MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2010) **10**, 421–430 doi: 10.1111/j.1755-0998.2009.02785.x

DNA BARCODING

DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts

ARIAGNA LARA,* JOSÉ LUIS PONCE DE LEÓN,† RODET RODRÍGUEZ,† DIDIER CASANE,‡ GUILLAUME CÔTÉ,§ LOUIS BERNATCHEZ§ and ERIK GARCÍA-MACHADO*
Centro de Investigaciones Marinas, Universidad de La Habana, Calle 16, No. 114 Entre 1ra y 3ra, Miramar, Playa, Ciudad Habana 11300, Cuba, †Facultad de Biología, Universidad de La Habana, Calle 25, No. 455 Entre J e I, Vedado, Ciudad Habana 10400, Cuba, ‡Laboratoire Evolution Génomes et Spéciation (UPR9034), CNRS, 91198 Gif-sur-Yvette Cedex, France, §Institut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène Marchand, Université Laval, Quebec, QC, Canada G1V 0A6

Abstract

Despite ongoing efforts to protect species and ecosystems in Cuba, habitat degradation, overuse and introduction of alien species have posed serious challenges to native freshwater fish species. In spite of the accumulated knowledge on the systematics of this freshwater ichthyofauna, recent results suggested that we are far from having a complete picture of the Cuban freshwater fish diversity. It is estimated that 40% of freshwater Cuban fish are endemic; however, this number may be even higher. Partial sequences (652 bp) of the mitochondrial gene COI (cytochrome *c* oxidase subunit I) were used to barcode 126 individuals, representing 27 taxonomically recognized species in 17 genera and 10 families. Analysis was based on Kimura 2-parameter genetic distances, and for four genera a character-based analysis (population aggregation analysis) was also used. The mean conspecific, congeneric and confamiliar genetic distances were 0.6%, 9.1% and 20.2% respectively. Molecular species identification was in concordance with current taxonomical classification in 96.4% of cases, and based on the neighbour-joining trees, in all but one instance, members of a given genera clustered within the same clade. Within the genus *Gambusia*, genetic divergence analysis suggests that there may be at least four cryptic species. In contrast, low genetic divergence and a lack of diagnostic sites suggest that *Rivulus insulaepinorum* may be conspecific with *Rivulus cylindraceus*. Distance and character-based analysis were completely concordant, suggesting that they complement species identification. Overall, the results evidenced the usefulness of the DNA barcodes for cataloguing Cuban freshwater fish species and for identifying those groups that deserve further taxonomic attention.

Keywords: COI, DNA barcode, freshwater fish, molecular taxonomy

Received 5 June 2009; revision received 25 August 2009; accepted 3 September 2009

Introduction

Cuban freshwater ichthyofauna is the richest among the Antillean islands (Rosen & Bailey 1963; Robins & Ray 1986; Vergara 1992; Burgess & Franz, 1989). They are classified into 10 orders, 14 families and 35 genera, occurring in rivers, streams, lakes and coastal regions. A total of 57 species have been identified from Cuban freshwater ecosystems in some stage of their life, of which 23 (40.35%) are endemic (Vales *et al.*, 1998). However, Vergara (1992) recognized only 38 strictly freshwater dwelling species, based on their apparent limited dispersal ability and their occurrence in freshwater during all or most of their life cycle. In the present study, we follow Vergara's criteria.

The family Poecilidae is the dominant Cuban fish fauna in terms of number of species. This group includes two endemic genera (*Girardinus* and *Quintana*) and distinctive endemic species of the genera *Gambusia* and *Limia*, some of them are widespread while others are restricted to local or regional areas on the main island (Rivas 1958; Barus *et al.* 1980; Rauchenberger 1989;

Correspondence: Erik García-Machado, Fax: 537 2025223;
E-mail: egarcia@cim.uh.cu

Vergara 1992). *Girardinus* and *Gambusia* apparently underwent extensive radiations and the number of recognized species or subspecies in these genera has increased recently (Rauchenberger 1989; Barus & Wohlgemuth 1994; Lucinda 2003; Doadrio *et al.* 2009). Conversely, *Quintana*, *Limia* and some other genera (*Agonostomus*, *Alepidomus*, *Anguilla*, *Atractosteus*, *Cubanichthys*, *Cyprinodon*, *Dormitator*, *Eleotris*, *Fundulus*, *Gobiesox*, *Gobiomorus*, *Joturus*, *Kryptolebias*, *Ophisternon* and *Sicydium*) are represented by a single species (Vergara 1992). The remaining genera are *Nandopsis* (two endemic species), *Lucifuga* (four endemic species) and *Rivulus* (two endemic species). These contrasting patterns of diversification probably reflect the chronology of colonization events by the different taxa, or the results of the differential success in resources exploitation and adaptability (Briggs 1984; Vergara 1992).

In spite of the accumulated knowledge of the systematics of Cuban freshwater ichthyofauna, recent results suggest that we are far from having a complete picture. For example, putative new species of the genera *Girardinus* (Doadrio *et al.* 2009) and *Lucifuga* (García-Machado, Hernández, García-Debrás, Chevalier-Monteagudo, Bernatchez, Casane, unpublished results) have recently been identified. A good understanding of the systematics of Cuban freshwater fish is important for a number of reasons. Contamination, habitat degradation, overexploitation and introduction of alien species have impacted native ichthyofauna at different scales. Any conservation plan needs to be based on a solid knowledge of the existing evolutionary units (DeSalle & Amato 2004). In addition, various species (e.g. killifishes) have been used as natural controls of insect vectors of disease (García & Koldenkova 1990; Rodríguez *et al.*, 2004; Menéndez *et al.*, 2007). However, there is a complete lack of information on population structure or what independent evolutionary lineages are present, which would serve as guidelines for local introductions.

The DNA barcoding initiative offers the opportunity for a standardized system of species identification based on the analysis of small fragments of DNA (Caterino *et al.*, 2000; Hebert *et al.* 2003). The basic rationale for barcodes is that intraspecific genetic distances should be lower than those estimated between congeneric species (Johns & Avise 1998). A 650-bp segment of the 5′ region of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene is currently used for cataloguing animal biodiversity (Hebert *et al.*, 2003; Hebert *et al.*, 2004; Blaxter *et al.*, 2005; Barber & Boyce, 2006; Pfenninger *et al.* 2007; Hubert *et al.* 2008; Tavares & Baker 2008). In spite of its success, this is still much controversy surrounding this methodology, for example in terms of species definition (Tautz *et al.*, 2003; Blaxter, 2004; Vences *et al.*, 2005; DeSalle, 2006; Cognato, 2006; Rubinoff, 2006), data relevance, distance vs. character analysis (DeSalle *et al.* 2005), and taxonomical background of the analysed group (Meyer & Paulay 2005). However, there are a growing number of examples demonstrating the usefulness of DNA barcoding in different fields of biological sciences including assigning individuals to their corresponding taxa (Hebert *et al.* 2003; Hebert *et al.*, 2004; Lee & Foighil, 2004; Ward *et al.* 2005; Barber & Boyce, 2006; Hajibabaei *et al.*, 2006b,c; Hubert *et al.* 2008), identification of extinct specimens or specimens from collection (Hajibabaei *et al.*, 2006a), forensic analysis (Nelson *et al.*, 2007), detection in markets and food products of regulated species (Smith *et al.* 2007; Yancy *et al.*, 2008) and control of invasive species (Chown *et al.* 2008).

In this study, we use established DNA barcoding methodology to catalogue the freshwater fishes of Cuba using distance and character-based criteria for species identification. We also contribute to the knowledge of this group of animals in the archipelago by uncovering potential cryptic variation as well as conflicting patterns of divergence with respect to current taxonomic designations.

Materials and methods

Males and females, of 27 taxonomically recognized species representing 17 genera and 10 families of Cuban freshwater fishes were collected using hand nets, at different localities across the archipelago (Fig. 1). We also included two putative new species of the genera *Gambusia* and *Lucifuga* and individuals of the species *Lophogobius cyprinoides*, which inhabit fresh to fully marine waters. Fish were transported alive to the laboratory, or directly preserved in 95% ethanol. Species level identification was carried out using morphological characters according to original descriptions.

Total DNA was extracted from muscle (~5 mm^3) or small fin fragments using the DNAeasy tissue kit (QIAGEN). A 652-bp segment from the 5′ region of the mitochondrial COI gene was amplified using the following primers: FishCOIf: 5′-AAYCAYAAAGAYGGYACCCT-3′; FishCOIr: 5′-CNGGRTGNCCRAAGAAYCA-3′. Polymerase chain reactions (PCR) were performed in 50 μL of final volume, containing: 1 μL of total DNA solution, 1X of amplification buffer [10 mM Tris–HCl (pH 9.0), 50 mM KCl, 1% Triton X-100], 200 μM of dNTPs, 200 μM of each one of the primers, 1.5 mM of $MgCl_2$ and a unit of *Taq* DNA polymerase (Promega). The thermal cycling profile was as follows: 94 °C initial denaturing for 5 min; 35 cycles of 94 °C of denaturing during 45 s, 48 °C annealing for 45 s, and 72 °C extension for 1 min and 30 s; and a final extension at 72 °C for 10 min.

Polymerase chain reaction products were purified using the NucleoSpin Extract II kit (Macherey-Nagel),

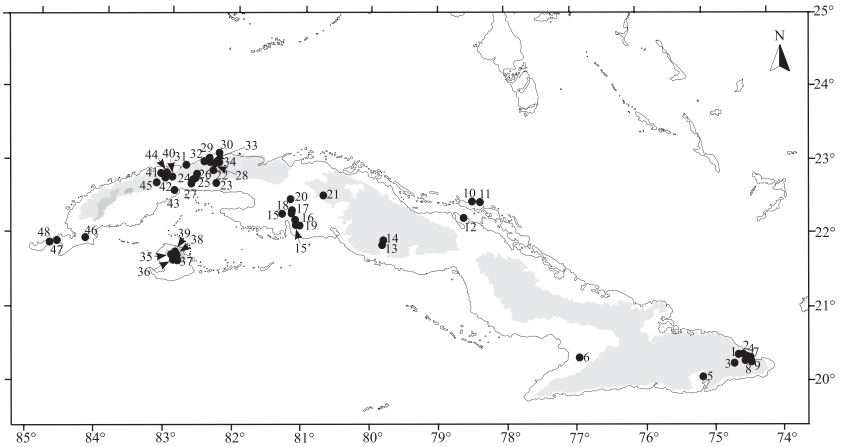

Fig. 1 Sampling sites for the species included in the present study. The numbers correspond to the locality names in Data S1.

and cycle-sequenced in both directions using the primers described previously and an ABI Prism Big Dye terminator sequencing kit V.3 (PerkinElmer). The fragments were resolved using an ABI 3100 automated sequencer (Applied Biosystems).

Sequences were aligned using the BIOEDIT 5.0.9. (Hall 1999). MEGA version 4 (Tamura *et al.* 2007) was used to calculated sequence divergences according to the Kimura 2-parameter (K2P) model (Kimura 1980), and to estimate a neighbour-joining (NJ) tree (Saitou & Nei 1987) to provide a graphic representation of species divergence. Complementary to genetic distance analysis, a population aggregation analysis (PAA) (Davis & Nixon 1992) was carried out for the most species-rich genera (*Gambusia*, *Girardinus*) and for *Lucifuga* and *Rivulus.*

As complementary information for certain comparisons and validations, the homologous sequences of 16 species available from GenBank database were included. These species were: *Anguilla rostrata* (Accession no. EU524442), *Atractosteus spatula* (Accession no. NC_008131), *Cataetyx rubrirostris* (Accession no. AP004407), *Cubanichthys pengelleyi* (Accession no. AY356593), *Diplacanthopoma brachysoma* (Accession no. AP004408), *Dormitator maculatus* (Accession nos AY722137 and AY722143), *Gambusia affinis* (Accession no. NC_004388), *Gobiomorus dormitor* (Accession nos AY722134 and AY722144), *Kryptolebias marmoratus* (Accession no. AF283503), *Lepisosteus oculatus* (Accession no. NC_004744), *Lepisosteus osseus* (Accession no. NC_008104), *Lophogobius cyprinoides* (Accession no. AF391362), *Nandopsis haitiensis* (Accession no. DQ119213), *Nandopsis managuensis* (Accession no. DQ119203), *Nandopsis ramsdeni* (Accession no. DQ119211) and *Nandopsis tetracanthus* (Accession no. DQ119212).

Individual samples were deposited in the collection of Museum of Natural History 'Felipe Poey' (MFP) of the University of Havana. Each studied fish was identified by a voucher showing species name, date of collection and locality. Data for each individual [identification, sex, barcoding number, museum number, sampling location and GenBank Accession numbers (544239–544257 and 545579–545685)] are presented in Data S1 and Fig. 1. Sequences and trace files are also available at Centro de Investigaciones Marinas, University of Havana under request to the authors.

Results

Of the 38 'true' freshwater fishes recognized in Cuba (Vergara 1992), we considered only 35 as valid. We excluded *Girardinus serripenis* Rivas, 1958; which is a synonym of *Girardinus creolus* Garman, 1895, *Lucifuga teresinarum* Diaz Peréz, 1988 that represents a synonymy of *Lucifuga subterranea* Poey, 1958 (García-Machado, Hernández, García-Debrás, Chevalier-Monteagudo, Bernatchez, Casane, unpublished results), and *Rivulus garciai* (de la Cruz & Dubitsky 1976) that was not found at its single highly degraded type locality. Similarly, in the case of the five members of the *puncticulata* species group, we only considered *Gambusia puncticulata* in our study as there is no consensus about the taxonomic status (species or subspecies) for the remaining members (*baracoana*, *bucheri*, *howelli*, *monticola*). A total of 27 (77.1%) of the 35 species were included in the present study (Data S1). Nineteen of these are endemic species, representing 90.5% of Cuba freshwater fish endemic species. The putative *Gambusia* and *Lucifuga* species, described above, were not included in the statistics. Three species (*Girardinus cubensis*, *Gobiesox nudus* and *Joturus pichardii*) were not available for analysis, while available individuals for other species (*Eleotris pisonis*, *Lucifuga simile*, *Ophisternon aenigmaticum* and *Sicydium plumieri*) were formalin fixed or provided low DNA quality resulted in poor or no PCR amplification.

A 652-bp COI fragment, starting at nucleotide position 6452 of the L-strand of the *Cyprinus carpio* mitogenome (Chang *et al.* 1994), was successfully amplified for 126 individuals. As expected from other fish species studies (e.g. Hubert *et al.* 2008), no sequence length differences

were observed. Translation of sequences did not result in stop codons indicating that the amplified domains were functional. In 96.4% of the cases analysis of evolutionary divergence using the K2P model, and the 3% cut-off criteria suggested for species level divergence (Hebert *et al.* 2003), indentified clades that were in concordance with recognized taxonomic units based on morphological characters.

The level of divergence among congeneric species was about 10 times higher than among conspecifics, and divergence levels between confamiliars were about two times higher than congenerics. Thus, the mean conspecific, congeneric and confamiliar genetic distances were 0.4%, 8% and 20.4% respectively (Table 1). Nevertheless, category overlap occurred in some cases as the distribution of the intraspecific K2P distances ranged between 0% and 3.3% and interspecific distances between 1.6% and 18.1% (Fig. 2, Data S2).

The NJ tree (Data S3) based on the K2P distance matrix essentially illustrates the relationships among the intrageneric units. In all but one case, species were clustered within their currently accepted genus.

Novelties and conflicting taxonomical designations

Gambusia. Three *Gambusia* species are recognized for Cuba, one of which (*Gambusia punctata*) is endemic. However, in the present study, we have provisionally designated some individuals, collected at La Jenifer hole at Key Coco, as *Gambusia* sp., taking in to account that they show a slender body shape quite different from that of *G. punctata* and *Gambusia rhizophorae*.

Table 1 Summary of genetic divergences (K2P model) for each taxonomic level of comparison

Comparisons within	Taxa (*n*)	Mean	Min.	Max.	SE
Species*	29	0.004	0	0.033	0.001
Genus, among species†	7	0.080	0.016	0.181	0.011
Family, among genus‡	5	0.204	0.125	0.234	0.019

*The GeneBank available sequences of the species *Anguilla rostrata*, *Dormitator maculatus*, *Kryptolebias marmoratus*, *Lophogobius cyprinoides*, *Nandopsis ramsdeni*, and *Nandopsis tetracanthus* (see Materials and methods) were included to estimate divergence values (see Data S2 and S3, for details).
†As above, the species *Atrastosteus spatula*, *Cubanichthys pengelleyi*, *Nandopsis ramsdeni* and *Nandopsis tetracanthus* (see Materials and methods) were included to estimate divergence values.
‡As above, the species *Atrastosteus spatula*, *Lepisosteus oculatus*, *Lepisosteus osseus* (Lepisosteidae); *Kryptolebias marmoratus* (Rivulidae), *Cataetyx rubrirostris* and *Diplacanthopoma brachysoma* (Bythitidae) (see Materials and methods) were included to estimate divergence values.

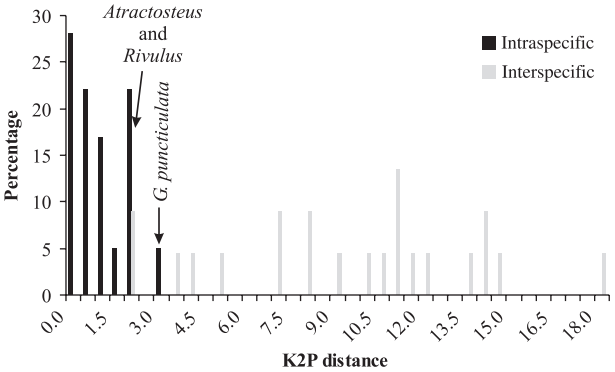

Fig. 2 Distribution of the genetic distances (K2P) within and between species.

The conspecific genetic distances in *G. rhizophorae* and *Gambusia* sp. showed a mean of 0.1 ± 0.1% and 0.0 respectively (Data S2). However, *G. punctata* and *Gambusia puncticulata* showed values of 3.1 ± 0.4% and 3.3 ± 0.4%, near the cut-off for the designated species limit. The smallest interspecies genetic divergence values of *Gambusia* were observed between *G. punctata* and *Gambusia* sp. (4.9 ± 0.8%), and between *G. punctata* and *G. rhizophorae* (5.2 ± 0.8%), whereas the remaining pairwise comparisons among species ranged from 6.2% to 11.3%.

The NJ tree (Data S3) also clearly showed genetic subdivisions within *G. punctata* and *G. puncticulata*. Surprisingly, three individuals of *G. punctata* sampled at the Yara River in Guantánamo province had identical sequences and clustered together within *Gambusia* sp. Hereafter, we include this group within *Gambusia* sp. Sequence divergence between *G. punctata* and the redefined *Gambusia* sp. was 6.4 ± 1.0%, while between *G. punctata* and *G. rhizophorae*, and *Gambusia* sp. and *G. rhizophorae*, it was 4.9 ± 0.8% and 6.2 ± 1.0% respectively.

In the case of *G. puncticulata*, four clades were identified, all supported by high bootstrap values: one consisted of samples from the most wide-ranging taxon recognized as *G. puncticulata* (*sensu stricto*); a second one included samples from Coco Key [hereafter designated as *G. puncticulata* (A)]; a third included samples from Río Baracoa and Guahanahacabibes [hereafter designated as *G. puncticulata* (B)]; and a fourth group included samples from Manglarito River in Baracoa [hereafter designated as *G. puncticulata* (C)] (Data S1, Fig. 1). Based on these results, these groups were treated as discrete units, and the pairwise level of divergence was recalculated. The results showed that the level of divergence among the new recognized groups was higher than 3% in all cases: *G. puncticulata–G. puncticulata* (A) (7.2 ± 1%); *G. puncticulata–G. puncticulata* (B) (4.7 ± 0.8%); *G. puncticulata–G. puncticulata* (C) (6.3 ± 1%); *G. punticulata* (A)–

G. punticulata (B) (8.0 ± 1.2%); *G. punticulata* (A)–*G. punticulata* (C) (3.6 ± 0.7%); and *G. puncticulata* (B)–*G. punticulata* (C) (7.8 ± 1.2%). In contrast, intraspecific genetic distances fell between 0.0% and 0.2% in all four groups.

A total of 132 variable sites were identified amongst the *Gambusia* spp. Following PAA, four sites were indentified as diagnostic for *G. punctata*, five for *G. rhizophorae*, 10 for *Gambusia* sp. and 26 for *G. puncticulata* (*sensu lato*). However, iterative grouping further identified four subgroups within *G. puncticulata*, in concordance with those identified by genetic distance analysis. *G. puncticulata* (*sensu stricto*) was differentiated from the other Cuban *Gambusia* by three diagnostic sites, *G. puncticulata* (A) by 11 sites, *G. puncticulata* (B) by 11 sites, and *G. puncticulata* (C) by six sites (Fig. 3).

Rivulus. The two Cuban species of *Rivulus* (*Rivulus cylindraceus* and *Rivulus insulaepinorum*) were analysed in the present survey. The estimated level of sequence divergence (1.8 ± 0.4%) between these species was similar to that observed at the intraspecific level for *R. cylindraceus* (1.6 ± 0.4%). In addition, the NJ tree did not show reciprocal monophyly for these species, and a similar result was obtained by PAA. Only 25 variable sites were found for *Rivulus* and none of them was diagnostic for species identification. However, using distance analysis, two subgroups were observed within *R. cylindraceus* (d = 3.0 ± 0.6%) (Data S3), differentiated by 17 sites. One subgroup consisted of the samples from Havana and Pinar del Rio localities, while the second one included those from Zapata swamp. *Rivulus insulaepinorum* had eight polymorphism sites but none of the sites as diagnostic. One individual of this species clustered within *R. cylindraceus* group from the Zapata swamp.

Cubanichthys. In Cuba, the genus *Cubanichthys* is represented by a single endemic species, *Cubanichthys cubensis* Eigenmann, 1903. A second species, *Cubanichthys pengelleyi* Fowler, 1939 has been identified in Jamaica. Comparison of the sequence divergence between these species was estimated to be 18.1 ± 1.9%, which is the maximum sequence divergence found in this study among congeneric species (Table 1, Data S2). This value was only approached by *Lucifuga* species (11.1%).

Lucifuga. Of the seven recognized species for this genus, four are endemic to Cuba. Here, we analysed two of them, *Lucifuga dentata* and *Lucifuga subterranea*, as well as a putative new species from a locality at the Western extreme of the island. The level of sequence divergence between the three taxa was amongst the highest observed between species within genera (mean 11.1%, range 10.8–11.5%). Correspondingly, the number of diagnostic characters was relatively high for all three taxa (*L. dentata*, 36 sites; *L. subterranean*, 33 sites; *Lucifuga* sp. n., 32 sites).

The two endemic genera

Girardinus *and* Quintana. Six of the seven recognized species of *Girardinus* were included in the present study (Data S1). It should be noted that although we made an effort to collect *G. cubensis* at its type localities we failed to find this species. Distance analysis was in complete concordance with current classification based on morphological characters. The highest intraspecific divergence was found among *Girardinus denticulatus* conspecifics (2.3 ± 0.5%), followed by *Girardinus microdactylus* (1.9 ± 0.4%), while congeneric divergences ranged from 4.1 ± 0.8% between *Girardinus falcatus* and *Girardinus uninotatus*, and 14.7 ± 1.5% between *G. falcatus* and *G. denticulatus*. A total of 147 variable sites were observed between *Girardinus* sequences, five were diagnostic for *G. uninotatus*, six for *G. creolus*, six for *G. microdactylus*, 11 for *Girardinus metallicus*, 10 for *G. falcatus* and

Fig. 3 Diagnostic sites defining the different evolutionary partitions inside *Gambusia* genus in Cuba. When a given site was polymorphic inside a given taxa, conventional designations (R = G or A, Y = C or T) were used to indicate the type of nucleotide base observed at the site.

```
                                    Nucleotide position
                                    1111111111222223333444445555555556
                                    3356899990012556679014771167578889002555679 4
Species                             8245123489065117345621114234648067326036987 29

G. punctata                         cgcggtgagatctcttagaYaccctcgcatcaagRagaYtaYcaa
                                       *  *                          *     *
G.puncticulata                      cgtgatgtgatcacttagagaccatcgcatcaaacagaacaacaa
                                          *           *                      *
G.puncticulata (A)                  catgacagagccacttagacaccattgcgttaaatgtaacaccaa
                                     *    * ***             *  * *   ***
G.puncticulata (B)                  tgggacgggattatctaagaatcatcgaatctaacagaacaccaa
                                    * *        * **  *** *     *   *
G.puncticulata (C)                  cgtgacggcctcacttagacactatcacatcagacagagcaccaa
                                          **           *   *     *     *
G.rhizophorae                       cgcgaagggatctcttagataccgccgcatcaaacagatcattag
                                         *                **                * *
Gambusia sp.                        cgcaatgggatccctaggacgccctcgcaccaaaaagttcggcga
                                       *        *  **   *        *       *  ** *
```

26 for *G. denticulatus*. Finally, *Quintana atrizona*, the other Cuban member of the tribe Girardininii *sensu* Rosen & Bailey (1963), showed a mean divergence value of 20.0 ± 1.8% (range 19.3 ± 1.8% to 21.1 ± 2.0%) with respect to *Girardinus* (Data S2).

Discussion

The present study represents the first molecular survey of Cuban freshwater ichthyofauna. Amplification of the COI 5′ region (652 bp) was successful for all assayed individuals. We did not detect nuclear copies of COI gene, either as frame shifts in the inferred protein sequence, or as mutation bias. Taxonomical confusion because of hybridization did not appear to be a problem; however, as will be noted below, in one case this could not be completely ruled out.

The level of genetic variation observed for the COI gene fragment was highly congruent with the taxonomic level. Thus, a 3% of cut-off appears generally adequate to discriminate species of Cuban freshwater fishes. In general, the average conspecific, congeneric and confamilial K2P distances were within the range observed among the Canadian freshwater fishes (0.27%; 8.37%, and 15.38% respectively) (Hubert *et al.* 2008) and Australian marine fishes (0.39%, 9.93%, and 15.46% respectively) (Ward *et al.* 2005). The character-based analysis results were completely in agreement with the genetic distance approach for species identification in all cases when both approaches were used. This supports the idea that both approaches used together can complement and facilitate species identification (Waugh *et al.* 2007).

Overlapping of conspecific and congeneric levels of divergence was minimal. Only two genera, *Atrastosteus* (i.e. between the Cuban gar *Atrastosteus tristoechus* and the Alligator gar *Atrastosteus spatula*) and *Rivulus*, showed interspecific divergences that fell below the threshold chosen to differentiate inter and intraspecific divergences. Mutation rate variation among genes and species has been recognized as a major problem for sequence divergence estimation and phylogenetic reconstructions (Ayala 1997; Johns & Avise 1998; Baer *et al.* 2007), and may sometimes hinder the use of the DNA barcoding for cataloguing species diversity (Stoeckle 2003; DeSalle *et al.* 2005; Smith *et al.* 2007). This appears to be the reason for low levels of sequence divergence that characterizes most comparisons of congeneric Lepisosteids, because a similar estimate (0.9 ± 0.4%) was obtained among the sequences of *Lepisosteus oculatus* and *Lepisosteus osseus* (this study). In addition, the family Lepisosteidae showed the smallest confamiliar divergence levels (*Atractosteus* vs. *Lepisosteus*, 12.5 ± 1.5%) (Data S2). Bermingham & Avise (1986) found that low levels of sequence divergence characterized geographic subdivided populations of the bowfin *Amia calva* (Aiiformes), compared with teleost fish species with similar geographic distributions. Similarly, Krieger & Fuerst (2002) showed that mutation rates are consistently lower for nuclear and mitochondrial genes in Acipenceriforms (sturgeons and paddlefish). Lower rates of sequence divergence in lepisosteids strengthens the suggestion that differences in mutation rates and/or historical effective population sizes may exist between teleost and nonteleost fishes (Bernatchez & Wilson 1998). Nonetheless, some teleost fishes also show very low interspecies divergence. For example, tunas of the genus *Thunnus* have a mean interspecies distance of only 1.04% (Ward *et al.* 2005), probably due to a recent radiation of this genus.

However, a similar argument cannot apply to the case of *Rivulus*. The two Cuban sister species *Rivulus cylindraceus* and *Rivulus insulaepinorum* showed a sequence divergence of 1.8%, similar to the 1.6% intraspecific comparisons of *R. cylindraceus*. Character-based analysis also failed to distinguish the two species, and suggests a closer relationship between *R. insulaepinorum* and *R. cylindraceus* from the Zapata swamp locality. In contrast to the situation in the Lepisosteids, an increase in substitution rates has been reported for mitochondrial genes in annual Rivulids relative to nonannual *Rivulus* species (Murphy *et al.* 1999). The sequence divergence observed between the nonannuals, *Rivulus* and *Kryptolebias*, was 23.1%, an estimate that was among the highest observed for confamilial comparisons in this study (Data S2). Moreover, Johns & Avise (1998) in an analysis of all available cytochrome *b* genetic distance data, found that for 98% of sister species the genetic divergence was >2%. Although fishes showed a large range in variation, the minimum estimate for *Rivulus* species in their study was >3%.

Meyer & Paulay (2005), applying barcoding methods to a well-studied group of cypraeid marine gastropods, demonstrated that well-defined species can show K2P estimates as low as 1.2%. The authors concluded that barcoding can only be reliable if based on solid taxonomic foundations. However, the original description of *R. insulaepinorum* was based on the analysis of only four adult females and two juveniles, and a very limited number of morphological diagnostic characters were described (de la Cruz & Dubitsky 1976). Moreover, the new species was reported in sympatry with respect to the more widely distributed *R. cylindraceus*. The lack of a clear molecular distinction between these species could have three possible explanations: (i) the time since the separation of populations has been too short for reciprocal monophyly; (ii) the occurrence of relatively recent introgressive hybridization among sympatric populations; (iii) or, simply, an erroneous taxonomic designation. While

the first two explanations cannot be ruled out, a critical reanalysis of the morphological characters together with the evaluation of ecological, reproductive and ethological data is necessary to determine the taxonomic status of *R. insulaepinorum*. In addition, analysis of both nuclear and mitochondrial genetic markers is essential to test whether there has been introgressive hybridization between these species.

In an opposite extreme from the genera discussed above, which have reduced levels of inter-specific divergence; we found that the genus *Cubanichthys* is composed of two unusually highly divergent species: *Cubanichthys cubensis* and *Cubanichthys pengelleyi*, from Cuba and Jamaica respectively. The level of divergence estimated between them (18.1 ± 1.9%) was the highest of all congeneric comparisons made in this study, as well as much higher than the mean estimates for Canadian freshwater fishes (mean = 8.37%) (Hubert *et al.* 2008), and Australian marine fishes (mean = 9.93%) (Ward *et al.* 2005). This value was also higher than the maximum value observed among the Cuban *Gambusia* and *Nandopsis* sequences and the available sequences of congenerics (*Gambusia affinis*, *Nandopsis haitiensis* and *Nandopsis managuensis*) from Central and North America, and Haiti (Data S2). We further checked the sequences for nucleotide composition bias and there were no significant differences among them. On the other hand, all nucleotide substitutions (99 in total) were conservative, showing a moderate transition/transversion rate (2.4). A maximum of 19.3% of divergence for congeneric comparisons has been reported (Stoeckle 2003; DeSalle *et al.* 2005; Smith *et al.* 2007; Hubert *et al.* 2008). Although we cannot rule out an accelerated rate of sequence evolution in these species, our result strongly suggests that the taxonomy of this genus deserves further attention, particularly the possible assignment of the Jamaican species to a new genus.

Of all the genera in this study, *Gambusia* is the most speciose. In addition to the three nominal Cuban species of the genus (*Gambusia punctata*; *Gambusia puncticulata* and *Gambusia rhizophorae*), four new evolutionary distinct entities were indentified by both distance and character-based analysis of the COI sequence.

An a priori morphologically distinctive group, designated here as *Gambusia* sp., was collected for the first time, in a brackish water pond at Key Coco locality and identified as a separate evolutionary lineage from the three species above by distance (range 6.2–13.8%) and PAA (10 diagnostic nucleotide sites) analysis. However, the most remarkable result was that it is indistinguishable, at the sequence level, from other *Gambusia* individuals sampled at River Yara (Eastern Cuba) which are characterized by a *G. punctata*-like morphology. These two localities are separated by about 300 km and are completely unconnected at the present time (Coco key is in the Northeastern archipelago and the Yara River flows southwards through the Gramma province and surroundings) (Fig. 1). As far as we know, *G. puntata* is a strictly freshwater fish, so a possible translocation of this species to a brackish water environment appears unlikely a priori. However, this possibility cannot be ruled out because, in this study, we also report *Nandopsis tetracanthus* from Coco key. Disentangling this unexpected pattern of relationships is necessary, and may reveal historical connections related to the complex geological formation of Cuba.

Cryptic diversity was identified within what, a priori, was identified as *G. puncticulata*. Four groups with genetic divergences between 3.6% and 8.0%, and a relatively large number of diagnostic sites, were identified. As expected, *G. puncticulata sensu stricto* showed the widest distribution across Cuba, while *G. puncticulata* A, *G. puncticulata* B, and *G. puncticulata* C may represent new unnamed taxa with more restricted distributions (Data S1, Fig. 1). This finding was not surprising as five species or subspecies have been recognized in Cuba, most of them described for single localities in Eastern Cuba and the Isle of Pines (Rivas 1944; Rosen & Bailey 1963; Rauchenberger 1989). Vergara (1992) has suggested that in Cuba the *G. puncticulata* group shows a complex pattern of relationships highly consistent with a common, central species and several derived peripheral species which may have originated in allopatry because of the eustatic fluctuations that occurred during quaternary glaciations. However, it should be noted that we collected samples from Miel River, the type locality of *Gambusia baracoana* Rivas, 1944; and we identified these individuals as *G. puncticulata* (BC 117 and 118, Data S1) using morphological and nucleotide sequence characters (COI). Similarly, a specimen from Isla de la Juventud, which is within the physical range of *G. howelli* fell within the morphology and molecular range of *G. puncticulata*. The results suggest that the *G. puncticulata* species group also deserves further systematic attention.

Among the Cuban poecilids, *Girardinus* is the richest genus in terms of species. Its taxonomy had also been rich in disagreements, with some authors suggesting that *Girardinus* is really several genera (Eigenmann 1903; Rivas 1958) while others (i.e. Rosen & Bailey 1959, 1963) recognize it as a single genus (*Girardinus*). Recently, Doadrio *et al.* (2009), using cytochrome *b* genetic uncorrected *p*-distances, concluded that the observed estimates [mean 9.6% (7.1–12.9)] justify a re-splitting of *Girardinus* into the previously described genera: *Dactylophallus*, *Girardinus*, *Glaridicththys* and *Toxus*. However, such a recommendation may be premature without a critical comparative appraisal of genetic divergences. Namely, the authors did not use the available information

from other close-related groups, which were included as outgroups in their study. For instance, from Doadrio *et al.* (2009; Figure 6), representatives of the monophyletic genus *Limia* (Hamilton 2001) show higher intraspecific divergence levels than those estimated among some of the above-mentioned genera. We recalculated the levels of divergence from the available sequences (*Limia dominicensis* Accession no. EF017533, *Limia melanogaster* Accession no. EF017534, *Limia tridens* Accession no. EF017535, *Limia vittata* Accession nos FJ178765 and FJ178766). Interspecific divergence estimates ranged from 6.2% to 11.1%. Our COI sequence results show that the range of genetic divergence (4.4–14.6%) among *Girardinus* spp. is similar to that observed among the Cuban *Gambusia* (3.6–13.8%) and from other fish barcoding studies (Ward *et al.* 2005; Hubert *et al.* 2008) and is consistent with a single genus. We agree with the statement of Rosen & Bailey (1963) that '…the genus can and should serve to express relationships…' In our view then, the generic name *Girardinus* better expresses the evolutionary distinctiveness of this endemic group of poecilids.

Lucifuga is another genus for which this study suggests a need for new taxonomic designations. Although not all species were included in this survey, our analysis of COI sequence from *Lucifuga dentata* and *Lucifuga subterranea* was in complete concordance with those from morphological analyses A third new entity was also identified, which had a level of genetic divergence as high as that observed between *L. dentata* and *L. subterranea*. This individual has unique morphological attributes (e.g. the occurrence of relatively well-developed eyes together with a lack of palatine teeth) that distinguish it from the rest of the described Cuban species.

In summary, the present study provides a clear example of the usefulness of barcoding for cataloguing the diversity of Cuban freshwater fishes. The discovery of several putative new species for several genera suggests that our current knowledge of Cuban ichthyofauna is far from being complete. Admittedly, this study is not sufficient in itself to solve those taxonomic issues. However, it serves to pinpoint those taxa requiring further analysis, as well as providing the basic information on those species that may represent good models for comparative phylogeographic studies. A better knowledge of the systematics of Cuban native freshwater fishes may also contribute to improving conservation programmes, as well as monitoring the impact of ecological changes. For example, the introduction of alien species in freshwater ecosystems has been a common practice in Cuba for the last two centuries. At present, around 28 introduced species have been reported, and there are no published studies critically evaluating their impact on native ichthyofauna.

Acknowledgements

We are really grateful to Eduardo Abreu, Daniel Rodríguez Vázquez, Rafael Cardet Sánchez, Rodolfo Rodríguez, Miguel Bayona and Gunnary Leon who provided us important fish specimens and field assistance. We thank two anonymous referees and Brian Golding (Subject editor) for the useful comments that improved the manuscript. We also thank Cushla Metcalfe for the revision of the English language. Financial support for the sequencing analyses of this project came from L.B.'s Canadian Research Chair in Genomics and Conservation of Aquatic Resources and the D.C.'s Equipe ATIP 'Evolution moléculaire et fonctionnelle des familles multigéniques' Laboratoire Evolution Génomes et Spéciation (UPR9034). Sampling was partially financed by Rufford Small Grants for Nature Conservation assigned to J.L.P.L. Other financial support came from IFS projects assigned to E.G.M. laboratory.

References

Ayala FJ (1997) Vagaries of the molecular clock. *Proceedings of the National Academy of Sciences, USA*, **94**, 7776–7783.

Barber P, Boyce SL (2006) Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. *Proc. R. Soc. B*, **273**, 2053–2061.

Baer CF, Miyamoto MM, Denver DR (2007) Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nature Reviews Genetics*, **8**, 619–631.

Barus V, Wohlgemuth E (1994) Two proposed subspecies in *Girardinus microdactylus* (Pisces: Poeciliidae) from isla de Cuba and isla de la Juventud. *Folia Zoologica*, **43**, 245–254.

Barus V, Liborsvársky J, De la Cruz J (1980) Observations on *Limia vittata* (Poeciliidae) from Cuba. *Folia Zoologica*, **29**, 267–287.

Bermingham E, Avise JC (1986) Molecular zoogeograhy of freshwater fishes in the southeastern United States. *Genetics*, **113**, 939–963.

Bernatchez L, Wilson C (1998) Comparative phylogeography of Neartic and Paleartic fishes. *Molecular Ecology*, **7**, 431–452.

Blaxter ML (2004) The promise of a DNA taxonomy. *Phil. Trans. R. Soc. Lond. B*, **359**, 669–679.

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. *Phil. Trans. R. Soc. B*, **360**, 1935–1943.

Briggs JC (1984) Freshwater fishes and biogeography of Central America and the Antilles. *Systematic Zoology*, **33**, 428–435.

Burgess GH, Franz R (1989) Zoogeography of the Antillean freshwater fish fauna. In: *Biogeography of the West Indies: past, present and future*. (ed Woods CA.), pp. 263–304. Sand Hill Crane Press, Gainesville, Florida.

Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving tower of babel. *Annual Review of Entomology*, **45**, 1–54.

Chang YS, Huang FL, Lo TB (1994) The complete nucleotide sequence and gene organization of carp (*Cyprinus carpio*) mitochondrial genome. *Journal of Molecular Evolution*, **38**, 138–155.

Chown SL, Sinclair BJ, Jansen van Vuuren B (2008) DNA barcoding and the documentation of alien species establishment on sub-Antarctic Marion Island. *Polar Biology*, **31**, 651–655.

Cognato AI (2006) Standard percent DNA sequence difference for insects does not predict species boundaries. *Journal of Economic Entomology*, **99**, 1037–1045.

de la Cruz J, Dubitsky AM (1976) Dos nuevas especies de peces dulceacuícolas del género *Rivulus*, Poey (Cyprinodontidae) de Cuba e Isla de Pinos. *Poeyana*, Series A, **155**, 1–6.

Davis JI, Nixon KC (1992) Populations, genetic variation, and the delimitation of phylogenetic species. *Systematic Biology*, **41**, 421–435.

DeSalle R (2006) Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. *Conservation Biology*, **20**, 1545–1547.

DeSalle R, Amato G (2004) The expansion of conservation genetics. *Nature Reviews Genetics*, **5**, 702–712.

DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation, and DNA barcoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1905–1916.

Doadrio I, Perea S, Alcaraz L, Hernandez N (2009) Molecular phylogeny and biogeography of the Cuban genus *Girardinus* Poey, 1854 and relationships within the tribe Girardinini (Actinopterygii, Poeciliidae). *Molecular Phylogenetics and Evolution*, **50**, 16–30.

Eigenmann CH (1903) The fresh-water fishes of western Cuba. *Bulletin of the U. S. Fish Commission*, **22**, 211–236.

García I, Koldenkova L (1990) *Clave pictórica para las principales especies de peces larvívoros de Cuba*. Instituto de Medicina Tropical Pedro Courí, Cuba.

Hajibabaei M, Smith MA, Janzen DH, Rodríguez JJ, Whitfield JB, Hebert PDN (2006a) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006b) DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA*, **103**, 968–971.

Hajibabaei M, Singer GAC, Hickey DA (2006c) Benchmarking DNA barcodes: an assessment using available primate sequences. *Genome*, **49**, 851–854.

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

Hamilton A (2001) Phylogeny of *Limia* (Teleostei: Poecilidae) based on NADH dehydrogenase subunit 2 sequences. *Molecular Phylogenetics and Evolution*, **19**, 277–289.

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, Cywinska A, Ball SL, deWaard J (2004) Identification of birds through DNA barcodes. *PLoS Biol*, **2**, e312.

Hubert N, Hanner R, Holm E *et al.* (2008) Identifying Canadian freshwater fishes through DNA barcodes. *PLoS ONE*, **3**, e2490, doi: 2410.1371/journal.pone.0002490.

Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution*, **15**, 1481–1490.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **15**, 111–120.

Krieger J, Fuerst PA (2002) Evidence for a slowed rate of molecular evolution in the Order Acipenseriformes. *Molecular Biology and Evolution*, **19**, 891–897.

Lee T, Foighil DO (2004) Hidden Floridian biodiversity: mitochondrial and nuclear gene trees reveal four cryptic species within the scorched mussel, *Brachidontes exustus*, species complex. *Molecular Ecology*, **13**, 3527–3542.

Lucinda PHF (2003) Family Poeciliidae. In: *Check List of the Freshwater Fishes* (eds Reis RE, Kullander SO & Ferraris C Jr), pp. 556–582. Porto Alegre, Edipucrs, 729 pp.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.

Menéndez Z, Manso E, Castex M, Fuentes O, Hernández N, García I (2007) Algunos aspectos ecológicos de la ictiofauna larvífaga presente en Cayo Santa María, Cuba. *Rev. Cubana Med. Trop.*, **3**, 59.

Murphy WJ, Thomerson JE, Collier GE (1999) Phylogeny of the Neotropical Killifish Family Rivulidae (Cyprinodontiformes, Aplocheiloidei) inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **13**, 289–301.

Nelson LA, Wallman JF, Dowton M (2007) Using COI barcodes to identify forensically and medically important blowflies. *Medical and Veterinary Entomology*, **21**, 44–52.

Pfenninger M, Nowak C, Kley C, Steinke D, Streit B (2007) Utility of DNA taxonomy and barcoding for the inference of larval community structure in morphologically cryptic *Chironomus* (Diptera) species. *Molecular Ecology*, **16**, 1957–1968.

Rauchenberger M (1989) Systematics and biogeography of the genus *Gambusia* (Cyprinodontiformes: Poecilidae). *American Museum Novitates*, **2951**, 1–74.

Rivas LR (1944) Contributions to the study of the Poecilid fishes of Cuba. I. Descriptions of six new species of the subfamily Gambusiinae. *Proceedings of the New England Zool. Club*, **23**, 41–53.

Rivas LR (1958) The origin, evolution, dispersal and geographical distribution of the Cuban poeciliid fishes of the tribe Girardinini. *The American Philisophical Society*, **102**, 281–320.

Robins CR, Ray GC (1986) *A Field Guide to Atlantic Coast Fishes of North America*. Houghton Mifflin Company, Boston, MA, USA, 354 pp.

Rodríguez J, González E, Hernández N, Capó V, García I (2004) Comparación morfológica e histológica del tubo digestivo de *Gambusia puncticulata* y *Girardinus metallicus*, peces utilizados en el control biológico de mosquitos. *Rev. Cubana Med. Trop.*, **56**, 73–76.

Rosen DE, Bailey RM (1959) Middle-American poeciliid fishes of the genera *Carlhubbsia* and *Phallichthys*, with descriptions of two new species. *Zoological*, **44**, 1–44.

Rosen DE, Bailey RM (1963) The poecilid fishes (Cyprinodontiformes), their structure, zoogeography and systematics. *Bulletin of the American Museum of Natural History*, **126**, 1–176.

Rubinoff D (2006) Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*, **20**, 1026–1033.

Saitou N, Nei M (1987) The Neighbor-Joining method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Smith MA, Poyarkov NA, Hebert PDN (2007) CO1 DNA barcoding amphibians: take the chance, meet the challenge. *Molecular Ecology Notes*, **7**, 235–246.

Stoeckle M (2003) Taxonomy, DNA and the bar code of life. *BioScience*, **53**, 2–3.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends Ecol. Evol.*, **18**, 70–74.

Tavares ES, Baker AJ (2008) Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. *BMC Evolutionary Biology*, **8**, 81.

Vales MA, Álvarez MA, Montes L, Ávila A (1998) *Estudio nacional sobre la diversidad biológica en la República de Cuba*. PNUMA, Centro Nacional de Biodiversidad, Instituto de Ecología y Sistemática, CITMA, La Habana, Cuba, 480 pp.

Vences M, Thomas M, Van der Meijden A, Chiari Y, Vieites DR (2005) Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front. Zool.*, **2**, 1–12.

Vergara RR (1992) *Principales características de la ictiofauna dulceacuícola cubana*. Editorial Academia, La Habana, Cuba, 27 pp.

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

Waugh J, Huynen L, Millar C, Lambert D (2007) DNA barcoding of animal species – response to DeSalle. *BioEssays*, **30**, 92–93.

Yancy HF, Zemlak TS, Mason JA, Washington JD, Tenge BJ, Nguyen NL, Barnett JD, Savary WE, Hill WE, Moore MM, Fry FS, Randolph SC, Rogers PL, Hebert PDN (2008) Potential use of DNA barcodes in regulatory science: applications of the Regulatory Fish Encyclopedia. *J. Food Prot.*, **71**, 210–217.

Supporting Information

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

Data S1 Summary of the specimens information including sampling sites, GenBank Accession numbers, taxonomy, barcode number and vouchers accession numbers in the collection of the Felipe Poey Museum, Faculty of Biology, University of Havana.

Data S2 Detailed genetic divergences (K2P model) for each taxonomic level of comparison.

Data S3 Neighbour-joining tree, constructed from K2P distance matrix, of the 27 freshwater fish species included in the analysis. Bootstrap values (1000 replicates) for a particular node are placed above the branches.

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