Molecular Ecology Resources (2009) **9** (Suppl. 1), 237–242 doi: 10.1111/j.1755-0998.2009.02649.x

BARCODING VERTEBRATES

DNA barcoding reveals overlooked marine fishes

TYLER S. ZEMLAK,*§ ROBERT D. WARD,† ALLAN D. CONNELL,‡ BRONWYN H. HOLMES† and PAUL D. N. HEBERT*

Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, 579 Gordon Street, Guelph, ON, Canada N1G 2W1, †CSIRO Wealth from Oceans Flagship, CSIRO Marine and Atmospheric Research, GPO Box 1538, Hobart, Tas 7001, Australia, ‡South African Institute of Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa

Abstract

With more than 15 000 described marine species, fishes are a conspicuous, diverse and increasingly threatened component of marine life. It is generally accepted that most large-bodied fishes have been described, but this conclusion presumes that current taxonomic systems are robust. DNA barcoding, the analysis of a standardized region of the cytochrome *c* oxidase 1 gene (COI), was used to examine patterns of sequence divergence between populations of 35 fish species from opposite sides of the Indian Ocean, chosen to represent differing lifestyles from inshore to offshore. A substantial proportion of inshore species showed deep divergences between populations from South African and Australian waters (mean = 5.10%), a pattern which also emerged in a few inshore/offshore species (mean = 0.84%), but not within strictly offshore species (mean = 0.26%). Such deep divergences, detected within certain inshore and inshore/offshore taxa, are typical of divergences between congeneric species rather than between populations of a single species, suggesting that current taxonomic systems substantially underestimate species diversity. We estimate that about one third of the 1000 fish species thought to bridge South African and Australian waters actually represent two taxa.

Keywords: *CO1*, conservation, cryptic species, DNA barcoding, marine

Received 10 October 2008; revision received 17 November 2008; accepted 28 November 2008

Introduction

Many marine fishes are thought to have broad geographical distributions, but for most species, this conclusion rests primarily on morphological congruence rather than upon hard evidence of gene exchange. Studies of population differentiation conducted to date are often limited in scope, only considering part of a species range and focusing on well-recognized taxa. Tagging studies provide the strongest basis for population connectedness on a large scale (Block *et al*. 2001; Bonfil *et al*. 2005), revealing that adults of some pelagic predators undertake long-distance migrations in oceanic waters. However, many other widely distributed species are shallow or mid-water forms whose adults cannot easily move between populations on different continental shelves because they are unknown to cross-abyssal regions. Many of these are bony fishes with planktonic larval and juvenile stages that might have the potential to bridge the abyssal divide, but studies on a few reef species have revealed unexpectedly strong site fidelity (Jones *et al*. 1999; Swearer *et al*. 1999). As a consequence, larval and juvenile dispersal may be a weaker force in maintaining cohesion across abyssal barriers than generally assumed, setting the stage for regional genetic divergence and ultimately speciation. In short, because we cannot directly evaluate dispersal and resultant gene flow, there is a need to examine patterns of genetic divergence in fishes with apparently broad distributions.

Here we analyse 35 commercially harvested species of Indo-Pacific fish that occur in both Australian and South African waters. Because of their economic importance, fished species represent an assemblage whose taxonomic status has been evaluated more critically than is the case for most other marine fishes. Our study examines five species

Correspondence: P. D. N. Hebert, Fax: 519-824-5703;
E-mail: phebert@uoguelph.ca
§Present address: Department of Biology, Dalhousie University, 1355 Oxford Street, Halifax, NS, Canada B3H 4J1.

© 2009 Blackwell Publishing Ltd

characteristic of the open oceans (offshore), nine inshore/offshore species found in both coastal and offshore waters, and 21 inshore species restricted to coastal habitats. The 7000-km breadth of the abyssal zone between Australia and South Africa must be a substantial barrier to inshore fishes, although gene flow in these species could, in principle, be mediated either by direct dispersal of planktonic larvae or by stepping-stone dispersal among populations scattered along the 19 000-km coastal arc from Africa to Australia.

Our study employs DNA barcoding to provide a standardized measure of sequence divergence within and between Australian and South African populations of 35 selected species. Specifically, we examine the extent of sequence diversity in the 648-bp region of the cytochrome *c* oxidase 1 (COI) gene that has been adopted as the standard barcode for members of the animal kingdom (Hebert *et al*. 2003a). Past studies have established two important facts — the regular presence of substantial sequence divergence between congeneric taxa and a relative dearth of variation within species (Hebert *et al*. 2003a, b; Hebert *et al*. 2004a, b; Hogg & Hebert 2004; Barrett & Hebert 2005; Smith *et al*. 2005; Ward *et al*. 2005; Hajibabaei *et al*. 2006). For example, a study of barcode variation in > 200 Australian teleosts and chondrichthyans revealed that intra-specific variation averaged 0.39%, 25-fold less than the average congeneric separation of 9.93% (Ward *et al*. 2005). A subsequent study of > 200 chondrichthyans gave slightly lower averages of 0.37% and 7.48% respectively (Ward *et al*. 2008). Based on these and similar results with other animal groups, we examined patterns of CO1 divergence in 35 species found in both Australian and South African waters. If their component populations consititute a single species, there should be little sequence divergence between representatives from the two areas. Conversely, evidence of deep genetic divergences between these regional populations would suggest that current taxonomic systems overlook species.

Materials and methods

Approximately one-quarter of the sequences considered in this study derive from earlier work on Australian fishes (Ward *et al*. 2005); the remainder represent new records (Table S1, Supporting information). Sampling strategies followed traditional barcoding objectives of analysing five individuals per species where possible. However, sometimes only one or two individuals per region could be collected. This limited sampling precludes accurate calculations of population parameters (e.g. gene flow, regional variation metrics), but is highly unlikely to have affected the main conclusions of this study aimed at testing species limits. A total of 229 individuals from 35 species were examined. Each specimen was expertly identified as an accurate representative from regional populations of each fish species and linked to a voucher specimen deposited at institutions in both South Africa and Australia. Voucher specimens for all South African specimens, excepting the largest species, were deposited in the South African Institute of Aquatic Biodiversity in Durban. The Australian samples derive from the tissue repository at the Commonwealth Scientific and Industrial Research Organization (CSIRO) Marine Laboratory in Hobart with one or more voucher specimens per species deposited in the Australian National Fish Collection. Taxonomic assignments follow FishBase (Froese & Pauly 2006).

DNA extraction, polymerase chain reaction (PCR), and sequencing were performed at the University of Guelph or the CSIRO Marine Laboratory following standard DNA barcoding methods (Hajibabaei *et al*. 2005). Total DNA was extracted from frozen or ethanol-preserved muscle or gill samples using Chelex dry release (Hajibabaei *et al*. 2005). Approximately 648 bp were amplified from the 5′ region of the COI gene using various combinations of the fish-specific primers described in Ward *et al*. (2005): FishF1-TCAACCAACCACAAAGACATTG-GCAC, FishF2-TCGACTAATCATAAAGATATCGGCAC, FishR1-TAGACTTCTGGGTGGCCAAAGAATCA, FishR2-ACTTCAGGGTGACCGAAGAATCAGAA. The 25-μL PCR mixes included 18.75 μL of ultrapure water, 2.25 μL of 10× PCR buffer, 1.25 μL of $MgCl_2$, 0.25 μL of each primer (0.1 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. 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 4.0 °C. Products were labelled using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) and sequenced bidirectionally using the ABI 3730 capillary sequencer following manufacturer's instructions. Sequences, electropherograms and primer sequences are available in the completed project file 'Overlooked Fishes in Marine Settings' on the Barcode of Life Data System (BOLD, www.barcodinglife.org). All sequences have also been deposited in GenBank (Table S1). A Kimura 2-parameter (K2P) distance metric was employed for sequence comparisons (Kimura 1980), including genetic distance calculations and neighbour-joining (NJ) analysis (Taxon ID tree; Table S2, Supporting information), using the BOLD Management and Analysis System.

We employed a sequence divergence of 3.5% as a screening threshold for overlooked species. This metric was based on the recommendation by Hebert *et al*. (2004b) that an appropriate screening threshold for flagging provisional species is 10× that of the average intra-species variation for a focus group. Three and a half per cent approximates 10× the average within-species COI sequence typical of marine fishes (Ward *et al*. 2005; Ward *et al*. 2008; Steinke *et al*. 2009; T. Zemlak and R. Ward, unpublished data), and is a

© 2009 Blackwell Publishing Ltd

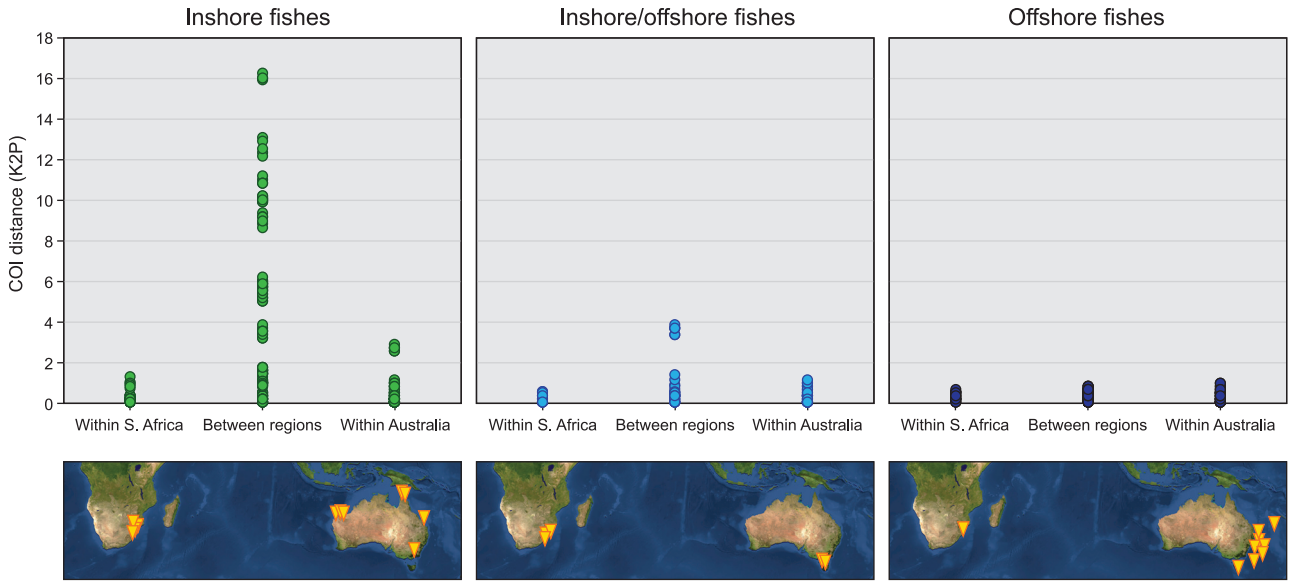

Fig. 1 Within-species COI sequence divergence (K2P) for 35 marine fishes collected on either side of the Indian Ocean (collection locations indicated by yellow triangles).

conservative threshold. We are confident that this approach presents a reliable first approach to evaluating the species status in marine fishes considering that surveys of COI variation already corroborate very well with traditional taxonomic assignments for several animal groups. In addition, most provisional species flagged by COI intraspecific thresholds ultimately gain species-level status following taxonomic revision (e.g. Hebert *et al*. 2004a, b; Ward *et al*. 2005, 2007; Kerr *et al*. 2007; Smith *et al*. 2008). Furthermore, a recent review by Zink & Barrowclough (2008) corroborates the power of mtDNA-based studies to accurately depict patterns of population history and species limits, and report that the signal from mtDNA loci is robust and rarely contradicted by supplemental analyses at nuclear markers.

Results

On a regional scale, genetic variation at COI was low for most taxa (Fig. 1). The one exception was the needlescaled queenfish, *Scomberoides tol*, which included two lineages in each region with 8.70% divergence. Intraspecific sequence divergences were otherwise low — Australian populations of the remaining 34 species averaged 0.28% (range 0–2.82%) and South African populations averaged 0.21% (range 0–1.27%). Barring *S. tol*, within-region intraspecific variation was typical of that expected among conspecific marine fishes.

By contrast, comparisons between Australian and South African populations varied dramatically, from a low of 0.05% to a high of 16.2% (Table 1). Species were allocated to one of three habitat groups — inshore, inshore/offshore, and offshore. While limited genetic divergence was detected for species in each group, the proportion of such species varied widely among groups (Fig. 1). Inshore fishes had the greatest number of exceptions, with eight out of 21 species comparisons between Australia and South Africa exceeding the 3.5% sequence divergence threshold. These high divergences ranged from *Argyrops spinifer* (5.30%) and *Lethrinus nebulosus* (5.68%) to the more marked divergences detected within *Priacanthus hamrur* (8.91%), *Rhabdosargus sarba* (10.01%), *Platycephalus indicus* (11.03%), *Bodianus perditio* (12.48%), *Parupeneus heptacanthus* (16.00%) and *Otolithes ruber* (16.24%). A ninth species, *Ariomma indica* (3.39%), also showed a high divergence. Inshore/offshore species revealed fewer exceptions; seven of nine species showed low divergence, but two, *Sphyrna lewini* (3.5%) and *Scomberomorus commerson* (3.6%), showed much higher divergence. All five oceanic species showed very limited genetic divergence, ranging from less than 0.10% in the tunas (*Thunnus albacares*, *Euthynnus affinis*) to a maximum of 0.53% between individuals of opah (*Lampris guttatus*). Overall, inshore fishes showed the highest average divergence (5.10%), followed by an intermediate level of variation among inshore/offshore taxa (mean = 0.84%), and very limited divergences among offshore species (mean = 0.26%).

Discussion

Our study has provided evidence for a substantial number of overlooked species in commercially important fishes. One of the 35 species targeted in our study, *Scomberoides tol*, appears to represent a broadly distributed sibling species pair whose distribution spans the Indian Ocean. However,

© 2009 Blackwell Publishing Ltd

Table 1 Pairwise intraspecific divergences between regions for specimens of species common to both South Africa and Australia

	Species	No. of individuals (Aus.)	No. of individuals (S.A.)	Percentage of sequence divergence ($\bar{X}$ | ± SE)
I Inshore	*Lutjanus rivulatus*	1	3	0.05 ± 0.05
	Carcharhinus amboinensis	2	1	0.08 ± 0.08
	Cephalopholis miniata	2	2	0.09 ± 0.05
	Caranx ignobilis	1	2	0.16 ± 0.16
	Lethrinus rubrioperculatus	1	2	0.40 ± 0.06
	Chanos chanos	3	2	0.48 ± 0.01
	Lutjanus argentimaculatus	5	2	0.59 ± 0.11
	Parupeneus indicus	3	3	0.60 ± 0.14
	Chelidonichthys kumu	5	3	1.12 ± 0.15
	Epinephelus rivulatus	1	6	1.95 ± 0.28
	Cephalopholis sonnerati	5	3	2.00 ± 0.18
	Ariomma indica	3	3	3.39 ± 0.09
	Argyrops spinifer	5	2	5.30 ± 0.09
	Lethrinus nebulosus	4	3	5.68 ± 0.04
	*Scomberoides tol**	2	3	8.70 ± 0.05
	Priacanthus hamrur	3	2	8.91 ± 0.08
	Rhabdosargus sarba	5	3	10.01 ± 0.03
	Platycephalus indicus	5	1	11.03 ± 0.07
	Bodianus perditio	4	2	12.48 ± 0.14
	Parupeneus heptacanthus	5	1	16.00 ± 0.04
	Otolithes ruber	1	2	16.24 ± 0.37
				5.10 ± 0.37
II Inshore/Offshore	*Carcharhinus obscurus*	2	2	0.00
	Galeocerdo cuvier	2	3	0.00
	Hoplostethus mediterraneous	4	4	0.13 ± 0.04
	Carcharhinus limbatus	5	3	0.26 ± 0.02
	Pristipomoides filamentosus	3	3	0.34 ± 0.06
	Carcharodon carcharias	1	2	1.00 ± 0.03
	Pomatomus saltatrix	5	1	1.35 ± 0.00
	Sphyrna lewini	2	1	3.54 ± 0.00
	Scomberomorus commerson	5	1	3.55 ± 0.10
				0.84 ± 0.15
III Offshore	*Thunnus albacares*	5	5	0.08 ± 0.02
	Euthynnus affinis	5	3	0.09 ± 0.03
	Xiphias gladius	5	3	0.31 ± 0.05
	Coryphaena hippurus	5	5	0.46 ± 0.05
	Lampris guttatus	5	1	0.53 ± 0.09
				0.26 ± 0.03

*Omitted from calculations (see text).

all other cases of newly revealed diversity involve lineages that share a species epithet, but show marked genetic divergence between Australian and South African waters. These cases of genetic subdivision appear linked to biological attributes.

Five of the species (i.e. tunas, swordfish, opah, dolphinfish) that we examined spend most of their time in oceanic habitats. Populations of these pelagic species showed little or no genetic divergence between South Africa and Australia. Seven of nine species typical of inshore/offshore environments shared this lack of regional divergence, but the scalloped hammerhead shark, *Sphyrna lewini*, and the narrow-barred Spanish mackerel, *Scomberomorus commerson*, possessed divergences equalling or greater than the threshold value of 3.5%, suggesting that they too represent a sibling species pair. Deep divergences within *Sphyrna lewini* have recently been commented on, with one lineage suspected to be a new species (Quattro *et al*. 2006); we suspect that the Australian–South African divergence is consistent with there being yet another species of hammerhead.

Evidence for overlooked species was far more frequent for inshore species with eight of 20 taxa (45%) showing greater than 3.5% sequence divergence between Australian and South African lineages, and a ninth species with 3.4%

© 2009 Blackwell Publishing Ltd

divergence. Such deep divergences characterize congeneric species rather than population differentiation within species. This suggests that current taxonomic systems for marine fishes substantially underestimate species diversity. These newly revealed putative species undoubtedly have population sizes that are smaller and ranges that are more limited than those of their 'parent' taxa. Because these overlooked species have long histories of independent evolution and because they are commercially harvested, this revelation of overlooked diversity raises a new layer of conservation concerns. The low genetic divergences in some inshore species likely reflect gene flow mediated through short-distance dispersal events, a hypothesis that could be validated through studies along the coastal arc connecting Australia and South Africa.

Viewed collectively, our results reveal that many fishes thought to be shared between South Africa and Australia most likely represent different species. Our study has examined only a small segment of the shared fauna — nearly 1000 fish species are thought to occur in both South African and Australian waters, 53% of which possess inshore lifestyles (Froese & Pauly 2006). Based on our results, we estimate that future DNA barcoding studies will reveal deep genetic divergences in about 300 of these taxa, suggesting that they are actually species pairs. Moreover, we expect further revelations of overlooked diversity as barcode studies assess genetic divergences in fish species currently thought to span other abyssal divides, investigations that will result from the ongoing effort to gather DNA barcodes for all marine fishes (www.fishbol.org, see Ward *et al*. 2009).

Acknowledgements

We are deeply grateful to Phil Heemstra, Peter Last and Gus Yearsley for validating identifications of the specimens used in this study. We thank Nataly Ivanova, Jeremy deWaard and Janet Topan for aid in the DNA analyses and both Rob Dooh and Sujeevan Ratnasingam for assistance in data entry and analysis. We also thank Nicolas Bailly for his assistance with FishBase related queries. Funding for this research was provided by the Gordon and Betty Moore Foundation, NSERC, Genome Canada through the Ontario Genomics Institute and the CSIRO Wealth from Oceans Flagship program. We also acknowledge and thank NSERC and OGS for fellowships awarded to T.S.Z.

Conflicts of interest

The authors have no conflict of interest to declare and note that the funders of this research had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Barrett RDH, Hebert PDN (2005) Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, **83**, 481–491.

Block BA, Dewar H, Blackwell SB *et al*. (2001) Migratory movements, depth preferences, and thermal biology of the Atlantic bluefin tuna. *Science*, **293**, 1310–1314.

Bonfil R, Meÿer M, Scholl MC *et al*. (2005) Transoceanic migration, spatial dynamics, and population linkages of white sharks. *Science*, **310**, 100–103.

Froese R, Pauly D (eds) (2006) *Fishbase: World Wide Web Electronic Publication*. Available from URL: http://www.fishbase.org

Hajibabaei M, deWaard JR, Ivanova NV *et al*. (2005) Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1959–1967.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences, USA*, **103**, 968–971.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **270**, S96–S99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004b) Identification of birds through DNA barcodes. *Public Library of Science Biology*, **2**, 1657–1663.

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

Jones GP, Milicich MJ, Emslie MJ, Lunow C (1999) Self-recruitment in a coral reef fish population. *Nature*, **402**, 802–804.

Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN (2007) Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, **7**, 535–543.

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

Quattro JM, Stoner DS, Driggers WB *et al*. (2006) Genetic evidence of cryptic speciation within hammerhead sharks (Genus *Sphyrna*). *Marine Biology*, **148**, 1143–1145.

Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1825–1834.

Smith P, Steinke D, Mcveagh S, Stewart AL, Struthers C, Roberts C (2008) Molecular analysis of Southern Ocean skates (*Bathyraja*) reveals a new species of Antarctic skate. *Journal of Fish Biology*, **73**, 1170–1182.

Steinke D, Zemlak TS, Boutillier JA, Hebert PDN (2009) DNA barcoding Pacific Canada's fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, submitted.

Swearer SE, Caselle JE, Lea DW, Warner RR (1999) Larval retention and recruitment in an island population of coral-reef fish. *Nature*, **402**, 799–802.

© 2009 Blackwell Publishing Ltd

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.

Ward RD, Holmes BH, Zemlak TS, Smith PJ (2007) Part 12 — DNA barcoding discriminates spurdogs of the genus *Squalus*. In: *Descriptions of New Dogfishes of the Genus Squalus (Squaloidea: Squalidae)* (eds Last PR, White WT, Pogonoski JJ), pp. 117–130. CSIRO Marine and Atmospheric Research Paper 014, Tasmania, Australia.

Ward RD, Holmes BH, White WT, Last PR (2008) DNA barcoding Australasian chondrichthyans: results and potential uses in conservation. *Marine and Freshwater Research*, **59**, 57–71.

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107–2121.

Supporting Information

Additional supporting information may be found in the online version of this article:

Table S1 List of GenBank and BOLD accession numbers. The symbol (*) indicates those sequences derived from previously published work (Ward *et al*. 2005).

Table S2 BOLD Taxon ID tree of 229 individuals from 35 species.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2009 Blackwell Publishing Ltd