Molecular Ecology Notes (2007) **7**, 730–734 doi: 10.1111/j.1471-8286.2007.01863.x

BARCODING ARTICLE

Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome *b* and nuclear rhodopsin genes

RAFAEL G. SEVILLA,* AMALIA DIEZ,* MICHAEL NORÉN,† OLIVIER MOUCHEL,‡ MARC JÉRÔME,‡ VÉRONIQUE VERREZ-BAGNIS,‡ HILDE VAN PELT,§ LAURENCE FAVRE-KREY,¶ GRIGORIOS KREY,¶ THE FISHTRACE CONSORTIUM†† and JOSÉ M. BAUTISTA*

**Department of Biochemistry and Molecular Biology IV, Universidad Complutense de Madrid (UCM), Facultad de Veterinaria, 28040 Madrid, Spain, †Swedish Museum of Natural History, Frescativaegen 40, PO Box 50007, SE-104 05, Stockholm, Sweden, ‡French Research Institute for the Exploitation of the Sea (IFREMER), Centre de Nantes Rue de l'île d'Yeu, 44311, Nantes cedex 3, France, §Netherlands Institute for Fisheries Research, Haringkade 1, 1970 AB IJmuiden, The Netherlands, ¶National Agricultural Research Foundation-Fisheries Research Institute, Nea Peramos, Kavala, GR-64007, Greece*

Abstract

This report describes a set of 21 polymerase chain reaction primers and amplification conditions developed to barcode practically any teleost fish species according to their mitochondrial cytochrome *b* and nuclear rhodopsin gene sequences. The method was successfully tested in more than 200 marine fish species comprising the main Actinopterygii family groups. When used in phylogenetic analyses, its combination of two genes with different evolutionary rates serves to identify fish at the species level. We provide a flow diagram indicating our validated polymerase chain reaction amplification conditions for barcoding and species identification applications as well as population structure or haplotyping analyses, adaptable to high-throughput analyses.

Keywords: DNA barcoding, fish, mitochondrial cytochrome *b*, PCR primers, rhodopsin, teleost

Received 6 January 2007; revision accepted 10 May 2007

Teleosts account for > 95% of the estimated 30 000 fish species alive today (Miya *et al*. 2003; Nelson 2006). The unequivocal identification and classification of living organisms to the species level frequently relies on genetic evidence. Specific DNA sequences act as unrepeatable signatures and therefore constitute a unique DNA barcode for each species. Initiatives, such as The Barcode of Life Database (http://www.barcodinglife.org) including The Fish Barcode of Life (http://www.fishbol.org), use a DNA-based identification system based on a relatively small fragment of the mitochondrial cytochrome *c* oxidase subunit I (COI). This short DNA sequence provides sufficient identification labels in terms of nucleotide positions (Hebert *et al*. 2003) to discriminate even between congeneric fish species, despite only 2% sequence divergence found in 98% of these species (Ward *et al*. 2005). It is nevertheless clear that longer length DNA barcodes will provide more efficient identification labels. Barcode efficiency can be further improved by the simultaneous use of two genes showing different evolutionary rates and genomic positions. The mitochondrial cytochrome *b* gene (cyt*b*) and the nuclear rhodopsin gene (rhod) fulfil these requirements. The cyt*b* gene, whose phylogenetic performance is comparable to that of COI (Zardoya & Meyer 1996), has been

Correspondence: José M. Baustista, Fax: +34913943824;
E-mail: jmbau@vet.ucm.es

††The FishTrace Consortium (www.fishtrace.org) comprises 53 members from the following institutions: University Complutense of Madrid; Joint Research Centre of the European Commission; Swedish Museum of Natural History; Canarian Institute of Marine Sciences; French Research Institute for the Exploitation of the Sea; Netherlands Institute for Fisheries Research; Natural History Museum of Funchal; Natural History Museum of Tenerife; Fisheries Research Institute of Kavala; and National Natural History Museum of Paris

Table 1 Fish-versatile primers. (A) Primer pairs for the amplification of mitochondrial cytochrome *b* (1141 bp). (B) Primers for cyt*b* sequencing purposes. Fish-seq and 7F-seq were, respectively, used for sequencing FishcytB-F and CytBI-7F amplification products. (C) Primers pairs for amplification of the targeted fragment in the rhodopsin nuclear gene (460 bp)

(A)

(a)	Name (b)	Sequence (5′–3′) (b)	Location (c)	Size (bp)	Percentage GC	T_m (°C) (d)
1	GluFish-F	AACCACCGTTGTTATTCAACTACAA	15329	25	36.0	57.7
2	FishcytB-F	ACCACCGTTGTTATTCAACTACAAGAAC	15330	28	39.3	60.7
3	CytBI-6F	TTCTCAGTAGACAACGCCACCCT	15862	23	52.2	61.0
4	CytBI-7F	CTAACCCGATTCTTTGCCTTCCACTTCCT	15883	29	48.3	68.3
5	CytBI-1F	CGATTCTTCGCATTCCACTTCCT	15889	23	47.8	62.5
6	*CytBI-5R*	*GGTCTTTGTAGGAGAAGTATGGGTGGAA*	16018	28	46.4	63.5
7	*CytBI-3R*	*GGGGTAAAGTTGTCTGGGTCTCC*	16111	23	56.5	60.9
8	*CytBI-2R*	*GCGGGGGTAAAGTTGTCTGGGTC*	16114	23	60.9	65.5
9	*CytBI-4R*	*AGGAAGTATCATTCGGGCTTAATATG*	16159	26	38.5	58.9
10	*TruccytB-R*	*CCGACTTCCGGATTACAAGACCG*	16528	23	56.5	64.6
11	*THR-Fish2-R*	*AACCTCCGACATCCGGCTTACAAGACCG*	16528	28	57.1	72.1
12	*THR-Fish-R*	*ACCTCCGATCTTCGGATTACAAGACC*	16529	26	50.0	64.4

(B)

(a)	Name	Sequence (5′–3′)	Location (c)	Size (bp)	Percentage GC	T_m (°C) (d)
13	Fish-seq	CCACCGTTGTTATTCAACTACAAG	15331	24	41.7	56.6
14	7F-seq	CTAACCCGATTCTTTGCCTTC	15883	21	47.6	56.7

(C)

(a)	Name (b)	Sequence (5′–3′) (b)	Location (e)	Size (bp)	Percentage GC	T_m (°C) (d)
15	RHO-30F:	CCNTAYGAYTAYCCNCARTAYTA	67	23	41.3	53.5
16	Rod-F2B:	GTCTGCAAGCCCATCAGCAACTTCCG	415	26	57.7	71.0
17	Rod-F2w:	AGCAACTTCCGCTTCGGTGAGAA	430	23	52.2	65.1
18	Rod-F2x:	AGCAACTTCCGCTTCGGCGAGAA	430	23	56.5	68.8
19	Rod-F2:	AGCAACTTCCGCTTCGGAGAGAA	430	23	52.2	64.4
20	*Rod-R4n*:	*GGAACTGCTTGTTCATGCAGATGTAGAT*	913	28	42.9	63.6
21	*Rod-4R*:	*CTGCTTGTTCATGCAGATGTAGAT*	913	24	41.7	57.2
22	*Rod-5R*:	*GGTGGTGATCATGCAGTGGCGGAA*	937	24	58.3	70.7
23	*RHO-319R*:	*TTNCCRCARCAYAANGTNGT*	955	20	45.0	66.6

(a): Numbers correspond to primers in Fig. 1.
(b): Reverse primers in italics.
(c): Nucleotide position corresponding to the 5′ position in the *Oncorhynchus mykiss* mitochondrial genome (GenBank Accession no: NC_001717). Locations given for the reverse primers are based on the reverse-complementary primer sequence position.
(d): T_m calculated using PRIMEREXPRESS 2.0 (Applied Biosystems).
(e): Nucleotide position corresponding to the 5′ position in the *Astyanax mexicanus* genomic rhodopsin gene (GenBank Accession no: U12328). Locations given for the reverse primers are based on the reverse-complementary primer sequence position.

widely used for identifying fish species and resolving fish phylogenies (Zardoya & Doadrio 1999; Farias *et al*. 2001; Chen *et al*. 2003; Dettai & Lecointre 2005). The intronless teleost fish rhod gene (Venkatesh *et al*. 1999) provides quantitatively equal interspecies identification labels of targeted nuclear polymerase chain reaction (PCR) amplification products throughout its coding sequence. This gene has also been used in fish phylogenetic studies (Chen *et al*. 2003). In addition, the nuclear and the mitochondrial genes serve mutually as an internal phylogenetic control to validate sequences obtained from a large number of samples.

Herein, we describe the use of 21 PCR primers capable of robustly and consistently amplifying targeted DNA sequences of practically any teleost fish species, and thus generating DNA sequence collections for species identification and phylogenetic purposes. The primers, 12 specific for cyt*b* (Table 1A) and nine for rhod (Table 1C), target regions of low variability, flanking the PCR amplification

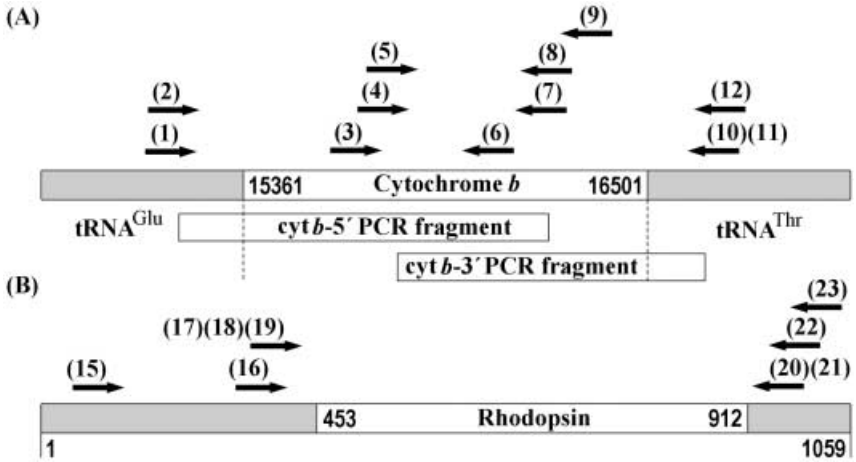

Fig. 1 Amplification schemes: targeted gene regions and primer positions. (A) Typical vertebrate cytochrome *b* organization indicating flanking genes (tRNA^{Glu}, tRNA^{Thr}) and their nucleotides positions (15361–16501) within the *Oncorhynchus mykiss* mitochondrial genome (GenBank Accession no: NC_001717). The relative lengths of targeted cyt*b*-5′-and cyt*b*-3′ PCR fragments are represented. Detailed information on represented primer pairs is given in Table 1A. (B) Rhodopsin amplification scheme. Targeted 460-bp length fragment and primer location given as that corresponding to the 5′ position in the *Astyanax mexicanus* rhodopsin gene (GenBank Accession no: U12328). Detailed information on the represented primer pairs is given in Table 1B.

areas of interest in the two genes (Fig. 1). The entire cyt*b* coding sequence (1141 bp) can be amplified in a single reaction. However, for improved amplification efficiencies two separate reactions, one for each of the 5′ (~750 bp) and the 3′ (~700 bp) fragments, with significant overlapping between the two regions (Fig. 1A), are suggested. The rhod-specific primers are used for single-step amplification of 460 bp in the gene coding sequence (Fig. 1B).

DNA from tissue samples (mostly white muscle) obtained from fish specimens was extracted using standard phenol–chloroform procedures (Sambrook *et al*. 1989), a DNA isolation station (ABI PRISM 6100 Nucleic Acid PrepStation; Applied Biosystems, Inc.), or commercial column kits (QIAGEN DNeasy Kit, QIAGEN DNeasy Tissue Kit and QIAamp DNA mini kit; QIAGEN GmbH). DNA quality was checked on 0.8% agarose gels and DNA concentration was determined using the PicoGreen DNA quantification kit (Molecular Probes) in a 96 multi-well microplate fluorometer reader, and a standard curve (0.2 ng/μL to 140 ng/μL).

Both genes were amplified in 25-μL reaction mixtures containing: 1 μL DNA template (concentration range 20 ng/μL to 30 ng/μL); 1× PCR buffer; 0.4 mM dNTPs; 2.5 mM $MgCl_2$; and 1.25 U *Taq* DNA polymerase. Forward and reverse primer concentrations were optimized and adjusted to 0.25 ng/μL for cyt*b* and 0.5 ng/μL for rhod. For improved amplification efficiencies and automated sequencing signal quality, a nested or seminested PCR step was found necessary in many cases. For nested and seminested PCR, 1–2 μL of the product from the first reaction was used as template for the subsequent amplification of targeted fragments. Table 2 provides alternative PCR protocols in case of amplification failure. Details of these protocols including 33 alternative amplification conditions can be found in Table S1, Supplementary material.

The final PCR products obtained were always of the expected length as determined by agarose gel electrophoresis. After purification, products were processed for sequencing using the same forward and reverse primers employed for amplification, except the the FishcytB-F and CytBI-7F amplification products that were sequenced using Fish-seq and 7F-seq, respectively (Table 1B). The optimal sequencing DNA concentration was estimated at 20 ng/μL. PCR products were bidirectionally sequenced using an ABI 3730 capillary sequencer.

Table 2 provides a flow diagram of the protocol options available using the designed primers. These protocols were tested on the complete list of species provided in FishTrace (www.fishtrace.org) including 1028 teleost specimens comprising 220 species from 17 Actinopterigii orders (Anguilliformes, Clupeiformes, Osmeriformes, Salmoniformes, Aulopiformes, Ophidiiformes, Gadiformes, Batrachoidiformes, Lophiiformes, Atheriniformes, Beloniformes, Beryciformes, Zeiformes, Scorpaeniformes, Perciformes, Pleuronectiformes and Tetraodontiformes), 75 families and 112 genera.

The first option (A1) in Table 2 was successful at amplifying either of the two cyt*b* fragments or the rhod fragment in > 60% of all barcoded FishTrace species. Approximately half of the remaining species were barcoded using option A2 and the rest using any of the other options. Overall, the above procedure yielded > 99.9%

(A)

Cyt*b*: consecutive options 1st reaction[1]			
A	B	C	D
FishcytB-F + TruccytB-R	GluFish-F + TruccytB-R	FishcytB-F + THR-Fish-R	GluFish-F + THR-Fish-R

Cyt*b*-5′: consecutive opt. 2nd reaction[2]	
1	2
FishcytB-F + CytBI-5R	FishcytB-F + CytBI-4R

Cyt*b*-3′: consecutive options 2nd reaction[2]			
1	2	3	4
CytbI-7F + THR-Fish-R	CytbI-7F + TruccytB-R	CytbI-6F + THR-Fish-R	CytbI-6F + TruccytB-R

(B)

Rhod: consecutive options 1st reaction[3]			
A	B	C	D
Rod-F2B + Rod-5R	Rod-F2B + Rod-5R	RHO-30F + RHO-319R	Rod-F2B + Rod-5R

Rhod: consecutive opt. 2nd reaction[4]		
1	2	3
Rod-F2W + Rod-R4n	Rod-F2X + Rod-R4n	Rod-F2 + Rod-R4n

PCR programs given as "temperature in °C - seconds" as follows: Initial Denaturation / (Denaturation / Annealing / Extension) x Number of Cycles / Final Extension: [1] 95-420 / (94-30 / 55-35 / 72-120) x35 / 72-420; [2] 95-420 / (94-30 / 55-35 / 72-45) x38 / 72-420; [3] 95-420 / (94-30 / 62-30 / 72-30) x40 / 72-420; [4] 95-420 / (94-30 / 56-30 / 72-30) x40 / 72-420.

Table 2 Flow diagram providing alternative PCR protocols for amplification of the target fragments of the cyt*b* and rhodopsin genes. Nested or seminested PCRs comprise two reactions. The first reaction uses a pair of *outer* primers (designated A to D in both genes), which in cyt*b* flank the whole gene. The second reaction has specific *inner* primers (numbered) for each gene fragment. Preferential reactions are designated with a letter followed by a number e.g. for cyt*b*-5′-A1 is the first choice and B1 the third choice. (A) Preferential PCR conditions (primer pair and thermocycling program) for the amplification of cyt*b* as two PCR fragments: cyt*b*-5′- and cyt*b*-3′. (B) Preferential PCR conditions for the amplification of the targeted fragment of the rhod gene

successful amplifications. After validating the sequencing data, approximately 3% did not match the expected phylogeny, mainly because of sampling (misidentified specimens) or amplification-sequencing errors. In these cases, repeating the procedure using newly extracted DNA or new samples was sufficient to successfully amplify and sequence the target genes.

In conclusion, the protocol proposed is a powerful tool for barcoding practically all teleost fish species and was successfully used here to provide fully validated sequence data for the FishTrace genetic catalogue (www.fishtrace.org).

Acknowledgements

We thank S. Pérez-Benavente for skilful technical assistance. Financial support was provided by FEDER programme and MCyT-Spain (1FD97-1235-C04 MAR and CAL01-020-C3) and European Commission (FishTrace contract, QLRI-CT-2002–02755).

References

Chen WJ, Bonillo C, Lecointre G (2003) Repeatability of clades as a criterion of reliability: a case study for molecular phylogeny of Acanthomorpha (Teleostei) with larger number of taxa. *Molecular Phylogenetics and Evolution*, **26**, 262–288.

Dettai A, Lecointre G (2005) Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. *Comptes Rendus Biologies*, **328**, 674–689.

Farias IP, Orti G, Sampaio I, Schneider H, Meyer A (2001) The cytochrome *b* gene as a phylogenetic marker: the limits of resolution for analyzing relationships among cichlid fishes. *Journal of Molecular Evolution*, **53**, 89–103.

Hebert PD, Ratnasingham S, deWaard JR (2003) Barcoding animal life: cytochrome *c* oxidase subunit I divergences among closely related species. *Proceedings. Biological Sciences / the Royal Society*, **270** (Suppl. 1), S96–S99.

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.

Nelson JS (2006) *Fishes of the World*, 4th edn. John Wiley and Sons, Inc, New York.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Venkatesh B, Ning Y, Brenner S (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. *Proceedings of the National Academy of Sciences, USA*, **96**, 10267–10271.

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

Zardoya R, Doadrio I (1999) Molecular evidence on the evolutionary and biogeographical patterns of European cyprinids. *Journal of Molecular Evolution*, **49**, 227–237.

Zardoya R, Meyer A (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, **13**, 933–942.

Supplementary Material

The following supplementary material is available for this article:

Table S1. PCR conditions detailing direct and nested amplifications, and alternative strategies for fish DNA barcoding. (A) Cytochrome *b*. (B) Rhodopsin 460-bp length fragment.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1471-8286.2007.01863.x
(This link will take you to the article abstract).

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