Molecular Ecology Notes (2007) **7**, 544–548 doi: 10.1111/j.1471-8286.2007.01748.x

BARCODING

Universal primer cocktails for fish DNA barcoding

NATALIA V. IVANOVA,* TYLER S. ZEMLAK, ROBERT H. HANNER and PAUL D. N. HEBERT
Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract

Reliable recovery of the 5′ region of the cytochrome *c* oxidase 1 (COI) gene is critical for the ongoing effort to gather DNA barcodes for all fish species. In this study, we develop and test primer cocktails with a view towards increasing the efficiency of barcode recovery. Specifically, we evaluate the success of polymerase chain reaction amplification and the quality of resultant sequences using three primer cocktails on DNA extracts from representatives of 94 fish families. Our results show that M13-tailed primer cocktails are more effective than conventional degenerate primers, allowing barcode work on taxonomically diverse samples to be carried out in a high-throughput fashion.

Keywords: COI, degenerate primers, M13-tailed primers, species identification

Received 11 January 2007; revision accepted 7 February 2007

Introduction

Fishes comprise nearly half of all vertebrate species; the group includes approximately 15 700 marine and 13 700 freshwater species (FishBase: www.fishbase.org). The Fish Barcode of Life Initiative (FISH-BOL; www.fishbol.org) is a collaborative international research effort, which seeks to establish a reference library of DNA barcodes for all fish species derived from voucher specimens with authoritative taxonomic identifications (Hanner *et al*. 2005). Once completed, FISH-BOL will enable a fast, accurate, and cost-effective system for molecular identification of the world's ichthyofauna. The benefits of this work include facilitating species identification, flagging potentially previously unrecognized species, and enabling identifications where traditional methods are not applicable, such as for immature stages or body fragments. FISH-BOL will also provide a powerful tool for enhanced understanding of the natural history and ecological interactions of fish species.

Obtaining high-quality sequence records from the barcode region of COI is a key requirement for the FISH-BOL enterprise. Moreover, with the goal of analysing at least 10 specimens per species, this effort will likely involve sequencing more than 0.5 million specimens. Ward *et al*. (2005) carried out a proof-of-principle study that compiled barcodes for 200 species of commercially important Australian marine fishes. Since then, an additional 5000 barcodes have been generated from over 2000 species. However, there has not been a serious effort to hone analytical protocols, a gap that this study addresses. Specifically, we seek to identify protocols that enable both efficient polymerase chain reaction (PCR) amplification of the barcode region and that deliver high quality sequence data.

Ward *et al*. (2005) used two forward and two reverse primers in all four pairwise combinations to amplify DNA barcodes from the 200 species in their study. This approach delivered amplicons for all but one of the target species, demonstrating the reliable amplification of COI from a diversity of fishes with only a few primers. However, the need for four PCR amplifications of each specimen is undesirably complex. Two possible solutions include the generation of a single primer set with degenerate sites or the techniques in this study assembly of a cocktail whose component primers are tailed with M13 to facilitate high throughput sequencing. We employ both, and test their effectiveness on a single species from each of 94 different fish families from diverse habitats (i.e. freshwater, diadramous, marine) and divergent evolutionary lineages.

Correspondence: N.V. Ivanova, Fax: (519) 824 5703; E-mail: nivanova@uoguelph.ca

Materials and methods

Primer design

COI sequences from all 159 mitochondrial fish genomes (GenBank, January 2006) were aligned in BIOEDIT (Hall 1999). Potential primer regions were analysed in CODEHOP (Rose *et al*. 1988) available at (http://blocks.fhcrc.org/

Table 1 PCR primer sets or cocktails used to amplify either 16S rDNA or COI. M13 tails are highlighted when present (*indicates original reference for the untailed version of each primer)

Name	Ratio	Cocktail name/Primer sequence 5′–3′	Product/primer position	References
16S			**2974–3546**	
16Sar-5′	1	CGCCTGTTTATCAAAAACAT	2954–2973	Palumbi (1996)
16Sbr-3′	1	CCGGTCTGAACTCAGATCACGT	3568–3547	Palumbi (1996)
COI-1			**6472–7126**	
FF2d	1	TTCTCCACCAACCACAARGAYATYGG	6446–6471	This study
FR1d	1	CACCTCAGGGTGTCCGAARAAYCARAA	7152–7127	This study
COI-2		**C_VF1LFt1–C_VR1LRt1**	**6472–7129**	
LepF1_t1	1	TGTAAAACGACGGCCAGTATTCAACCAATCATAAAGATATTGG	6446–6471	*Hebert *et al*. 2004
VF1_t1	1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAAAGACATTGG	6446–6471	*Ivanova *et al*. 2006
VF1d_t1	1	TGTAAAACGACGGCCAGTTCTCAACCAACCACAARGAYATYGG	6446–6471	*Ivanova *et al*. 2006
VF1i_t1	3	TGTAAAACGACGGCCAGTTCTCAACCAACCAIAAIGAIATIGG	6446–6471	*Ivanova *et al*. 2006
*Lep*RI_t1	1	CAGGAAACAGCTATGACTAAACTTCTGGATGTCCAAAAAATCA	7155–7130	*Hebert *et al*. 2004
VR1d_t1	1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCRAARAAYCA	7155–7130	*Ivanova *et al*. 2006
VR1_t1	1	CAGGAAACAGCTATGACTAGACTTCTGGGTGGCCAAAGAATCA	7155–7130	*Ward *et al*. 2005
VR1i_t1	3	CAGGAAACAGCTATGACTAGACTTCTGGGTGICCIAAIAAICA	7155–7130	*Ivanova *et al*. 2006
COI-3		**C_FishF1t1–C_FishR1t1**	**6475–7126**	
VF2_t1	1	TGTAAAACGACGGCCAGTCAACCAACCACAAAGACATTGGCAC	6448–6474	*Ward *et al*. 2005
FishF2_t1	1	TGTAAAACGACGGCCAGTCGACTAATCATAAAGATATCGGCAC	6448–6474	*Ward *et al*. 2005
FishR2_t1	1	CAGGAAACAGCTATGACACTTCAGGGTGACCGAAGAATCAGAA	7152–7127	*Ward *et al*. 2005
FR1d_t1	1	CAGGAAACAGCTATGACACCTCAGGGTGTCCGAARAAYCARAA	7152–7127	This study
M13F (–21)		TGTAAAACGACGGCCAGT		Messing (1983)
M13R (–27)		CAGGAAACAGCTATGAC		Messing (1983)

blocks/codehop.html) with *Danio rerio* codon usage. Primers for each cocktail were designed at the same positions in the COI gene so that sequences generated would be readily interpretable. M13 tails were derived from Messing (1983), with minor changes: a subset of forward primers (VF1 and its modifications, VF2, FishF2, and FF2d) contained a 5′-T-nucleotide identical to the 3′ end of corresponding M13 tag and therefore this nucleotide was eliminated from the tag. To make the LepF1_t1 primer compatible with the VF1_t1 and the C_VF1LFt1 primer cocktail, we added an entire M13 (–21) forward tag sequence to LepF1 primer. For reverse primers, a similar approach was used to equalize the length of FishR2_t1 and FR1d_t1 primers with the M13 (–27) reverse tag. Primer sequences are shown in Table 1.

PCR amplification and sequencing

Whole genomic DNA was extracted from muscle tissue of 94 species (Appendix 1, available at www.dnabarcoding.ca/CCDB_DOCS/UNIVERSAL_PRIMER_COCKTAILS_FOR_FISH_DNA_BARCODING_Appendix_1.pdf) each belonging to a different family. These taxa included representatives from all major fish lineages (e.g. Myxini, Cephalaspidomorphi, Holocephali, Elasmobranchii, Actinopterygii). DNA was extracted using a standard glass fibre extraction protocol (Ivanova *et al*. 2006). To evaluate the universality of primers, these 94 samples were assembled along with two negative controls to create a 96-well test-plate. Four PCRs were conducted on this plate. The first reaction employed universal primers for the mitochondrial 16S rRNA gene (16Sar and 16Sbr; Palumbi, 1996) to provide a positive control for DNA extraction. The other three reactions targeted the COI barcode region using varied primer combinations (see Table 1 for details). All PCRs had a total volume of 12.5 μL and included: 6.25 μL of 10% trehalose, 2.00 μL of ultra pure water, 1.25 μL 10× PCR buffer [10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1% Triton X-100], 0.625 μL $MgCl_2$ (50 mM), 0.125 μL of each primer cocktail (0.01 mM), 0.0625 μL of each dNTP (10 mM), 0.0625 μL of *Taq* DNA Polymerase (New England Biolabs), and 2.0 μL of DNA template. The thermocycle profile for 16S, COI-1 and COI-3 consisted of 94 °C for 2 min, 35 cycles of 94 °C for 30 s, 52 °C for 40 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min. Conditions for COI-2 were: 94 °C for 1 min, five cycles of 94 °C for 30 s, 50 °C for 40 s, and 72 °C for 1 min, followed by 35 cycles of 94 °C for 30 s, 54 °C for 40 s, and 72 °C for 1 min, with a final extension at 72 °C for 10 min.

PCR products were visualized on a 2% agarose gel using an E-Gel96 Pre-cast Agarose Electrophoresis System

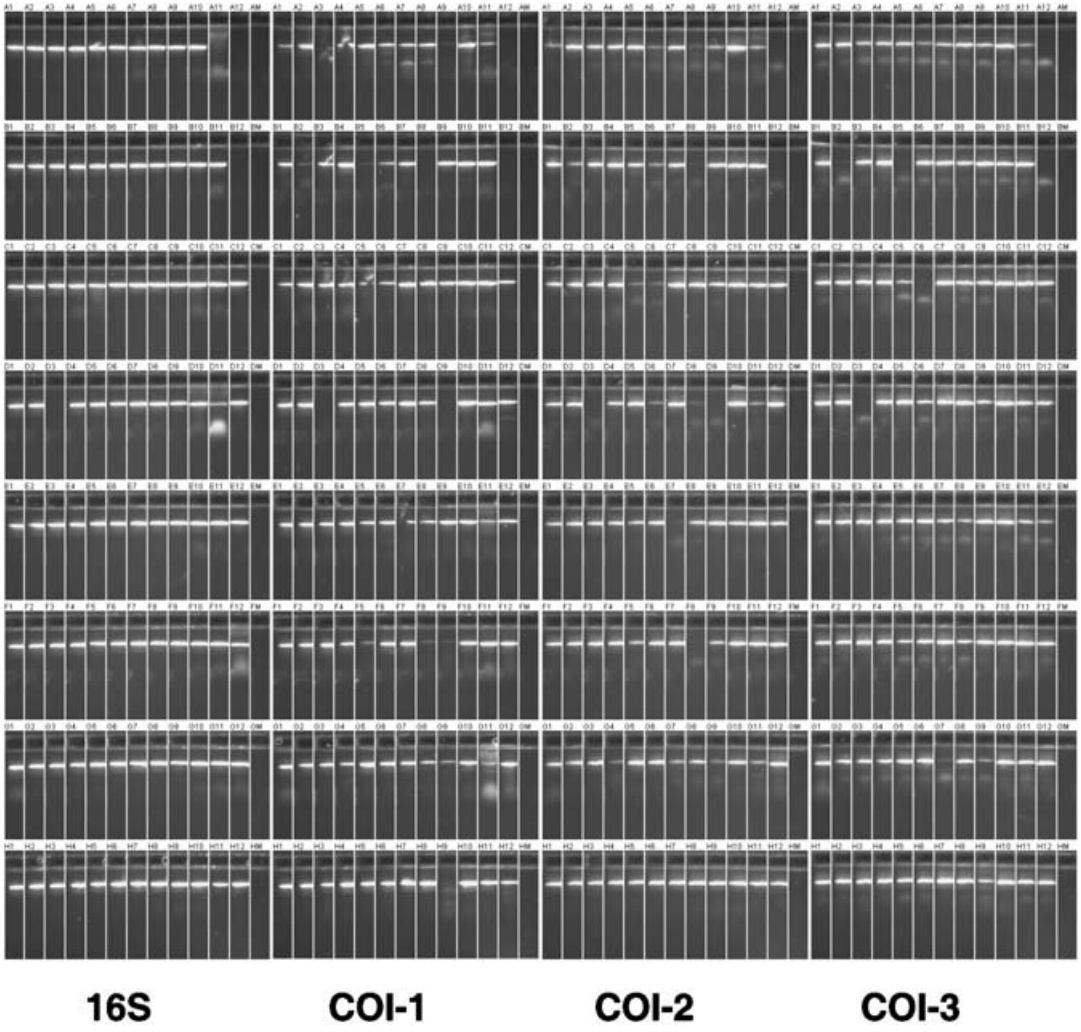

Fig. 1 Images of PCR amplicons for representatives of 94 fish families.

(Invitrogen) and bidirectionally sequenced using the BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems, Inc.) on an ABI 3730 capillary sequencer (see Hajibabaei *et al*. 2005 for details). PHRED scores and length of read (LOR) scores were generated using SEQUENCING ANALYSIS software version 5.1.1 (Applied Biosystems). Bidirectional sequences were assembled in SEQSCAPE version 2.1.1 (Applied Biosystems) and manually edited.

Results

Strong PCR products were generated with the 16S primers for all but two samples indicating that DNA templates were generally high quality (Fig. 1). The DNA extract from *Goniistius zonatus* failed to amplify in all PCRs and was omitted for further consideration. DNA from a second species (*Acipenser fulvescens*) failed to amplify for 16S, but was weakly amplified by all COI primer sets, suggesting DNA degradation.

Each of the three primer sets was very effective (96.8%) in amplifying the target region of COI (Fig. 1) and only one case of nonspecific amplification was detected (well G11, COI-1). Amplicon intensities varied with COI-3 generating the greatest proportion of high intensity bands (Fig. 2). Average sequencing success of amplicons was high for COI-3 (95.2%) and COI-2 (93.0%), but lower for COI-1 (86.0%). Average PHRED scores and read lengths varied substantially. Both quality scores (38) and read length (608 bp) were lowest for COI-1. The COI-2 cocktail delivered the longest average read length (645 bp) and a PHRED of 39. By comparison, the COI-3 cocktail had an intermediate read length (631 bp), but the highest quality (41). Barcodes were recovered from all 93 samples with nondegraded DNA, although this required 'cherrypicking' the few cases

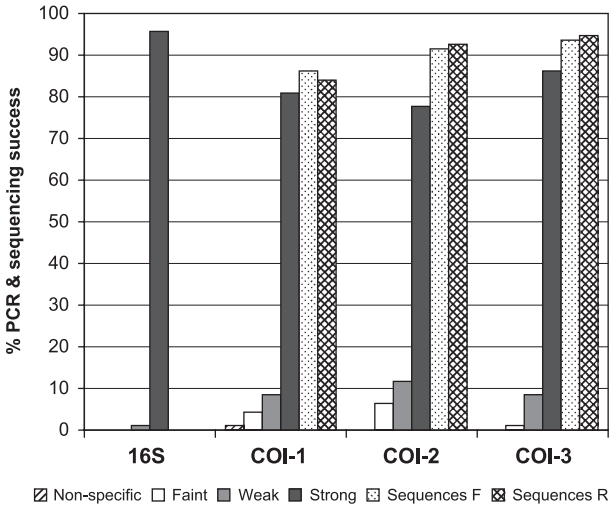

Fig. 2 Amplification and sequencing success with three COI primer cocktails in a test plate with single species from 94 fish families. Amplification of 16S rDNA was used to check DNA quality.

of failure with a specific primer set. It is worth noting that only the COI-3 cocktail amplified the sole Myxini tested (*Eptatretus cirrhatus*) and one of the two holocephalids (*Harriotta raleighana*). The target region for the other 91 species was amplified by at least two of the cocktails.

Discussion

The 5′ region of the COI gene was selected as the basis for a DNA barcoding system, in part, because of the availability of primers aiding its recovery from a broad range of taxa (Hebert *et al*. 2003). However, there is enough variation in the flanking segments on either side of the barcode region to require multiple primer combinations to gain COI amplicons from some taxonomic groups. The fishes represent one group where multiple rounds of PCR have been required to recover barcodes (Ward *et al*. 2005). The individual primer sets developed in this study largely overcome this difficulty, because each amplify the barcode region from most taxa, meaning that few samples require a secondary round of amplification. Considering the broad diversity of fish species examined in our study, we expect this success to extend to other fishes. Interestingly, the COI-2 primer cocktail was designed for amplification of the barcode region from mammals (untailed version — Ivanova *et al*. 2006; tailed version — this study; Clare *et al*. 2007), but this study shows that it also performs very well for fishes. Conversely, the COI-3 cocktail designed for fishes is also very effective for mammals, amphibians and reptiles (N.V. Ivanova, personal observation). Jointly, these cocktails have amplified the barcode region for every species (> 3000) in these groups that we have tested.

Amplification

In the absence of a positive control, failed amplifications for COI can be attributed to a primer mismatch when they simply reflect degraded template. Universal 16S primers provide a simple test for DNA quality because they amplify a product that is similar in size to the barcode region. As a result, screening samples that fail to amplify for COI with 16S provides a quick check for amplifiable DNA. When both COI and 16S fail, this likely reflects DNA degradation or the presence of PCR inhibitors.

Although we have not encountered any cases of failed amplification in fishes or other vertebrates, future instances may be resolved by adding a new primer to the existing cocktails. Amplification and sequencing protocols will remain unaffected so long as these new primers contain the same M13 tag and target the same amplicon. In the event that taxon-specific primer mismatches are detected, decisions on the nucleotide composition of new primers will be aided by the very large number of complete mitochondrial genomes available for fishes (Inoue *et al*. 2001; Miya *et al*. 2001; Miya *et al*. 2003). Cases of failed PCR amplification may also be resolved by employing degenerate or inosine-containing primers to overcome 3′-end mismatches (Batzer *et al*. 1991; Candrian *et al*. 1991; Christopherson *et al*. 1997; Sorenson *et al*. 1999). However, such primers may increase the chance of co-amplifying other gene regions (Zhang & Hewitt 1996) or segments of mitochondrial DNA that reside in the nucleus (NUMTs) (Lorenz *et al*. 2005). It is worth noting that NUMTs are rarely problematic in fishes (Bensasson *et al*. 2001; Richly & Leister 2004; Venkatesh *et al*. 2006), a conclusion reinforced in our study where we routinely recovered a single amplicon that showed the sequence characteristics expected of authentic COI.

Sequencing

Despite the high amplification success from all three primer sets, sequencing results were variable. Some reads were obscured with background signal, especially at the 5′-terminus. M13-tailed primer cocktails (COI-2, COI-3) consistently produced reads averaging 30 or more base pairs longer than those from untailed primers (COI-1). We suspect that conventional primers are more prone to form dimers, and, without clean up, these dimers are incorporated into sequencing reactions where they obscure the first 30–40 base calls at the 5′ end of the sequence. Longer reads are also likely a consequence of increasing amplicon size—M13 tags shift the read towards the 5′ end, delivering more overlap in the bidirectional reads that are a standard element of barcode analysis. More overlap means more reliable and longer sequence records in the reference database, but are also important for automating steps in the analytical

chain that otherwise require human intervention (e.g. automated sequence alignment and base calling).

In summary, the primer cocktails developed in this study are highly effective in generating amplicons that sequence cleanly for the DNA barcode region of diverse fish taxa and other groups of vertebrates.

Acknowledgements

This work was supported by grants to P.D.N.H. from the Gordon and Betty Moore Foundation, Genome Canada through the Ontario Genomics Institute, the Canada Foundation for Innovation, the Ontario Innovation Trust, the Canada Research Chairs Program and Natural Sciences and Engineering Research Council of Canada (NSERC). We thank Alex Borisenko for development of electronic laboratory books and comments on the manuscript. We also thank Nicolas Hubert, Peter Smith and Seinen Chow for providing samples for the test-plate.

References

Batzer MA, Carlton JE, Deininger PL (1991) Enhanced evolutionary PCR using oligonucleotides with inosine at the 3′-terminus. *Nucleic Acids Research*, **19**, 5081.

Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, **16**, 314–321.

Candrian U, Furrer B, Hofelein C, Luthy J (1991) Use of inosine-containing oligonucleotide primers for enzymatic amplification of different alleles of the gene coding for heat-stable toxin type I of enterotoxigenic *Escherichia coli*. *Applied Environmental Microbiology*, **57**, 955–961.

Christopherson C, Sninsky J, Kwok S (1997) The effects of internal primer-template mismatches on RT-PCR: HIV-1 model studies. *Nucleic Acids Research*, **25**, 654–658.

Clare EL, Lim BK, Engstrom MD, Eger JL, Hebert PDN (2007) DNA barcoding of Neotropical bats: species identification and discovery within Guyana. *Molecular Ecology Notes*, **7**, 184–190.

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.

Hall TA (1999) BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hanner RH, Schindel DE, Ward RD, Hebert PDN (2005) *FISH-BOL Workshop Report, August 26, 2005*. For the Workshop Held at the. University of Guelph, June 5–8, 2005. Ontario, Canada. Retrieved from http://www.fishbol.org/news.php on 1 August, 2006.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **270**, 313–322.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) 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.

Inoue JG, Miya M, Tsukamoto K, Nishida M (2001) A mitogenomic perspective on the basal teleostean phylogeny: resolving higher-level relationships with longer DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 275–285.

Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Lorenz JG, Jackson WE, Beck JC, Hanner R (2005) The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1869–1877.

Messing J (1983) New M13 vectors for cloning. *Methods in Enzymology*, **101**, 20–78.

Miya M, Kawaguchi A, Nishida M (2001) Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. *Molecular Biology and Evolution*, **18**, 1993–2009.

Miya M, Takeshima H, Endo H *et al.* (2003) Major patterns of higher teleostean phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **26**, 121–138.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics* (eds Hillis DM, Moritz C, Mable BK), pp. 205–247. Sinauer & Associates Inc., Sunderland, Massachusetts.

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

Rose TM, Schultz ER, Henikoff JG *et al.* (1988) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. *Nucleic Acids Research*, **26**, 1628–1635.

Sorenson MD, Ast JC, Dimcheff DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution*, **12**, 105–114.

Venkatesh B, Dandona N, Brenner S (2006) Fugu genome does not contain mitochondrial pseudogenes. *Genomics*, **87**, 307–310.

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.

Zhang DX, Hewitt GM (1996) Highly conserved nuclear copies of the mitochondrial control region in the desert locust *Schistocerca gregaria*: some implications for population studies. *Molecular Ecology*, **5**, 295–300.