

DNA barcodes identify marine fishes of São Paulo State, Brazil

AMANDA DE OLIVEIRA RIBEIRO,* RODRIGO ANTUNES CAIRES,† TATIANE CASAGRANDE MARIGUELA,* LUIZ HENRIQUE GARCIA PEREIRA,* ROBERT HANNER‡ and CLAUDIO OLIVEIRA*
*Laboratório de Biologia e Genética de Peixes, Instituto de Biociências, Universidade Estadual Paulista (UNESP), Botucatu, 18618-970, SP, Brazil, †Museu de Zoologia, Universidade de São Paulo (MUZUSP), São Paulo, 04263-000, SP, Brazil, ‡Department of Integrative Biology, Biodiversity Institute of Ontario, University of Guelph, Guelph, ON, Canada, N1G 2W1

Abstract

Anthropogenic impacts are an increasing threat to the diversity of fishes, especially in areas around large urban centres, and many effective conservation actions depend on accurate species identification. Considering the utility of DNA barcoding as a global system for species identification and discovery, this study aims to assemble a DNA barcode reference sequence library for marine fishes from the coastal region of São Paulo State, Brazil. The standard 652 bp 'barcode' fragment of the cytochrome *c* oxidase subunit I (COI) gene was PCR amplified and bidirectionally sequenced from 678 individuals belonging to 135 species. A neighbour-joining analysis revealed that this approach can unambiguously discriminate 97% of the species surveyed. Most species exhibited low intraspecific genetic distances (0.31%), about 43-fold less than the distance among species within a genus. Four species showed higher intraspecific divergences ranging from 2.2% to 7.6%, suggesting overlooked diversity. Notably, just one species-pair exhibited barcode divergences of <1%. This library is a first step to better know the molecular diversity of marine fish species from São Paulo, providing a basis for further studies of this fauna – extending the ability to identify these species from all life stages and even fragmentary remains, setting the stage for a better understanding of interactions among species, calibrating the estimations about species composition and richness in an ecosystem, and providing tools for authenticating bioproducts and monitoring illegal species exploitation.

Keywords: biodiversity, cytochrome *c* oxidase subunit I, mitochondrial DNA, species identification

Received 16 May 2012; revision received 15 July 2012; accepted 2 August 2012

Introduction

Species are the fundamental units of organismal biology and, although the species concept has not achieved a consensus, the activity of associating valid scientific names with groups of individuals is crucial to define priorities in biodiversity conservation (Balakrishnan 2005). Species description and identification are traditionally based on morphological traits. Nevertheless, the morphological approach alone has intrinsic limitations: phenotypic plasticity and genetic variability may mask the characters employed; cryptic *taxa* cannot be easily detected; identification keys typically do not treat all life stages and genders and the keys themselves may be very challenging to use, making the system heavily dependent on experts (Hebert *et al.* 2003b).

Over the last several decades, varied molecular tools have been used to augment morphological approaches to species identification and discovery. More recently, DNA-based methods have gained popularity because DNA is relatively stable, may be accessed from all life stages using tiny amounts of tissue, and because DNA sequences are highly reproducible (Ward *et al.* 2009). The current reduction in global diversity demands accelerating the rate at which knowledge about biodiversity is obtained (Savage 1995) and, although not intended to replace traditional morphological taxonomy, molecular methods arise as a new column to sustain an 'integrative taxonomy' approach (Padiál *et al.* 2010).

Despite their utility, the plethora of molecular markers employed by different researchers has hindered the development of an unifying molecular identification system. Thus, Hebert *et al.* (2003a,b) proposed the use of a standardized marker consisting of approximately 650 base pairs from the 5' end of the mitochondrial gene Cytochrome *c* oxidase subunit I (COI) as a universal

Correspondence: Amanda de Oliveira Ribeiro, Fax: 55-14-3880-0499; E-mail: aolribeiro@yahoo.com.br

identification system suitable for most animal species. For this gene fragment, species typically exhibit a unique sequence or a tight cluster of very similar sequences deeply divergent from those of other congeneric species, collectively known as 'DNA barcodes' which can be used for the identification of known species and the discovery of new ones. As COI sequences tend to vary among species, but are relatively constant within the species, an unknown specimen or sample – including eggs, larvae, juveniles and fragmented remains – can be identified by matching its sequence against a reference sequence library derived from expert-identified voucher specimens. The effectiveness of the barcode fragment for species identification has been tested in diverse animal groups, and its success owes to the adherence to data standards (reviewed in Teletchea 2010).

Once reference libraries are completed for a certain group, they facilitate further studies in molecular ecology and wildlife forensics. The ability to identify species from all life stages (Pegg *et al.* 2006; Valdez-Moreno *et al.* 2010) and even from fragmentary remains (Holmes *et al.* 2009) sets the stage for a better understanding of interactions among species and may be helpful to calibrate estimations about species composition and richness in an ecosystem (Valentini *et al.* 2009). Barcode libraries also help authenticate commercial products and aid in detecting intentional or unintentional frauds involving bio-products labelling (Wong & Hanner 2008; Barbuto *et al.*

2010) and may provide a tool for monitoring illegal species exploitation (Baker 2008).

Coastal zones in Brazil have experienced a strong process of urbanization after the industrial expansion in the mid-1950s. Large industrial and urban centres, like São Paulo State, have been showing signs of impacts and degradation, especially because of industrial and domestic residues, tourism activity and fisheries (Diegues 1999). To better evaluate these impacts on the stocks of marine fishes from the São Paulo State and to build a solid knowledge base for conservation initiatives, an accurate identification system for these species is crucial. Considering the promising use of DNA barcoding to identify species, this study aims to generate a comprehensive barcode reference library for marine fishes from the São Paulo coast and to calibrate the effectiveness of this tool for unambiguous species identification.

Material and methods

Specimen collection

For this study, a total of 678 fish specimens were obtained by performing bottom otter trawls or acquired in local markets or fishery landings at the south, north and central areas of the São Paulo State coast, or obtained from collaborator researchers (Fig. 1). A fin or muscle tissue fragment was removed from each individual and

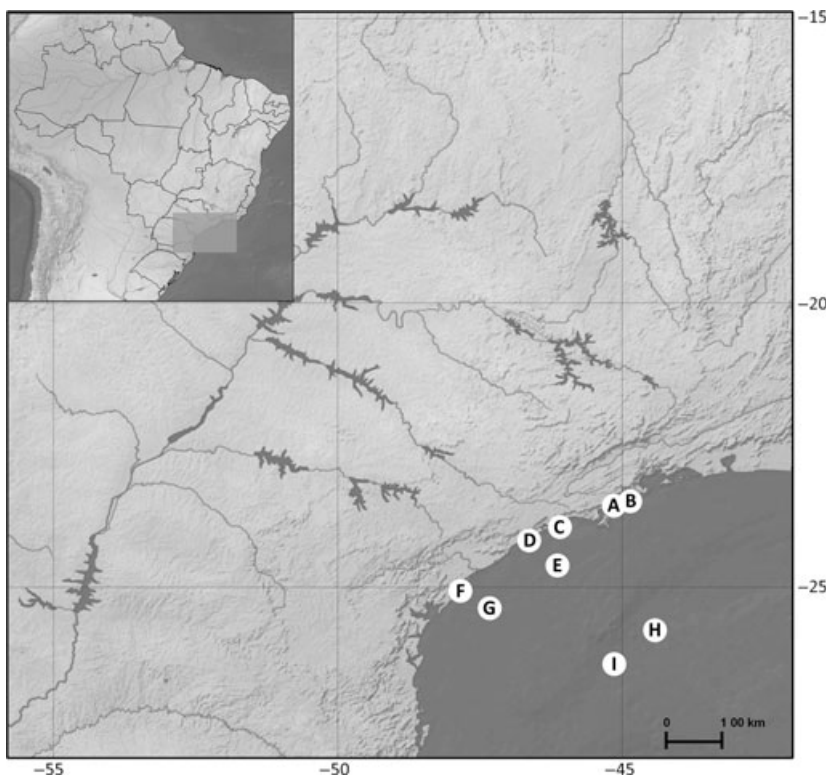


Fig. 1 Distribution of the collecting points in the São Paulo state area.

preserved in 95% ethanol at -20°C for molecular assays. All specimens were identified based on the morphological features, using identification keys and manuals (Figueiredo 1977; Carpenter 2002a,b,c), by a collaborator taxonomist or by the experts who provided some of the samples and, whenever possible, were deposited in the fish collection of Laboratório de Biologia e Genética de Peixes (LBP), UNESP, Botucatu, São Paulo, Brazil. Because of the restriction on the storage of large fishes, some specimens were not retained, but digital images or e-vouchers were kept following the FISH-BOL collaborators protocol (Steinke & Hanner 2010). Specimen provenance data, including information about the collection site, and specimen taxonomy are recorded in the BOLD project 'Marine Fishes from São Paulo, Brazil – Part 1' (project code: MFSP1) following established best practices for fish barcoding (Hubert *et al.* 2008; Ward *et al.* 2009). Sequences have also been deposited with GenBank (Appendix S1, Supporting Information).

In an attempt to take intraspecific variation into account, multiple specimens (mean = 5 specimens per species) were analysed for most species (119/135). However, species represented by a single specimen (16/135) were included in this analysis as a first step to understand their genetic identity.

Extraction, PCR amplification and DNA sequencing

Molecular assays were carried out at the Laboratório de Biologia e Genética de Peixes, UNESP, Brazil and at the Canadian Centre for DNA Barcoding (CCDB), Canada, using standard barcoding protocols (Hajibabaei *et al.* 2005). Total genomic DNA was isolated from each tissue sample using the DNeasy Tissue Kit (Qiagen, Hilden, Germany), salt extraction (Aljanabi & Martinez 1997), or through an automated glass fibre protocol (Ivanova *et al.* 2006).

A 652 bp fragment of the 5' end of the mitochondrial COI gene was amplified by PCR using different primer combinations: FishF1, FishR1, FishF2, FishR2 (Ward *et al.* 2005); the M13-tailed primer cocktails C_FishF1t1–C_FishR1t1 and C_VF1LFt1–C_VR1LRt1 (Ivanova *et al.* 2007); or L5698-Asn (Miya & Nishida 2000) and H7271-Coi (Melo *et al.* 2011). The Fish primer set was the first one to be used; if they failed, the cocktail primers were tested, and in case of suspected amplification of pseudogenes, the last pair of primers, targeting a larger area of the mitochondrial genome, was used. PCR mixes were based on Steinke & Hanner (2010), with minor adjustments according to the species being studied and the DNA polymerase used. Reactions were carried out in a thermocycler (Veriti® 96-well Thermal Cycler, Applied Biosystems™ or Mastercycler® EPGradient, Eppendorf), under the following general thermal conditions:

denaturation at 95°C , annealing at $48\text{--}56^{\circ}\text{C}$ (according to species and primer) and extension at 72°C .

Amplification success was checked via gel electrophoresis using Blue Green Loading Dye I (LGC Biotechnology) in 1% agarose gels or on precast agarose gels (E-gel®96; Invitrogen™). Amplified PCR products were purified with an ExoSap-IT® (USB Corporation) solution. Purified PCR products were labelled with the BigDye Terminator v3.1 Cycle Sequencing Ready Reaction kit (ABI). Labelled products were purified with ethylenediamine tetraacetic acid/sodium acetate/alcohol precipitation or with Sephadex® G-50 (Sigma-Aldrich), and then bidirectionally sequenced on automated DNA sequencers: 4-capillary ABI 3130 Genetic Analyzer (Applied Biosystems™), 8-capillary ABI 3500 Genetic Analyzer, or 48-capillary ABI 3730 DNA Analyzer.

Data analysis

Bidirectional sequence 'contigs' were assembled, and multiple sequence alignments were constructed using SEQSCAPE v2.6 (Applied Biosystems™), SEQUENCHER 4.8 (Gene Codes Corporation) or CODONCODE ALIGNER (Codon Code Corporation) to obtain consensus sequences, and BIOEDIT SEQUENCE ALIGNMENT EDITOR v7.0.5.3 (Hall 1999) was used to check them for indels or stop codons. Genetic distances among and within species were calculated using the Kimura 2-parameter (K2P) substitution model (Kimura 1980) and the divergence patterns suggested by those distances were graphically represented using a neighbour-joining (NJ) 1000 replicates bootstrap tree using MEGA v5.0 (Tamura *et al.* 2011). Sequence alignments were carried out using the ClustalW algorithm available on DAMBE software (Xia & Xie 2001) or with BOLD tools (Ratnasingham & Hebert 2007). The average base composition of sequences and the nearest-neighbour distance analysis were carried out using BOLD tools (Ratnasingham & Hebert 2007).

Results

Partial COI sequences were obtained for 678 fish specimens belonging to 62 families, 110 genera and 135 species (Appendix S1, Supporting Information). Sequence lengths averaged 649 bp (ranging from 457 to 652 bp), and 98.5% of them were longer than 600 bp. No indels or stop codons (that could suggest the amplification of NUMTs) were detected, and the sequences had <1% of ambiguous bases. Overall mean nucleotide frequencies were G (18.63%), C (28.59%), A (23.48%), T (29.12%).

Nearly all of the species (97%) could be discriminated by the barcode approach. The neighbour-joining tree (Figs 2 and S1, Supporting Information) shows that specimens assigned as conspecifics clustered in cohesive

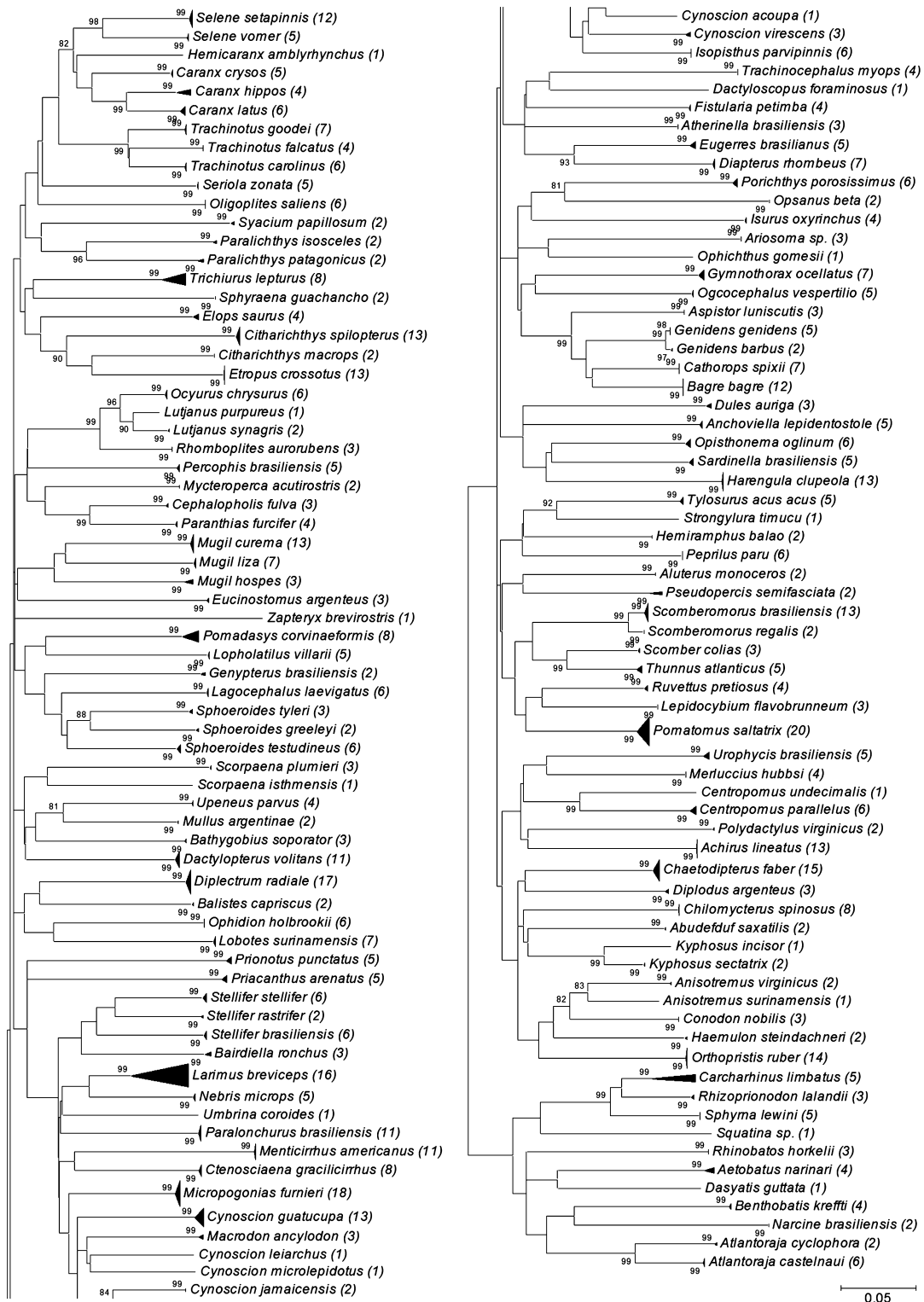


Fig. 2 Condensed neighbour-joining tree showing genetic distances (K2P) among 135 marine fish species from the São Paulo State area, Brazil. The number of fish specimens analysed for each species is between brackets.

Table 1 K2P genetic divergence values within different taxonomic levels for 678 specimens from the São Paulo State coast

	N	Taxa	Comparisons	K2P genetic divergence (%)			
				Minimum	Mean	Maximum	SE
Within species	657	118	2429	0	0.312	7.717 [†]	0.015
Within genus	670	101	696	0.617 [*]	13.295	20.725	0.167
Within family	645	57	8973	5.158	19.008	27.936	0.03

*Due to *Genidens barbatus* – *G. genidens* pair of species

[†]Due to *Larimus breviceps* species.

units. Mean K2P distance within species (0.31%) was about 43-fold less than the distance among congeners (13.29%), with genetic distance increasing further at higher taxonomic levels (Table 1). The total number of sequences presented here represents 136 barcode clusters or BINs (Barcode Index Numbers; BOLD 3.0) – 69 of them are taxonomically concordant with the other records from BOLD clustered in the same BIN, while 58 are discordant, in most cases at the species level, and 9 of them represent singletons.

Nearest-neighbour distance analysis revealed one pair of species with K2P distances <2.0% divergent from one another, but still divergent enough to discriminate each species as a molecular unit (*Genidens barbatus* and *Genidens genidens*; see Fig. 3).

Among the species with two or more specimens analysed, four showed intraspecific divergences higher than 2.0% (*Carcharhinus limbatus*, *Pomadasys corvinaeformis*, *Larimus breviceps* and *Trichiurus lepturus*), which were split into subclusters (Fig. 4). For *Carcharhinus limbatus*, one specimen (LBPV53117) diverged from the other four by 5.7%, whereas the K2P genetic distance within the subcluster composed of the remaining specimens was null. One specimen of *Larimus breviceps* (LBPV51368) was 7.6% divergent from the others, members of a cluster with only 0.1% internal divergence. One individual of *Pomadasys corvinaeformis* (LBPV51404) split from its conspecifics by 2.2%, whereas the distance among the remaining specimens in the subgroup was null. Finally,



Fig. 3 Neighbour-joining tree of COI sequences from two fish species showing interspecific divergences (K2P) lower than 2.0%.

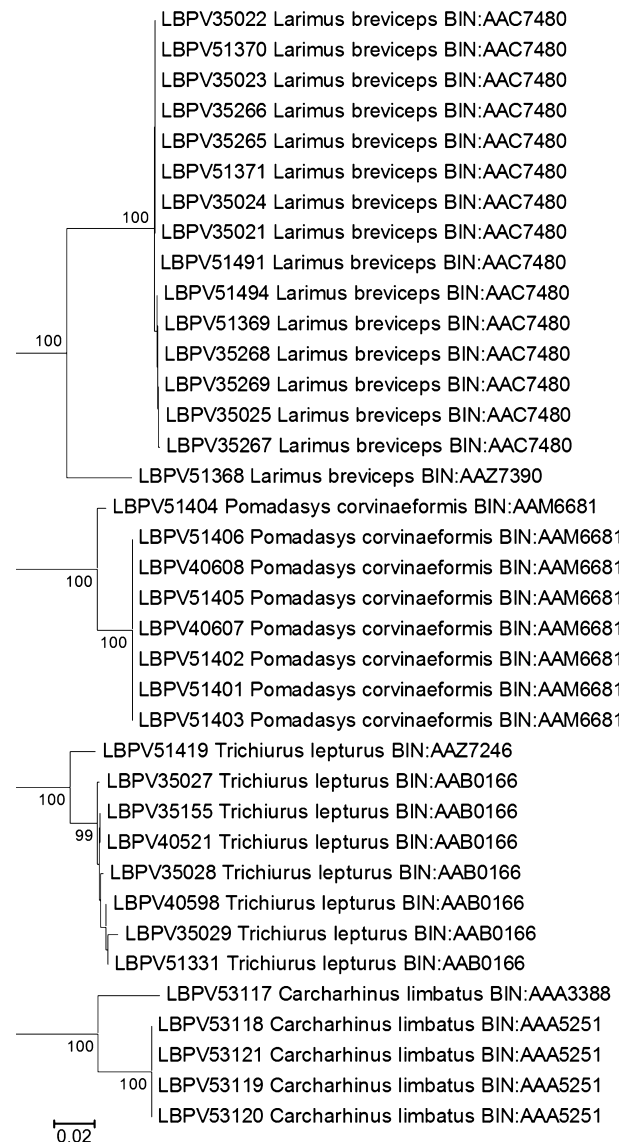


Fig. 4 Neighbour-joining tree of 37 COI sequences from four fish species showing intraspecific divergences (K2P) higher than 2.0%.

one specimen of *Trichiurus lepturus* (LBPV51419) had a genetic distance of 3.0% relative to the other members of the group, which exhibited only 0.4% divergence among

them. These values may suggest the existence of cryptic species or morphological misidentification; hence, these voucher specimens were re-examined. However, the outlier individuals maintained their initial morphological designation, indicating deep divergent lineages or hidden diversity in those groups.

The best nucleotide substitution model was calculated for both groups with low interspecific and deep intraspecific genetic distances, using MEGA 5.0 tools (data not shown). However, no significant differences between the distances calculated using these models and the K2P model were found.

Discussion

Regarding fishes, erroneous or imprecise identification of species is a significant problem for fisheries management (Lleonart *et al.* 2006) and morphologically exclusive approaches have limitations for identifying juveniles, damaged or fragmentary remains and cryptic *taxa*. The present investigation demonstrates that DNA barcoding can be used to discriminate among marine fishes from São Paulo and also contributes valuable reference data to the International Barcoding of Life (iBOL.org) project's Fish Barcode of Life (FISH-BOL) campaign via the Barcode of Life Data Systems (BOLD), yielding a barcode library of 135 marine fish species from the São Paulo State, Brazil, which represents about 25% of all species estimated for this area (Menezes 2011). From a zoogeographic point of view, the São Paulo State region represents a transactional zone where tropical and cold-water species coexist (Palacio 1982), resulting in a high level of fish biodiversity – encompassing almost half of the more than 1200 species estimated for the Brazilian marine fishes (Menezes *et al.* 2003).

Nearly all of the species could be discriminated by their divergent COI genetic distances. The K2P mean distance within species observed here (0.31%) is in accordance with that of other comprehensive analysis of fishes from Antarctica, 0.38% (Rock *et al.* 2008), Australia, 0.39% (Ward *et al.* 2005), China, 0.32% (Zhang 2011), India, 0.3% (Lakra *et al.* 2011), and Mexico, 0.31% (Valdez-Moreno *et al.* 2010), but slightly higher than for fishes from Argentina, 0.23% (Mabragaña *et al.* 2011), and Pacific Canada, 0.25% (Steinke *et al.* 2009a), and slightly lower than that for ornamental fish from Indian and Pacific Oceans, 0.42% (Steinke *et al.* 2009b). Nevertheless, in this study, K2P mean distance within genera (13.29%) was close only to the value found for the Mexican fishes (13.88%), being lower than the value of the Chinese study (15.74%) and higher than the values found for all the other data sets (ranging from 3.75% to 10.81%). Genetic mean distance increased with increasing taxonomic levels, supporting the idea of a significant change

in genetic divergence at the species boundaries (Hebert *et al.* 2003b).

Nearest-neighbour distance analysis revealed that only two of the species were <1.0% divergent from their nearest neighbour – the species *Genidens genidens* diverged from *G. barbatus* by only 0.62%. The other 98.5% of the species ranged from 4.14% to 23.63% divergent from their nearest neighbour. Although the distance between *G. genidens* and *G. barbatus* is short, morphological and cytogenetic evidences suggest that they are separate units. Among the individuals belonging to the two species, there are clear differences on their occipital processes and predorsal plates, on the shape of their premaxillary and palatine bones and on their vomerine teeth disposition, all diagnostic features for marine catfishes (Marceniuk & Menezes 2007) and, although the karyotype of both species show 56 chromosomes, they are structurally different and only *G. barbatus* individuals carry sexual chromosomes (Gomes *et al.* 1994; Szczepanski *et al.* 2010). Such evidences, combined with the presence of nucleotide substitutions in four positions along the barcode fragment of these species (Fig. S2, Supporting Information), all of them at the third codon position and silent, point the existence of two separated species with low interspecific divergence between them, suggesting recent separation.

Four species (*Carcharhinus limbatus*, *Pomadasys corvinaeformis*, *Larimus breviceps* and *Trichiurus lepturus*) showed markedly deeper COI variation, ranging from 2.2% to 7.6%, yet in each case, only one specimen (Fig. 4) split from the main group of its conspecifics. Mean K2P distance within the main groups in each of these species ranged from 0.0% to 0.5%, in accordance with typical intraspecific divergence values. There is no evidence supporting allopatric population divergence, because one or more individuals from the main group were also caught at the same location as the divergent specimens. Thus, the possibility of overlooked diversity needs to be further investigated. Two different BINs are linked to each one of the internal groups formed inside *C. limbatus*, *L. breviceps* and *T. lepturus*. For *Carcharhinus limbatus*, the BIN linked to the divergent specimen (AAA3388) and the one for the rest of the group (AAA5251) were both attributed to at least other five species belonging to the genus *Carcharhinus* by researchers from several parts of the world, indicating a high degree of difficulty in accurately identifying species belonging to this genus and/or possible gene flow between them. Also, previous studies revealed low divergences between *C. limbatus* and other congeneric species, especially *C. tilstoni* (Ward *et al.* 2008; Wong *et al.* 2009). Both BINs linked to *Larimus breviceps* were attributed only to this species, with the BIN associated with the divergent specimen (AAZ7390) being comprised of only this singleton. The same was

observed for *Trichiurus lepturus*: both BINs refer to this species and the divergent BIN (AAZ7246) is comprised of the singleton from this study. Studies from other regions involving the globally distributed hairtail fish *T. lepturus* suggest a high degree of intraspecific divergence exists within this species and additional mitochondrial data support the idea that *T. lepturus* represents a 'species complex' that may consist of three species (Hsu *et al.* 2009), indicating that this group could be a good target for a taxonomic revision.

The present study supports the idea that the DNA barcoding through part of the mitochondrial gene COI is an adequate tool to discriminate most species of fishes. DNA barcoding was proposed as a global system for animal identification (Hebert *et al.* 2003b) because of its evolutionary rate, absence of indels, limited occurrence of recombination and because of the existence of robust primers capable of recovering the barcode from diverse taxa. This study and others like this demonstrate the success of barcoding to accurately discriminate a great number of species. Interestingly, Lane (2009) suggests a reason deeper than coincidence or convenience for the performance of the DNA barcode based on COI – suggesting that mitochondrial genes might actually be powerful drivers in the process of speciation. Because ATP generation is critical to fitness and derives from both nuclear and mitochondrial genomes, the tight co-adaptation of the subunits encoded by mtDNA and nDNA may participate in the formation of reproductive barriers (Gershoni *et al.* 2009), suggesting a mechanistic basis for the efficacy of barcoding in delineating species. Notably, Puslednik *et al.* (2012) showed that mitochondrial COI amino acid residues that interact with amino acid residues from proteins of the oxidative phosphorylation system encoded by the nuclear genome have a faster evolutionary rate than those COI amino acids that interact with proteins encoded by other mitochondrial genes or those which do not interact with other proteins. This co-evolution between the mitochondrial and nuclear Cytochrome *c* oxidase subunits is not accounted for using simple models of nucleotide sequence evolution. The Kimura 2-parameter model, for example, assumes different substitution rates only for transitions and transversions (Kimura 1980), while a model that considers other sources of variation may be a better choice. However, our results using different substitution models were equivocal, perhaps because the actual number of interacting residues is relatively limited.

In summary, this study corroborates the suitability of the DNA barcode approach to discriminate and unambiguously identify most species of marine fishes examined to date. Patterns of COI divergence consistently matched morphologically recognized species, although

in a few cases, molecular data revealed hidden genetic divergences within a group and one case of low interspecific variation. As cryptic taxa are relatively common among marine animals, the possibility of overlooked diversity and the occurrence of species complexes have to be considered. For the purpose of the Barcoding of Life Initiative, the more complete the barcode library, the more efficient and useful it is (Ekrem *et al.* 2007). So, this kind of contribution is important to consolidate the DNA barcode as a global system of identification of life and it is a first step to better understand the genetic diversity of the Brazilian marine fishes. However, it is clear that even among expert taxonomists, the consistent application of names proves challenging, particularly when compounded with the presence of cryptic diversity, as evidenced by the conflicting names applied to specimens within the same BIN by different research groups. As more data accumulate globally, barcoding and BINs will play an important role in the integrative taxonomic feedback loop and will contribute significantly to standardizing the application of names internationally, which is of extreme significance for the sustainable management of the world's fisheries.

Acknowledgements

We thank Renato Devidé for supporting the collection of samples; Otto Bismarck Gadig, Fernando Mendonça, Renato Hajenius Freitas, Bruno Barbanti, and Alexandre Marceniuk for samples concession; and all BOLD/CCDB team for their technical support. Financial support for this study was provided by CNPq and FAPESP. We also acknowledge the Government of Canada for funding to the Biodiversity Institute of Ontario from Genome Canada (through the Ontario Genomics Institute) for the International Barcode of Life Project.

Conflicts of interest

The authors have no conflict of interest to declare. Institutions supporting this research had no influence in study design, data collection and analysis, or preparation of the manuscript.

References

- Aljanabi SM, Martinez I (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, **25**, 4692–4693.
- Baker CS (2008) A truer measure of the market: the molecular ecology of fisheries and wildlife trade. *Molecular Ecology*, **17**, 3985–3998.
- Balakrishnan R (2005) Species concepts, species boundaries and species identification: a view from the tropics. *Systematic Biology*, **54**, 689–693.
- Barbuto M, Galimberti A, Ferri E *et al.* (2010) DNA barcoding reveals fraudulent substitutions in shark seafood products: the Italian case of "palombo" (*Mustelus* spp.). *Food Research International*, **43**, 376–381.

- Carpenter KE (2002a) *The Living Marine Resources of the Eastern Central Atlantic. Volume 3: Bony Fishes part 2 (Opistognathidae to Molidae), Sea Turtles and Marine Mammals*. FAO, Rome.
- Carpenter KE (2002b) *The Living Marine Resources of the Western Central Atlantic. Volume 1: Introduction, Molluscs, Crustaceans, Hagfishes, Sharks, Batoid Fishes, and Chimaeras*. FAO, Rome.
- Carpenter KE (2002c) *The Living Marine Resources of the Western Central Atlantic. Volume 2: Bony Fishes Part 1 (Acipenseridae to Grammatidae)*. FAO, Rome.
- Diegues AC (1999) Human populations and coastal wetlands: conservation and management in Brazil. *Ocean & Coastal Management*, **42**, 187–201.
- Ekrem T, Willassen E, Stur E (2007) A comprehensive DNA sequence library is essential for identification with DNA barcodes. *Molecular Phylogenetics and Evolution*, **43**, 530–542.
- Figueiredo JL (1977) *Manual de peixes marinhos do Sudeste do Brasil*. Museu de Zoologia da Universidade de São Paulo, São Paulo.
- Gershoni M, Templeton AR, Mishmar D (2009) Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays*, **31**, 642–650.
- Gomes V, Phan VN, Passos MJDACR (1994) Karyotypes of three species of marine catfishes from Brazil. *Boletim do Instituto Oceanográfico*, **42**, 55–61.
- 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.
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London, Series 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. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**(Suppl.), S96–S99.
- Holmes B, Steinke D, Ward R (2009) Identification of shark and ray fins using DNA barcoding. *Fisheries Research*, **95**, 280–288.
- Hsu K, Shih N, Ni I, Shao K (2009) Speciation and population structure of three *Trichiurus* species based on mitochondrial DNA. *Zoological Studies*, **48**, 835–849.
- Hubert N, Hanner R, Holm E *et al.* (2008) Identifying Canadian Freshwater Fishes through DNA Barcodes. *PLoS ONE*, **3**, e2490.
- Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.
- Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.
- Lakra WS, Verma MS, Goswami M *et al.* (2011) DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, **11**, 60–71.
- Lane N (2009) On the origin of barcodes. *Nature*, **462**, 272–274.
- Leonart J, Taconet M, Lamboeuf M (2006) Integrating information on marine species identification for fishery purposes. *Marine Ecology Progress Series*, **316**, 231–238.
- Mabragaña E, Dias de Astarloa JM, Hanner R, Zhang J, González-Castro M (2011) DNA barcoding identifies Argentine fishes from marine and brackish waters. *PLoS ONE*, **6**, e28655.
- Marceñiuk AP, Menezes NA (2007) Systematics of the family Ariidae (Ostariophysi, Siluriformes), with a redefinition of the genera. *Zootaxa*, **1416**, 1–126.
- Melo BF, Benine RC, Mariguela TC, Oliveira C (2011) A new species of *Tetragonopterus* Cuvier, 1816 (Characiformes: Characidae: Tetragonopterinae) from the rio Jari, Amapá, northern Brazil. *Neotropical Ichthyology*, **9**, 49–56.
- Menezes NA (2011) Checklist dos peixes marinhos do estado de São Paulo, Brasil. *Biota Neotropica*, **11**, 1–15.
- Menezes NA, Buckup PA, Figueiredo JL, Moura RL (2003) *Catálogo das espécies de peixes marinhos do Brasil*. Museu de Zoologia da USP, São Paulo.
- Miya M, Nishida M (2000) Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Molecular Phylogenetics and Evolution*, **17**, 437–455.
- Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. *Frontiers in Zoology*, **7**, 1–14.
- Palacio FJ (1982) Revisión zoogeográfica marina del sur del Brasil. *Boletim do Instituto Oceanográfico*, **31**, 69–92.
- Pegg GG, Sinclair B, Briskey L, Aspden WJ (2006) MtDNA barcode identification of fish larvae in the southern Great Barrier Reef, Australia. *Scientia Marina*, **70S2**, 7–12.
- Puslednik L, Yeates DK, Faith DP, Ballard JWO (2012) Protein-protein interactions of the cytochrome c oxidase DNA barcoding region. *Systematic Entomology*, **37**, 229–236.
- Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.
- Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, Carvalho GR (2008) DNA barcodes of fish of the Scotia Sea, Antarctica indicate priority groups for taxonomic and systematics focus. *Antarctic Science*, **20**, 253–262.
- Savage JM (1995) Systematics and the biodiversity crisis. *BioScience*, **45**, 673–679.
- Sczepanski TS, Noleto RB, Cestari MM, Artoni RF (2010) A comparative study of two marine catfish (Siluriformes, Ariidae): cytogenetic tools for determining cytotaxonomy and karyotype evolution. *Micron*, **41**, 193–197.
- Steinke D, Hanner R (2010) The FISH-BOL collaborators' protocol. *Mitochondrial DNA*, **22**(Suppl. 1), 10–14.
- Steinke D, Zemlak TS, Boutillier JA, Hebert PDN (2009a) DNA barcoding of Pacific Canada's fishes. *Marine Biology*, **156**, 2641–2647.
- Steinke D, Zemlak TS, Hebert PDN (2009b) Barcoding nemo: DNA-based identifications for the ornamental fish trade. *PLoS ONE*, **4**, e6300.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.
- Teletchea F (2010) After 7 years and 1000 citations: comparative assessment of the DNA barcoding and the DNA taxonomy proposals for taxonomists and non-taxonomists. *Mitochondrial DNA*, **21**, 206–226.
- Valdez-Moreno M, Vásquez-Yeomans L, Elías-Gutiérrez M, Ivanova NV, Hebert PDN (2010) Using DNA barcodes to connect adults and early life stages of marine fishes from the Yucatan Peninsula, Mexico: potential in fisheries management. *Marine and Freshwater Research*, **61**, 665–671.
- Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**, 110–117.
- 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.
- 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 PD (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.
- Wong EHK, Hanner RH (2008) DNA barcoding detects market substitution in North American seafood. *Food Research International*, **41**, 828–837.
- Wong EH, Shivji MS, Hanner RH (2009) Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. *Molecular Ecology Resources*, **9**(Suppl. s1), 243–256.
- Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. *The Journal of Heredity*, **92**, 371–373.

Zhang J (2011) Species identification of marine fishes in China with DNA barcoding. *Evid Based Complement Alternat Med*, **2011**, 978253.

A.O.R., C.O., and R.H. participated equally in the design of the study. A.O.R., L.H.G.P., and T.C.M. did most of the laboratory experiments. R.A.C. identified morphologically the specimens. A.O.R. analyzed the data. All authors discussed results. A.O.R., R.H., and C.O. wrote substantial parts of the manuscript. All authors read and approved the final manuscript.

Data Accessibility

DNA sequences: GenBank accessions GU702318–GU702321; GU702322–GU702325; GU702327–GU702338; GU702341–GU702481; GU702483–GU702535; HM424133–HM424137; HQ991927–HQ991931; HQ991933–HQ991935; JN313783; JQ365199–JQ365619; JX033994–JX034022; Final DNA sequence assembly uploaded as online supplemental material; Voucher specimens details and sampling locations: BOLD project Marine Fishes from São Paulo, Brazil – Part 1 (project code: MFSPI), at www.boldsystems.org.

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1 Neighbour-joining tree showing the genetic distances (K2P) among 678 COI sequences belonging to 135 marine fish species from the São Paulo State area, Brazil.

Fig. S2 Diagnostic nucleotides from the COI sequences of *Genidens genidens* and *G. barbatus*.

Appendix S1 Classification, collecting locality, BOLD record numbers, GenBank accession nos and voucher specimen identifier for each of the 678 individual specimens analysed in this study.

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.