In our study, the five samples of *H. lemures* clustered tightly together (mean genetic distance of 0.21%), as did the four samples of *H. ogilbyi* (mean genetic distance of 0.19%), with a mean distance between the two taxa of 6.80%. Barcoding (and earlier protein fingerprinting; Yearsley *et al.* 1999) supports the contention that these are distinct species.

In several instances, we detected deep divergences among individuals that had been assigned to single

Table 2. Mean percentage base composition (with s.e.), comparing *cox1* sequences of chimaerids, sharks and rays, and teleosts. Where multiple individuals were taken for any one species, a single sequence was selected at random.

group	number of species	% of base			
		G	C	A	T
chimaerids	3	16.67±0.30	28.13±1.13	25.21±0.06	29.98±0.94
sharks & rays	61	16.75±0.10	25.44±0.25	25.33±0.15	32.48±0.24
teleosts	143	18.31±0.07	28.75±0.15	23.58±0.09	29.38±0.14

Table 3. GC content of the 1st, 2nd and 3rd codon positions (with s.e.). Where multiple individuals were taken for any one species, a single sequence was selected at random.

group	number of species	GC% and codon position		
		1st	2nd	3rd
chimaerids	3	52.24±0.40	42.33±0.08	34.97±2.40
sharks & rays	61	53.54±0.19	42.71±0.02	29.89±0.90
teleosts	143	57.11±0.10	42.63±0.02	41.05±0.50

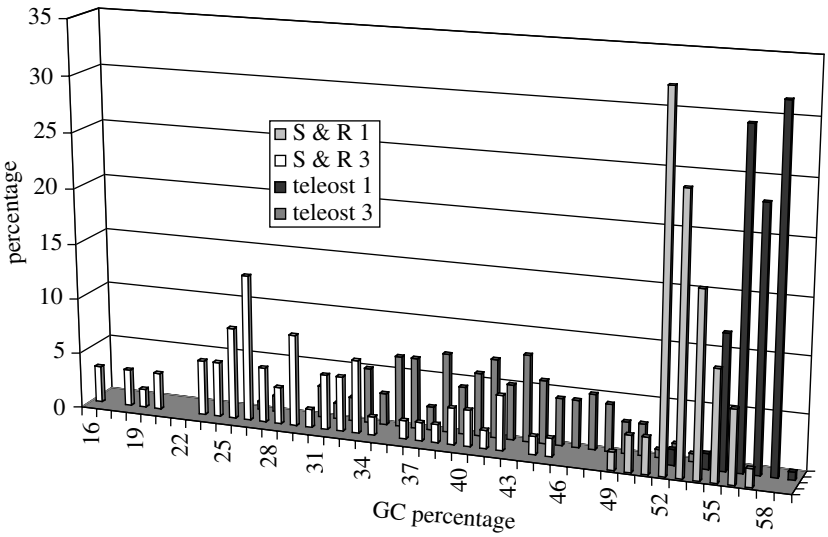

Figure 2. Variation in GC content for *cox1* among two groups of fishes. The first, second, third and fourth rows plot the GC content of sharks and rays codon position 1, sharks and rays codon position 3, teleosts codon position 1 and teleosts codon position 3, respectively. Codon position 2 is not shown as this shows very little variation within and among the two fish groups (table 3).

species. These deep divergences may flag previously unrecognized species and warrant further study. Two such examples (see figure 3) will be briefly described.

Two individuals of *Hydrolagus novaezelandiae* (short-nose chimaera), both collected off the east coast of the South Island of New Zealand, showed a divergence of 14.08%. One of these individuals was only tentatively ascribed to this species. It is likely, in retrospect, that this identification was incorrect or that an undescribed species exists. These two specimens also had distinct protein fingerprints (Yearsley *et al*. 2003). Clearly the incorrectly designated specimen is neither *H. lemures* nor *H. ogilbyi*, as both these species are distinct from both specimens of *H. novaezelandiae*.

Deep divergence was also seen between one specimen of the monotypic genus *Pastinachus* (cowtail stingrays) and two further specimens (within- and between-group divergences of 0.61% and 6.43% respectively). Recent studies have shown that *Parupeneus sephen* is a complex of morphologically distinct species in the Indo-Pacific (Last *et al*. 2005), and the high *cox1* divergence probably reflects the presence of two of these species.

Possible examples of shared haplotypes were seen in the genera *Pristiophorus* and *Plectropomus*. Five samples of each of two species of *Pristiophorus* were sequenced (*Pristiophorus cirratus*, the common sawshark, and *Pristiophorus nudipinnis*, the southern sawshark), but one of the *P. nudipinnis* samples had an identical *cox1* sequence to two of the *P. cirratus* samples. Excluding this aberrant sample, the mean genetic distance within species was 0.35% and the mean genetic distance between species was 10.94%. Similarly, one of two specimens of *Plectropomus leopardus* (common coral trout) clustered very closely with five specimens of *Plectropomus maculatus* (barcheek coral trout; mean sequence divergence of the six samples=0.19%) and away from the remaining *Pl. leopardus* sequence

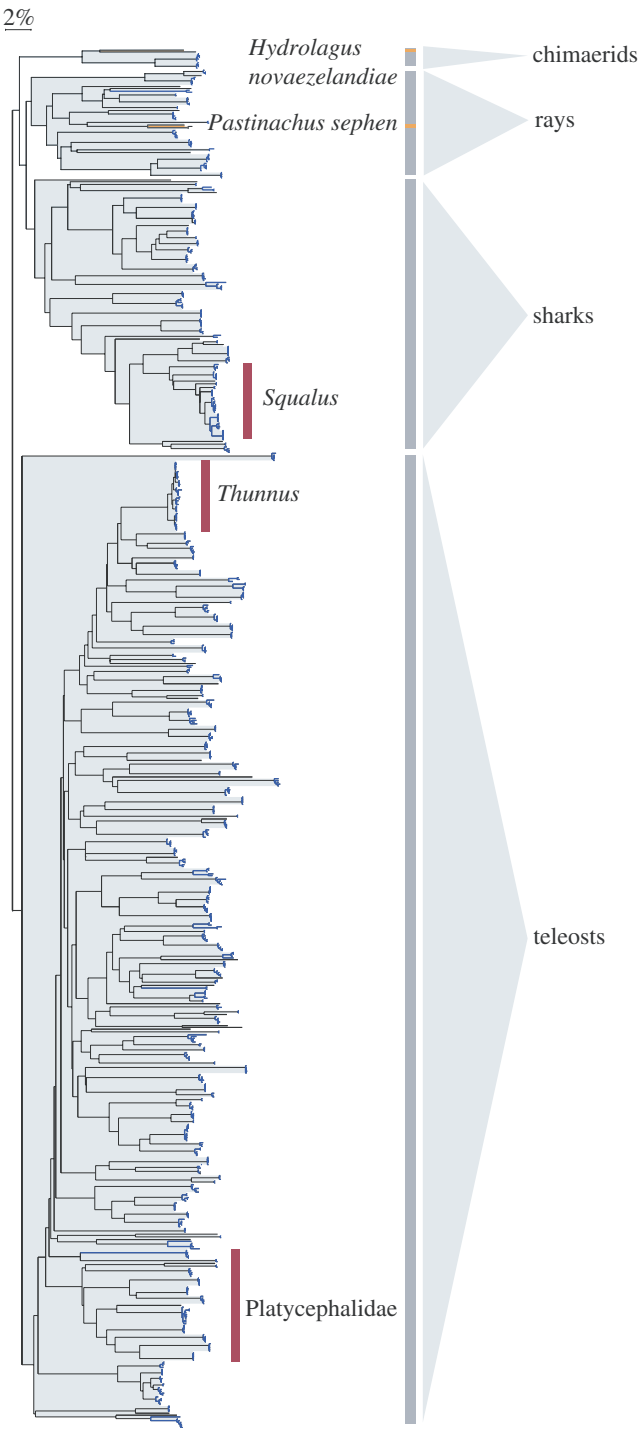

Figure 3. Neighbour-joining tree of 754 *cox1* sequences from 207 fish species, using K2P distances. Multiple specimens of individual species are marked in blue. The three instances of deep intra-specific divergence are identified in orange. The three subgroups examined in more detail are identified.

(divergence = 4.64%). In both genera, the aberrant samples might reflect shared haplotypes (perhaps from past introgressive hybridization between species—in both instances the species have overlapping ranges) or, possibly, misidentification of the original specimens.

(b) *Flatheads, family Platycephalidae*

Flatheads are scorpaeniform fish and members of the family Platycephalidae, of which there are some 60 species worldwide in 18 genera (see http://www.fishbase.org; also see Nelson 1994). We examined 15 mostly commercially significant species from three genera, but one species, *Neoplatycephalus conatus*, failed to amplify. Most of the species were represented by multiple specimens (figure 4). Two genera (*Platycephalus* and *Neoplatycephalus*) are members of the subfamily Platycephalinae; the third, *Cymbacephalus*, belongs to the subfamily Onigociinae (Matsubara & Ochiai 1955). Keenan (1988) suggested the subfamily Cymbacephalinae be revived for a small group of species including one examined here, *Cymbacephalus nematophthalmus*.

All assemblages of conspecific individuals had bootstrap values of 95–100%. K2P nucleotide diversity within species was extremely limited, ranging from 0% to 0.93%, with a mean of 0.22%. Ten individuals of *Platycephalus longispinis* were examined, five each from eastern and western Australia: there was no obvious east–west separation of these individuals in the *P. longispinis* clade.

Divergence between species was high, with average within-genus and within-family K2P distances of 15.55% and 22.07%, respectively. Two distinct clades representing the subfamilies Onigociinae (*Cymbacephalus*, two species) and Platycephalinae (*Platycephalus* and *Neoplatycephalus*, ten and two species respectively) were recognized with a bootstrap value of 99%. Four subclades with bootstrap support values of 90–100% can be seen within the Platycephalinae clade: (i) *P. longispinis*/*P. bassensis*; (ii) *P. caeruleopunctatus*/*P. speculator*; (iii) *P. fuscus*/*P. endrachtensis*/*P. indicus*; (iv) *N. aurimaculatus*/*N. richardsoni*.

(c) *Tunas of the genus* Thunnus

Tunas are large highly migratory fish of the family Scombridae. All are commercially important, some extremely so (FAO 2004). There are eight species in the genus *Thunnus* (we follow Collette *et al.* (2001) in separating *T. thynnus thynnus* and *T. thynnus orientalis* into two species, *T. thynnus* and *T. orientalis*, respectively), six of which are found in Australian waters: *T. obesus* (bigeye tuna), *T. alalunga* (albacore), *T. albacares* (yellowfin tuna), *T. maccoyii* (southern bluefin tuna), *T. orientalis* (northern bluefin tuna) and *T. tonggol* (longtail tuna). We sequenced five individuals of each of these species. We also sequenced four individuals of *T. atlanticus* (Atlantic blackfin tuna) and four individuals of *T. thynnus* (Atlantic bluefin tuna); for the latter, we downloaded data for a further three individuals from GenBank. Finally, we downloaded two additional sequences from GenBank for *T. alalunga* and added three additional sequences that we collected for *T. albacares* from South African waters.

The resulting phenogram (figure 5) of 46 sequences appears very different from the flathead phenogram. Each species clustered as a separate assemblage (no individuals were misplaced), but bootstrap values for species separation were mostly around the 60–70% level. Genetic differences were very small within species, with a mean K2P distance of 0.11%. The three South African individuals of *T. albacares* could not be separated from the five Australian individuals. The mean interspecies distance was very low at 1.04%. Only one clade had a high level of bootstrap support (99%), comprising *T. orientalis* and *T. alalunga*.

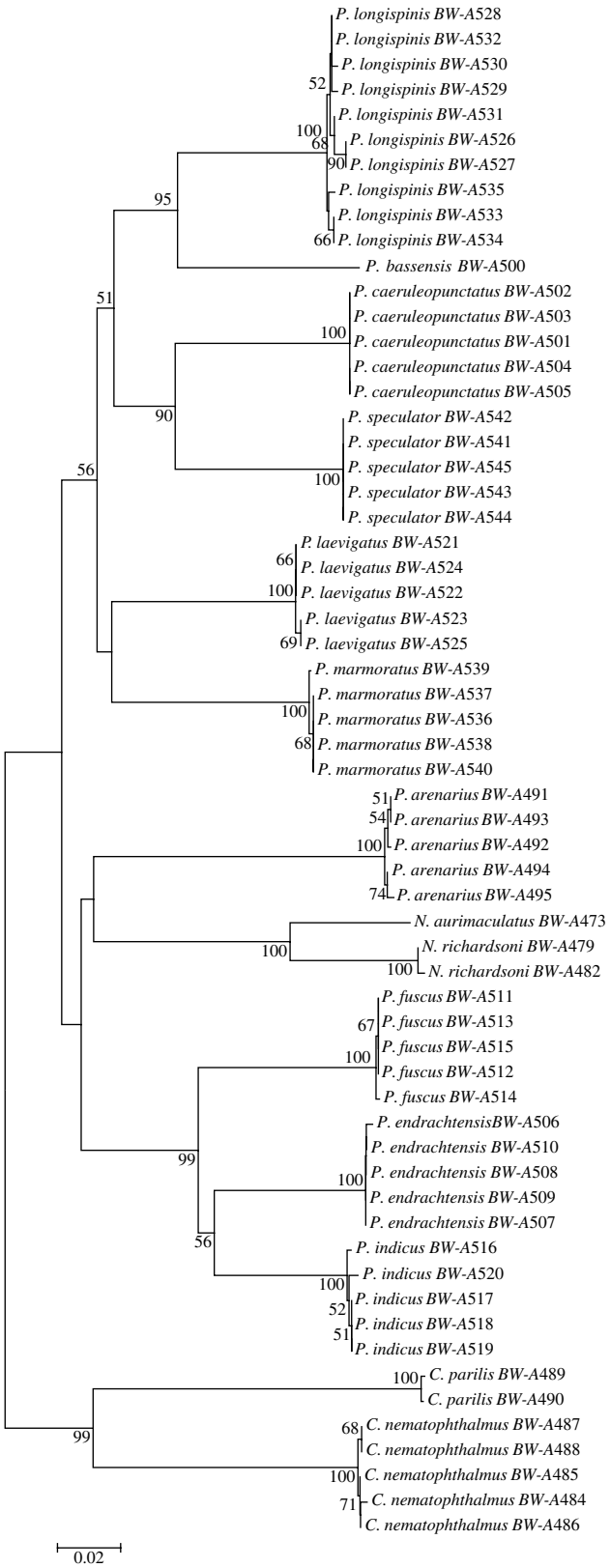

Figure 4. K2P distance neighbour-joining tree of 61 *cox1* sequences from 14 species of flathead (Platycephalidae, genera *Platycephalus*, *Neoplatycephalus* and *Cymbacephalus*). Bootstrap values greater than 50 shown. Specimen numbers for the Barcode of Life Database (BoLD, www.barcodinglife.org) given.

(d) *Dogfishes of the genus* Squalus

The genus *Squalus* is represented in Australian waters by three nominal species (*Squalus acanthias*, *Squalus megalops* and *Squalus mitsukurii*) and six species that have not been formally named (*Squalus* spp. A–F). These nine species are described in Last & Stevens (1994). We sequenced five *S. acanthias* (white-spotted dogfish), four *S. megalops* (spikey dogfish), ten *S. mitsukurii* (green-eye dogfish), four *S.* sp. B (eastern highfin spurdog), nine *S.* sp. C (western highfin spurdog), two *S.* sp. D (fatspine spurdog), two *S.* sp. E (western longnose spurdog) and five *S.* sp. F (eastern longnose spurdog), but did not have access to samples of *S.* sp. A.

The resulting phenogram of 41 sequences (figure 6) shows that there are three major clades: *S. acanthias, S. megalops,* and a *S. mitsukurii*/*S.* spp. B–F group. These separate out with bootstrap values of 94–100%. Species B, C, D, E and F form distinct subclades within the *S. mitsukurii*/*S.* spp. B–F clade, with bootstrap values of 92–100%. Four of the *S.* sp. C specimens were originally identified as *S. mitsukurii,* but were found to be *S.* sp. C following independent re-identification by a shark taxonomist; barcoding supported their identification as *S.* sp. C. All *S.* sp. C specimens have been collected from SW Australia. The single apparent *S. mitsukurii* in clade F was assigned to *S. mitsukurii* based on morphology; but it now appears likely that this specimen was in fact *S.* sp. F. This specimen, along with those of *S.* sp. F, came from the New South Wales coast off eastern Australia. There appear to be two remaining clades of *S. mitsukurii*: the cluster of five specimens with a bootstrap value of 99% were all from the Great Australian Bight, the cluster of four specimens with a bootstrap value of 78% comprised three from New South Wales and one from Western Australia. Whether these clades represent distinct undescribed species or geographic differentiation within a species remains to be assessed.

Omitting the one apparently misclassified specimen, genetic differences were again small within species, with a mean K2P distance of 0.33%. The mean interspecies distance within the genus was quite low at 4.17%.

4. DISCUSSION

This study has strongly validated the efficacy of *cox1* barcodes for identifying fish species. We sequenced (usually multiple) specimens of three species of chimaerids, 61 species of sharks and rays and 143 species of teleosts for the barcode region of *cox1*. With no exceptions, all 207 sequenced species could be discriminated. Nearly 98% of all species amplified with the one of two primer sets. Only five of 211 species failed to amplify with these protocols, and one of these amplified with a newly-designed primer set. The four failures came from varied fish groups and included congeners of species that amplified without problem; they may reflect either DNA degradation or primer mismatches. Since our two commonly-used primer sets are extremely similar, we are exploring the possibility of a single, quasi-universal fish primer set that incorporates inosine at the variable positions or has built-in degeneracy.

The GC content of the 655 bp mitochondrial *cox1* region was on average higher in the 143 species of Osteichthyes than in the 61 species of Chondrichthyes: 47.1% versus 42.2%. Saccone *et al.* (1999) reviewed

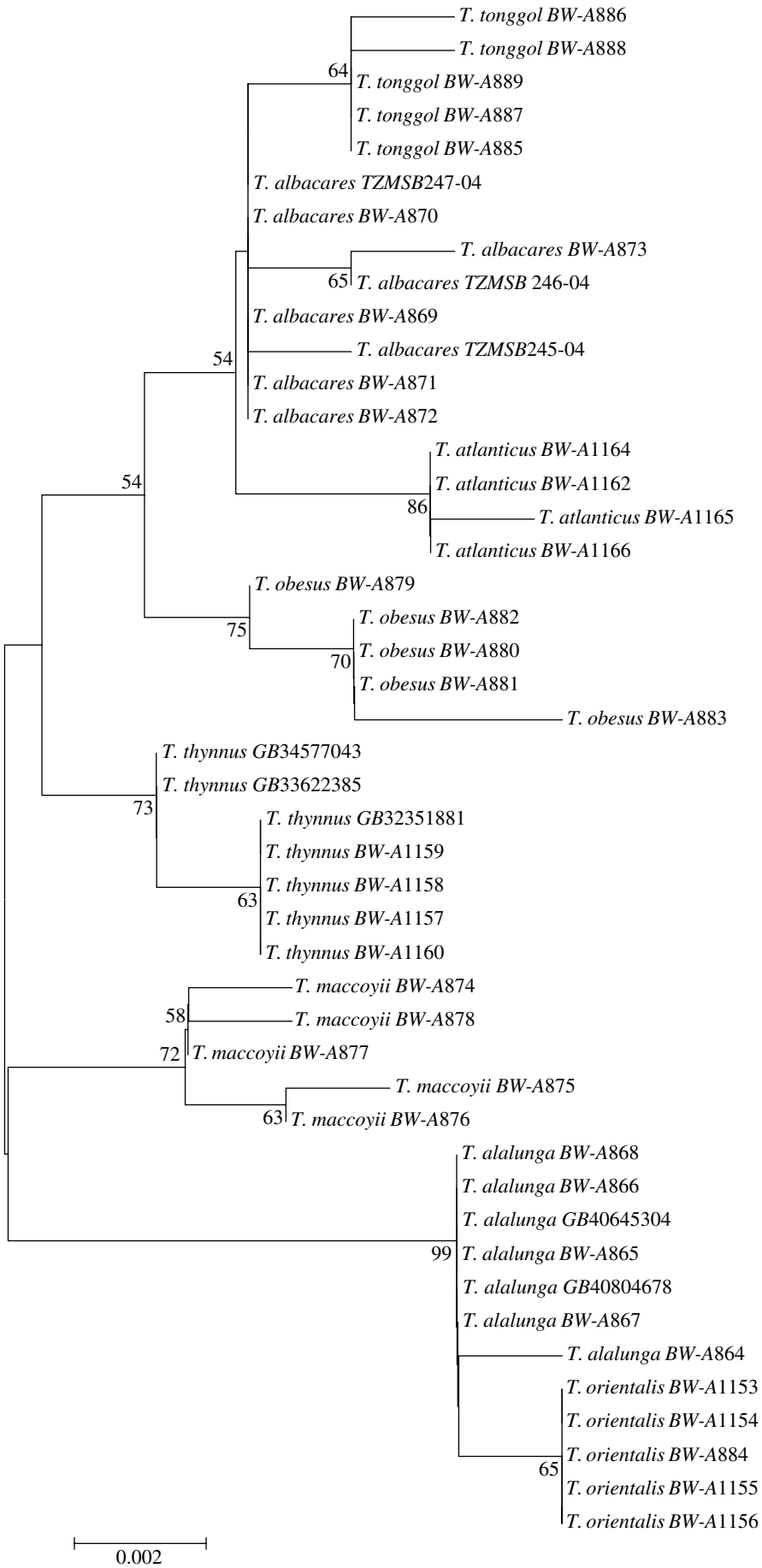

Figure 5. K2P distance neighbour-joining tree of 46 *cox1* sequences from the eight species of tuna of the genus *Thunnus*. Bootstrap values greater than 50 shown. Specimen numbers for the Barcode of Life Database (BoLD, www.barcodinglife.org) given.

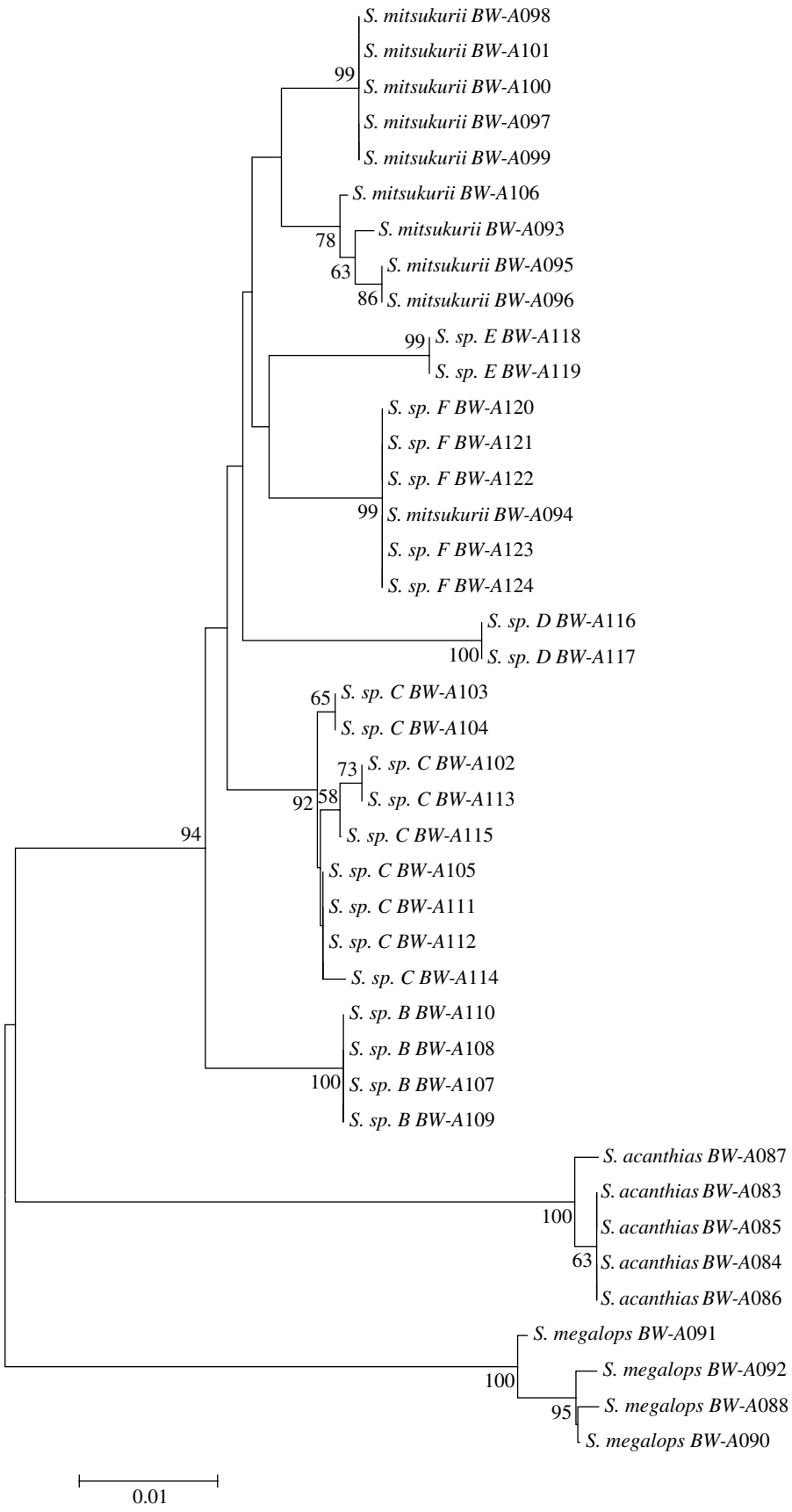

Figure 6. K2P distance neighbour-joining tree of 41 *cox1* sequences from eight species of dogfish of the genus *Squalus*. Bootstrap values greater than 50 shown. Specimen numbers for the Barcode of Life Database (BoLD, www.barcodinglife.org) given.

data from the complete mitochondrial genomes of nine Osteichthyes and three Chondrichthyes species, deriving GC contents of 43.2% and 38.4%, respectively. These values correspond reasonably well to ours, especially with respect to the higher GC content of the teleosts. We observed substantially more nucleotide changes at the 3rd codon position than the 1st, and more at the 1st than the 2nd. For example, the standard

errors of the GC percentages of the 3rd, 1st and 2nd bases of the Osteichthyes were 0.50, 0.10 and 0.02, respectively (see also figure 2). This reflects the fact that most synonymous mutations occur at the 3rd position, with a few at the 1st position and none at the 2nd. The higher GC content of the Osteichthyes compared with Chondrichthyes was largely due to 3rd base variation, with mean values of 41.1% and 29.9%, respectively, although 1st base differentiation was also observed. Within the Chondrichthyes, GC content of rays was higher than that of sharks (44.7% versus 41.0%), again largely due to GC_3 variation. The causes for the GC variation among teleosts, sharks and rays are not known.

No NUMTs (transfers of mtDNA *cox1* sequences into the nuclear genome) were observed. A review of the occurrence of NUMTs in plants and animals did not find any evidence of their existence in Actinopterygii (Bensasson *et al*. 2001), but a comparison of *Fugu rupripes* mtDNA and nuclear DNA sequences did detect seven or eight NUMTs (Richly & Leister 2004). This confirms the need for vigilance in examining fish amplicons for potential pseudogene status.

Confusion in taxonomic assignments as a result of inter-specific hybridization (Verspoor & Hammar 1991) does not seem to be a major issue—only two of 754 sequences (one in the genus *Pristiophorus* and one in *Plectropomus*) appeared in the 'wrong' congeneric species. This may represent either introgressive hybridization, or incorrect identification of the original specimen.

Although barcode analysis seeks only to delineate species boundaries, there is clearly some phylogenetic signal in *cox1* sequence data. For example, four major clusters were apparent in the NJ phenogram: chimaerids, rays, sharks and dogfish and teleosts. Congeneric species always clustered together and in most cases so did confamilial species. However, methodologies for phylogeny reconstruction from molecular data remain somewhat controversial, with a wide variety of disparate approaches possible (see, for example, Nei & Kumar 2000). We cannot hope to recover the true phylogeny of fishes from a 655 bp fragment of mitochondrial DNA through K2P distance and neighbour joining—rather more gene regions should be used (including nuclear genes) and additional analytic methods deployed including maximum parsimony and maximum likelihood.

Cox1 barcoding for species identification is far more powerful than, for example, protein fingerprinting. Reliable discrimination of *Thunnus* species using conventional protein electrophoresis is hard if not impossible (e.g. Yearsley *et al*. 1999), but we found that the same samples were readily identified by *cox1* sequencing.

Barcoding discriminated all of the fish species we examined, and would clearly be capable of unambiguously identifying individually isolated fish eggs, larvae, fillets and fins from these species. However, some taxa showed deeper divergence than others. For example, the average within-genus divergence of the flatheads (*Platycephalus, Neoplatycephalus, Cymbacephalus*) was 15.55%, considerably larger than the 4.17% of the genus *Squalus*, which is itself considerably larger than the within-*Thunnus* divergence of 1.11%. These differences among genera probably reflect the average age of species divergence, although within genera some species will be older than others. Nevertheless, it seems likely that, for example, the *Platycephalus* radiation preceded the *Thunnus* radiation. The large tunas have long been suspected of having diverged relatively recently (see, e.g. Elliott & Ward 1995 for allozyme evidence of limited nuclear DNA differentiation).

The *Thunnus* phenogram (figure 5) only clearly differentiates one clade (*T. alalunga* and *T. orientalis*, with 99% bootstrap support). The mitochondrial DNA similarity of *T. alalunga* and *T. orientalis* had been earlier described by Chow & Kishino (1995), from cytochrome b and ATPase sequencing. The three species suggested by Collette (1978) to comprise a separate subgenus *Neothunnus* (*Thunnus atlanticus*, *T. tonggol* and *T. albacares*) form a loosely defined clade in the *cox1* phenogram (56% bootstrap support). Our *cox1* phenogram is almost identical to one based on sequencing 400 bp of the mtDNA control region of the same eight species (Alvarado Bremer *et al*. 1997). One minor difference is that the control region tree gives a higher level of bootstrap support, 86%, to the proposed subgenus *Neothunnus*. The three *Neothunnus* species differ from the other five species (proposed subgenus *Thunnus*) in having central heat exchangers rather than lateral heat exchangers, and in being confined to more tropical waters (Collette 1978).

Many of the flathead species that were barcoded (figure 4) were earlier examined allozymically by Keenan (1991). His proposed cladogram of the Platycephalinae component is very similar to the Platycephalinae component of figure 3—the four subclades identified from *cox1* were also present in the allozyme tree. The mitochondrial and nuclear (allozyme) trees thus compare well. The genera *Neoplatycephalus* and *Cymbacephalus* appear to be monophyletic, *Platycephalus* being paraphyletic.

There do not appear to have been any prior surveys of genetic differentiation among species of the genus *Squalus*, although *S. acanthias* has been used as a model elasmobranch in some DNA sequencing studies (e.g. Stock & Powers 1995; Hong *et al*. 1996; Salaneck *et al*. 2003). *Squalus* was picked as a genus of particular interest as it included several provisional species (Last & Stevens 1994). The *cox1* data clearly supported the biological reality of the species, *Squalus* spp. B, C, D, E and F, as each of these species had bootstrap values close to 100% for their constituent individuals (figure 6). Indeed, there was greater genetic divergence among these species than among the well-recognized *Thunnus* species.

The various unresolved questions about specimen identification briefly presented here (for the genera *Squalus, Centrophorus, Hydrolagus, Pristiophorus*) indicate the need to retain whole voucher specimens wherever possible, or at least make an e-voucher from a photograph. While we retained a single voucher specimen for the majority of species discussed here, most other samples are only represented as small tissue samples. Retaining all specimens as vouchers will require significant infrastructure facilities as many fish are large: this may not be practical but it might be

feasible to retain whole specimens of most species at least until barcoding of those specimens is complete.

In our survey, conspecific samples often (but not always, see the *P. longispinis* example) came from adjacent areas. Thus we might have somewhat underestimated the extent of within-species diversity. However, any such effect is likely to be minor. Allozyme surveys of marine fish indicate that typically only about 5% percent of genetic variance comes from interpopulation differentiation. This percentage is appreciably higher for freshwater fish, around 20% on average (Ward *et al*. 1994). For freshwater fish, sampling should include individuals from different watersheds whenever possible.

Our results reveal that *cox1* barcoding will permit the unambiguous identification of the vast majority of fish species. We now intend to extend our survey to all Australian and all North American fish species. In the longer term, it is hoped that broader collaborations will enable the assembly of a global database of fish *cox1* sequences. This will mean collecting sequences from at least 25 000 species. Note that this will inevitably mean that for many species, multiple specimens from widely divergent locations will be sequenced, minimizing the concern expressed above about underestimating genetic diversity. With increasing application of DNA barcoding, many previously unrecognized fish species will be revealed through the discovery of deep divergence of *cox1* sequences within currently recognized species. There might also be instances of supposedly distinct species that have identical *cox1* sequences, suggesting the possibility of species fusion. Resolution of cases of this nature will require careful morphological analysis from expert taxonomists before any final recommendations can be made. Barcoding and morphological analysis should go hand-in-hand.

Once a global *cox1* barcode database has been established for fishes, anyone with direct or indirect access to a DNA sequencer will be able to identify, to a high degree of certainty, any fish egg, larva or carcass fragment. This will be an invaluable tool for fisheries managers, fisheries ecologists and fish retailers, and for those wishing to develop fish identification microarrays. The scientific and practical benefits of fish barcoding are manifold.

We thank those who helped with sequencing and data management at the University of Guelph (Janet Topan, Natalia Ivanova, Sujeevan Ratnasingham, Jeremy de Waard, Rob Dooh) and those that helped with specimen identification and sampling at CSIRO Marine and Atmospheric Research (Ross Daley, Alastair Graham, Tim Fountain, Gordon Yearsley). The Gordon and Betty Moore Foundation provided critical support for both the assembly of barcode sequences and the platform required for their analysis. The CSIRO Wealth from Ocean Flagship program facilitated the assembly and identification of specimens and some of the barcoding. Gordon Yearsley, Jawahar Patil and John Volkman provided useful comments on an earlier version of this manuscript.

REFERENCES

Alvarado Bremer, J. R., Naseri, I. & Ely, B. 1997 Orthodox and unorthodox phylogenetics relationships among tunas revealed by the nucleotide sequence analysis of the mitochondrial DNA control region. *J. Fish Biol.* **50**, 540–554.

Avise, J. C. 1994 *Molecular markers, natural history and evolution*. New York: Chapman & Hall.

Bensasson, D., Zhang, D-X., Hartl, D. L. & Hewitt, G. M. 2001 Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* **16**, 314–321. (doi:10.1016/S0169-5347(01)02151-6.)

Chow, S. & Kishino, H. 1995 Phylogenetic relationships between tuna species of the genus *Thunnus* (Scombridae: Teleostei): inconsistent implications from morphology, nuclear and mitochondrial genomes. *J. Mol. Evol.* **41**, 741–748. (doi:10.1007/BF00173154.)

Collette, B. B. 1978 Adaptations and systematics of the mackerels and tunas. In *The physiological ecology of tunas* (ed. G. D. Sharp & A. E. Dizon), pp. 7–39. New York: Academic Press.

Collette, B. B., Reeb, C. & Block, B. A. 2001 Systematics of the tunas and mackerels (Scombridae). In *Tuna: physiology, ecology and evolution* (ed. B. A. Block & E. D. Stevens), pp. 5–30. San Diego and London: Academic Press.

Elliott, N. G. & Ward, R. D. 1995 Genetic relationships of eight species of Pacific tunas (Teleostei, Scombridae) inferred from allozyme analysis. *Mar. Freshw. Res.* **46**, 1021–1032.

FAO 2002 *The state of world fisheries and aquaculture, part 1: world review of fisheries and aquaculture*. Rome: Food and Agricultural Organization, Fisheries Department.

FAO 2004 *Capture production 2002. FAO yearbook of fishery statistics 94/1*. Rome: Food and Agricultural Organization, Fisheries Department.

Hajibabaei, M., de Waard, 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. *Phil. Trans. R. Soc. B* **360**. (doi:10.1098/rstb.2005.1727.)

Hebert, P. D. N., Cywinska, A., Ball, S. L. & de Waard, J. R. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. B* **270**, 313–322. (doi:10.1098/rspb.2002.2218.)

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*. *Proc. Natl Acad. Sci. USA* **101**, 14 812–14 817. (doi:10.1073/pnas.0406166101.)

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

Hogg, I. D. & Hebert, P. D. N. 2004 Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. *Can. J. Zool.* **82**, 749–754. (doi:10.1139/z04-041.)

Hong, J., Salo, W. L., Chen, Y. Q., Atkinson, B. G. & Anderson, P. M. 1996 The promoter region of the carbamoyl-phosphate synthetase III gene of *Squalus acanthias*. *J. Mol. Evol.* **43**, 602–609.

Keenan, C. P. 1988 Systematics and evolution of Australian species of flatheads (Pisces, Platycephalidae). Ph.D. thesis, University of Queensland, Brisbane, Australia.

Keenan, C. P. 1991 Phylogeny of Australian species of flatheads (Teleostei, Platycephalidae) as determined by allozyme electrophoresis. *J. Fish Biol.* **39**(suppl. A), 237–249.

Kimura, M. 1980 A simple method of estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120. (doi:10.1007/BF01731581.)

Kumar, S., Tamura, K. & Nei, M. 2004 MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5**, 150–163. (doi:10.1186/1471-2105-5-150.)

Last, P. R. & Stevens, J. D. 1994 *Sharks and rays of Australia*. Melbourne, Australia: CSIRO Publishing.

Last, P. R., Manjaji, B. M. & Yearsley, G. K. 2005 *Pastinachus solocirostris* sp. nov., a new species of Stingray (Elasmobranchii: Myliobatiformes) from the Indo-Malay Archipelago. *Zootaxa* **1040**, 1–16.

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

Manwell, C. & Baker, C. M. A. 1963 A sibling species of sea-cucumber discovered by starch-gel electrophoresis. *Comp. Biochem. Physiol.* **10**, 39–53. (doi:10.1016/0010-406X(63)90101-4.)

Matsubara, K. & Ochiai, A. 1955 A revision of the Japanese fishes of the family Platycephalidae (the flatheads). *Mem. Coll. Agric., Kyoto Univ.* **68**, 1–109.

Moritz, C. & Cicero, C. 2004 DNA barcoding: promise and pitfalls. *PLoS Biol.* **2**, e354. (doi:10.1371/journal.pbio.0020354.)

Nei, M. & Kumar, S. 2000 *Molecular evolution and phylogenetics*. Oxford, UK: Oxford University Press.

Nelson, J. S. 1994 *Fishes of the world*, 3rd edn. New York: Wiley.

Pérez-Martin, R. I. & Sotelo, C. G. 2003 *Authenticity of species in meat and seafood products*. Eduardo Cabello, Spain: Association International Congress on Authenticity of Species in Meat and Seafood Products.

Richly, E. & Leister, D. 2004 NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* **21**, 1081–1084. (doi:10.1093/molbev/msh110.)

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G. & Reyes, A. 1999 Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* **238**, 195–209. (doi:10.1016/S0378-1119(99)00270-X.)

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

Salaneck, E., Ardell, D. H., Larson, E. T. & Larhammar, D. 2003 Three neuropeptide Y receptor genes in the spiny dogfish, *Squalus acanthias*, support en bloc duplications in early vertebrate evolution. *Mol. Biol. Evol.* **20**, 1271–1280. (doi:10.1093/molbev/msg133.)

Stock, D. W. & Powers, D. A. 1995 The cDNA sequence of the lactate dehydrogenase-A of the spiny dogfish (*Squalus acanthias*): corrections to the amino acid sequence and an analysis of the phylogeny of vertebrate lactate dehydrogenases. *Mol. Mar. Biol. Biotech.* **4**, 284–294.

Tautz, D., Arctander, P., Minelli, A., Thomas, R. H. & Vogler, A. P. 2002 DNA points the way ahead in taxonomy. *Nature* **418**, 479. (doi:10.1038/418479a.)

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

Verspoor, E. & Hammar, J. 1991 Introgressive hybridization in fishes: the biochemical evidence. *J. Fish Biol.* **39**(Suppl. A), 309–334.

Ward, R. D. & Grewe, P. M. 1994 Appraisal of molecular genetic techniques in fisheries. *Rev. Fish Biol. Fisheries* **4**, 300–325. (doi:10.1007/BF00042907.)

Ward, R. D., Woodwark, M. & Skibinski, D. O. F. 1994 A comparison of genetic diversity levels in marine, freshwater and anadromous fish. *J. Fish Biol.* **44**, 213–232. (doi:10.1111/j.1095-8649.1994.tb01200.x.)

Woese, C. R. & Fox, G. E. 1977 Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl Acad. Sci. USA* **97**, 8392–8396. (doi:10.1073/pnas.97.15.8392.)

Yearsley, G. K., Last, P. R. & Ward, R. D. (eds) 1999 *Australian seafood handbook: an identification guide to domestic species*, p. 461. Australia: CSIRO Marine Research. (Reprinted with minor corrections, 2001.)

Yearsley, G. K., Last, P. R. & Ward, R. D. (eds) 2003 *Australian seafood handbook: an identification guide to imported species*, p. 231. Australia: CSIRO Marine Research.

Zhang, D.-X. & Hewitt, G. M. 1996 Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol. Evol.* **11**, 247–251. (doi:10.1016/0169-5347(96)10031-8.)

The electronic supplementary material is available at http://dx.doi.org/10.1098/rstb.2005.1716 or via http://www.journals.royalsoc.ac.uk.