

# An optimised protocol for barcoding museum collections of decapod crustaceans: a case-study for a 10–40-years-old collection

Dario Zuccon<sup>A,C</sup>, Julien Brisset<sup>A</sup>, Laure Corbari<sup>B</sup>, Nicolas Puillandre<sup>B</sup>, José Utge<sup>A</sup> and Sarah Samadi<sup>A,B</sup>

<sup>A</sup>Service de systématique moléculaire (CNRS-MNHN, UMS2700), Département Systématique et Evolution, Muséum national d'Histoire naturelle, CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05, France.

<sup>B</sup>'Systématique, Adaptation et Evolution', UMR 7138 UPMC-IRD-MNHN-CNRS (UR IRD 148), Département Systématique et Evolution, Muséum national d'Histoire naturelle, CP 26, 57 Rue Cuvier, F-75231 Paris Cedex 05, France.

<sup>C</sup>Corresponding author. Email: [dario.zuccon@mnhn.fr](mailto:dario.zuccon@mnhn.fr)

**Abstract.** The sequencing of the crustacean collection of the MNHN, Paris, constitutes a promising yet very challenging barcoding project. For the collection's crustacean specimens preserved in ethanol, some of which were collected up to 40 years ago, the conventional COI barcoding procedure of amplification with Folmer primers failed for more than half of the specimens (58%,  $n = 1920$ ). We hypothesised that this failure may have been due to incompatible mismatches between the crustaceans targeted and the Folmer primer sequences and/or the amount of degradation of the DNA extracted from museum specimens. The comparison of the Folmer primers against the COI sequences from GenBank complete decapod mitochondrial genomes revealed that the annealing regions were, in fact, rather conserved, suggesting that the amplification failures were due more likely to the low quality of the DNA isolated. Using an alignment of all available decapod sequences we designed two internal primers in the middle of the barcoding COI region and also selected two additional external primers to be used as alternative to the standard Folmer primers. Using a two-overlapping-fragments amplification strategy and different primer combinations, our new protocol significantly increased the amplification success rate of the collection material from 42% with the Folmer primers to 84%, recovering an additional 364 complete barcodes and 443 minibarcodes (i.e. fragments of less than 400 base pairs), and expanding the species coverage from 254 to 397 barcoded crustaceans.

Received 17 April 2012, accepted 26 September 2012, published online 19 December 2012

## Introduction

The Muséum national d'Histoire naturelle, Paris (MNHN), has a long tradition of carcinological studies. Leading zoological figures of the 19th century such as P.A. Latreille and H. Milne-Edwards established the MNHN crustacean collection, which has grown into one of the more important international collections in existence today. Forty years of marine expeditions as part of the ongoing Tropical Deep-Sea Benthos program, led by the MNHN in collaboration with the Institut de Recherche pour le Développement, have assured a continuous flow of new material from the Pacific and Indian Oceans into the collection (Bouchet *et al.* 2008). According to conservative estimates the MNHN crustacean collection contains over 120 000 lots representing at least 20 000 species and includes ~8000 types.

We took advantage of the rich and comparatively recent material obtained during the Tropical Deep-Sea Benthos campaigns to initiate a DNA barcoding project at the MNHN, as part of the international Marine Barcode of Life project started in 2007 ([www.marinebarcoding.org](http://www.marinebarcoding.org)) (Bucklin *et al.* 2011). Two

main taxonomic groups, the marine molluscs and the decapod crustaceans, were targeted as part of the MNHN contribution to the MarBOL project. The decapods were specifically chosen as they represent one of the major components of the MNHN crustacean collection. Thanks to the efforts of several international taxonomists, the decapod collection at the MNHN is very well studied. Almost all specimens are identified to species level and numerous new species have been described using this material (among the most recently published papers are, for example, Castro 2010; Komai 2011; Chan *in press*). This situation offers excellent opportunities to develop a Barcoding reference database, relying on numerous specimens identified to species that are preserved in 75% ethanol. Given its richness in type specimens, it will allow us to test the species-delimitation hypotheses established on morphological grounds (Puillandre *et al.* 2011).

Despite the potential of the MNHN decapod collection to be an excellent reference source for developing a DNA barcoding database, it soon became clear from our preliminary analyses that

this material presented specific challenges in recovering COI sequences. More than half (58%) of the specimens chosen for initial barcoding analysis failed to amplify using the universal invertebrate COI Folmer primers (Folmer *et al.* 1994). Given that cross-species amplification success is dependent upon the number and position of primer-template mismatches (Housley *et al.* 2006) and that DNA degradation in museum specimens is a well known factor affecting the chance of recovering amplicons using standard PCR protocols (e.g. Dean and Ballard 2001; Mandrioli *et al.* 2006; Zimmermann *et al.* 2008), we formulated three hypotheses that might account for the amplification failures: (1) the Folmer primers are not compatible with decapod crustaceans as a whole; (2) the Folmer primers were suitable only for some groups but not all (a taxonomic bias within the decapods); and (3) despite the relatively young age of the specimens and the preservation conditions the DNA was degraded to the point that most DNA template molecules were of shorter size than the COI barcode fragment. Here we describe the procedure used to test these three hypotheses and to optimise a decapod-specific amplification protocol for collection material using internal and specific primers.

## Materials and methods

For a preliminary barcoding analysis we selected 1920 specimens representing 449 decapod species obtained mostly in the Indo-Pacific Ocean over the last 40 years and preserved in ethanol in the MNHN collections. The taxa selection is representative of the benthonic decapod diversity in the Indo-Pacific Ocean. We followed the MarBOL workflow methodology established at the MNHN and detailed in Puillandre *et al.* (2012), as well as the standard protocol of DNA sequencing in Barcoding projects for animals (Hebert *et al.* 2004). The DNA was extracted from a part of one pleiopod using the NucPrep (AB Applied Biosystem) or the NucleoSpin 96 Tissue (Macherey–Nagel) extraction kits, following the manufacturer's protocols. A preliminary screening with the Folmer primers (LCO1490 and HCO2198) resulted in only 805 barcodes (42%) from 254 species.

To test the hypothesis that the Folmer primers were not optimal for our target group as a whole, we aligned all COI sequences obtained from the complete decapod mitochondrial genomes available in GenBank. Then we compared the Folmer primer sequences with the COI alignment to verify the degree and pattern of mismatches.

To test a potential taxonomic bias in the primer mismatch, we verified whether any taxon-specific bias was associated with the amplification failure using a Chi-square test. As a null hypothesis we assumed that the proportion of barcodes obtained from each decapod infraorder (*sensu* De Grave *et al.* 2009) did not differ significantly from the mean success rate for all decapods.

With a Chi-square test we also verified whether the specimen age affects the barcode recovery rate using the Folmer primers, comparing the proportions of barcodes obtained from two age classes (0–20 years versus older).

Following these tests, all 805 barcodes obtained with the Folmer primers were aligned and the alignment was inspected by eye for conserved regions that were suitable for designing internal primers. The primers were checked for compatibility (primer dimer or hairpin formation and Tm) using the online

application NetPrimer ([www.premierbiosoft.com/netprimer/index.html](http://www.premierbiosoft.com/netprimer/index.html)). Then we tested the effectiveness of the new primer sets on the decapod specimens that failed to amplify with the Folmer primers.

The PCRs were conducted in 20- $\mu$ L reaction volume, containing 1–5 ng of DNA and to a final concentration of 1 $\times$  reaction buffer, 2.5 mM MgCl<sub>2</sub>, 0.26 mM dNTP, 0.3  $\mu$ M of each primer, 5% DMSO and 1.5 units of Qiagen Taq polymerase. For all primer pair combinations the amplification profile was: 5 min initial denaturation at 94°C, 35–40 cycles of 40 s at 94°C, 40 s at 48°C and 60 s at 72°C, followed by a final extension of 5 min at 72°C. PCR products were visualised on a 1.5% agarose gel stained with ethidium bromide and the positive PCRs were sequenced in both directions using the Sanger method.

As a case-study we illustrate the significant input provided by our specific primers in recovering a barcode library for the genus *Plesionika* Spence Bate, 1888 (Family Pandalidae). Pandalidae is the second most commercially important family of prawns after the Penaeidae. While the fishery for Penaeidae is mostly confined to the tropics and subtropics, that for Pandalidae is located in colder waters of both the Northern and Southern Hemispheres. All barcodes for *Plesionika* were aligned with Bioedit 7.0.9.0 (Hall 1999). The Neighbour-Joining tree was generated with PAUP\* (Swofford 2003) using the Kimura 2-parameter nucleotide-substitution model. The tree was rooted using three *Pandalus* species obtained from GenBank (accession numbers indicated in Fig. 1). The full specimen details and BOLD/GenBank accession numbers are supplied in Appendix 1.

## Results

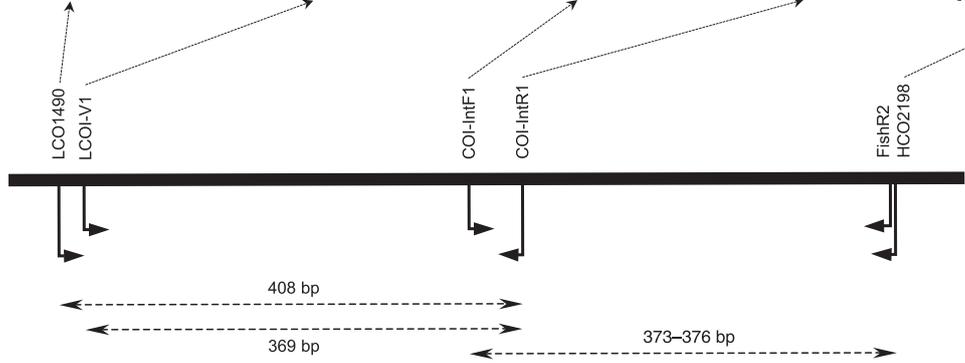
The alignment of COI sequences from the 28 complete mitochondrial genomes available in GenBank revealed that this gene is structurally conserved in decapods. It has a total length of 1534–1544 bp without any indels. The Folmer primers align well with the COI sequences, having, on average, only 4.1 and 3.3 mismatching bases, in the forward and reverse primers, respectively (Fig. 1, primers LCO1490 and HCO2198).

We did not detect any taxonomic bias in recovering the barcodes with the Folmer primers. While amplification success ranged between 26 and 66% across the different decapod infraorders (Fig. 2), the differences were not statistically significant ( $\chi^2 = 13.44$ , d.f. = 7,  $P > 0.05$ ). Conversely, sample age has a significant effect: we recovered barcodes with the Folmer primers from 45% of the more recent samples (<20 years) but from only 35% of the specimens collected >20 years ago ( $\chi^2 = 12.00$ , d.f. = 1,  $P < 0.01$ ).

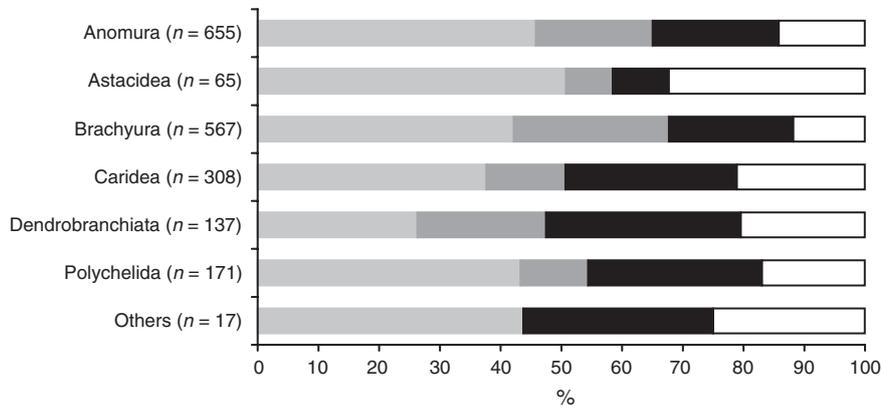
We designed two new internal primers specific for decapods (Table 1 and Fig. 1). One forward (COI-IntF1) and one reverse (COI-IntR1) primer are located near the centre of the barcoding region and they are intended to pair with the opposite Folmer primer. The barcode fragment of the COI gene is thus split into two overlapping fragments of ~350–400 bp each.

We observed that although the Folmer primers appear to be conserved within decapods as a whole, occasional mismatches near the 3' end of the primer might lower the annealing strength and significantly reduce the amplification success. Therefore we designed a new forward primer (LCOI-V1) annealing 39 bp downstream of LCO1490 which can be used as an alternative

Ac <i>Palinurus</i>	TT...A...AT.....	...C..A..A.....	.G..G..A.....C.....	.G.....C.....A..	..T..G.CC.C.ACA...A.....
Ac <i>Palinurus</i>	TT.....C.....	...C..A..A.....	..T..A.....C.....	.A.....C..C..G..	...G..C..C..A...A...C..
Ac <i>Palinurus</i>	TT.....C.....	...A..A.....	.G..A..G.....C..C.....	.G.....C..T..A..	...C..C..A...A...C..
An <i>Pagurus</i>	TC..T..T..C..C.....	..C..C..A.....	.G.....T.....	.....T.....	...C..C..A...A...C..
As <i>Cambaroides</i>	TT..T.....	.A..AA.....	..A.....	.....	..T.....
As <i>Cherax</i>	TT.....	.C..C..A..C..T.....	..T..G.....G.....	.....G..G..C..	...C.....G..G..G..C..
As <i>Homarus</i>	TT.....C.....	.C..C..A..A.....	..T..A..T.....	.....	...G.....A..T.....
As <i>Procambarus</i>	TT..T.....G.....	...A.....	.G..A..T.....G.....	.....G..A..	..T.....A..T.....
Br <i>Callinectes</i>	TT..T.....C.....	...A..A..T.....	..T..A..T.....	.....G..C..	..T.....C..A..T.....
Br <i>Charybdis</i>	TT..T.....C.....	..C..A.....T.....	.....	.....T..C..	..T.....C..A..T.....
Br <i>Eriocheir</i>	TT..C.....	..C..A.....	.....A..T..G.....	.....T..G..A..	..T.....C..A..T.....
Br <i>Gandalfus</i>	TT..T.....C.....	..C..A.....	.....T..G.....	.....C..A..	..T.....A.....
Br <i>Geothelphusa</i>	TT..C.....C.....C.....	..C..A.....	.....AA..T.....	.....C.....	..T.....G.....A.....
Br <i>Portunus</i>	TT..T.....	...A..A..T.....	.....	.....C..T..A..	..T.....A..T.....
Br <i>Pseudocarcinus</i>	TT..T.....C.....	...A..A.....	.G.....T.....	.....	..T.....A.....G.....
Br <i>Scylla</i>	TT..T.....C.....	..C..A..A..T.....	.G..A..G.....	.....C.....	..T.....C..G..T..A.....
Br <i>Xenograpsus</i>	TC..T.....	.A..C..G..C..G.....	..G..G..T.....G.....	.....A..C.....	..T.....C.....A.....G.....
Ca <i>Alpheus</i>	TT..T..T..C..C.....C.....	.A..C..G..C..G.....	..G..G..T.....G.....	.....A..C.....	..T.....C.....A.....G.....
Ca <i>Exopalaemon</i>	TT..C.....C.....	CC..A..A.....	..A..T.....	.A..G..C.....	..T.....A..T.....
Ca <i>Holocaridina</i>	TC..C..C..C.....	..C..C..A..A.....	..A..G..C.....	.T.....A..C.....	..T..G.....C..C..A...C..
Ca <i>Macrobrachium</i>	TC.....C.....	..C..C..A..A.....	..A..T..C.....	C.....C.....	..T..G..C.....C.....A..
De <i>Farfantepenaeus</i>	TT..T.....C.....	..C..C..G.....	.....G.....	.....	..T.....C.....A.....
De <i>Femneropenaeus</i>	TT..T.....C.....C.....	..C..A..C.....	.....G.....	.....C.....	..T.....C.....G..T..G..A..
De <i>Litopenaeus</i>	TT..T.....C.....C.....	..C..C..A.....	.....C.....	.....T.....	..T.....C.....A.....A.....
De <i>Marsupenaeus</i>	TC..T.....C.....C.....	..C..G..C.....	..A..T.....	.....A.....G..C.....	..T.....C..C..G...A..G..C..
De <i>Penaeus</i>	TT..T.....C.....C.....	..C..A.....	..A..G.....	.....T.....	..T.....C..C..T.....A.....
	GGTCAACAATCATAAAGATATTGG	TTTTTGGTGCTTGAGCNGGNATAGT	AAAGAGGTGTAGWACWGGWTGAAC	TYGCYCAYGAGGAGCTTC	TTCTGATTCTTCGGTCACCCCTGAAGT---



**Fig. 1.** Schematic representation of the primer positions on the COI gene. The amplicon lengths generated with the different combinations of internal primers are indicated below. The regions of the complete mitochondrial genomes alignment corresponding to the primer annealing sites are shown above, with the corresponding primer sequence at the bottom. For the three reverse primers the reverse complement sequence is shown. The two-letter code before the generic names indicates the major decapod subdivisions: Achelata, Anomura, Astacidea, Brachyura, Caridea, Dendrobranchiata. For improving the clarity some congeneric sequences have been omitted.



**Fig. 2.** Proportion of complete barcodes and minibarcodes recovered for each decapod subgroup. The sample size is indicated in parentheses. The heading 'Others' includes the few specimens of Stenopidea ( $n=3$ ) and Thalassinidea ( $n=14$ ). Light grey: barcodes obtained with the Folmer primers; dark grey: complete barcodes amplified with the specific primers; black: minibarcodes; white: missing barcodes, either full or mini.

to the latter. As an alternative to HCO2198 we selected the primer Fish-R2 (Ward *et al.* 2005), which aligns well with decapod sequences.

To maximise the effectiveness of the new primers we took care that they have similar melting temperatures and that they are all compatible, assuring that the same amplification conditions

**Table 1. Primer sequences used in this study with their annealing temperatures (T<sub>m</sub>, in °C)**

Primer name	Sequence	T <sub>m</sub>
COI-IntF1	AAAGAGGTGTAGGWACWGGWTGAAC	61
COI-IntR1	CTACTGAAGCTCCTGCRTGRCRA	65
LCO-V1	TTTTTGGTGCTTGAGCNGGNATAGT	67
LCO1490	GGTCAACAAATCATAAAGATATTGG	58
HCO2198	TAAACTTCAGGGTGACCAAAAAATCA	63
Fish-R2	ACTTCAGGGTGACCGAAGAATCAGAA	66

(Table 1) were effective for every possible pairing. All three new primers have two or three degenerate bases to compensate for variability of the 3rd codon position in the 3' half of the primer annealing site; the 1st and 2nd codon positions, though, appear highly conserved in the species analysed.

First we screened the specimens that failed to amplify with the Folmer primers using the primer pairs LCO1490/COI-IntR1 and COI-IntF1/HCO2198 (i.e. we employed one of the Folmer primers in combination with the respective internal primer specifically designed for the decapods). We obtained sequences from ~36% of samples from both primer combinations. Then with those samples that still failed to amplify we used the primer pairs LCO-V1/COI-IntR1 and COI-IntF1/Fish-R2, obtaining 26% and 23% new sequences, respectively. The success rate was rather homogeneous across the different decapod suborders (Table 2).

Using the new primers we sequenced 364 complete new barcodes and added 98 species to the barcode library obtained with the Folmer primers, while for an additional 443 specimens only partial barcodes were obtained, i.e. only one of the two fragments. Overall, we doubled the sequenced samples, from 42% to 84%, and increased the species coverage by 58%, from 254 to 397 barcoded taxa.

The example offered by the *Plesionika* shrimps illustrates clearly the effectiveness of our approach. In addition to the 25 individuals sequenced with the Folmer primers, our primer combinations generated 43 barcodes and account for 9 of the 20 barcoded species. Of the four barcodes recovered from type specimens, only one was sequenced with the Folmer primers (Fig. 3).

## Discussion

In over 200 years of activity, natural history museums have amassed an astonishing repository of the world's biological diversity. In terms of both taxonomic diversity and geographical coverage, museum collections offer unique opportunities for adding a molecular perspective to the traditional morphological approach in biodiversity investigations. At the same time, collections not established for molecular studies pose specific challenges for modern-day use (e.g. Martinková and Searle 2006; Lee and Prys-Jones 2007).

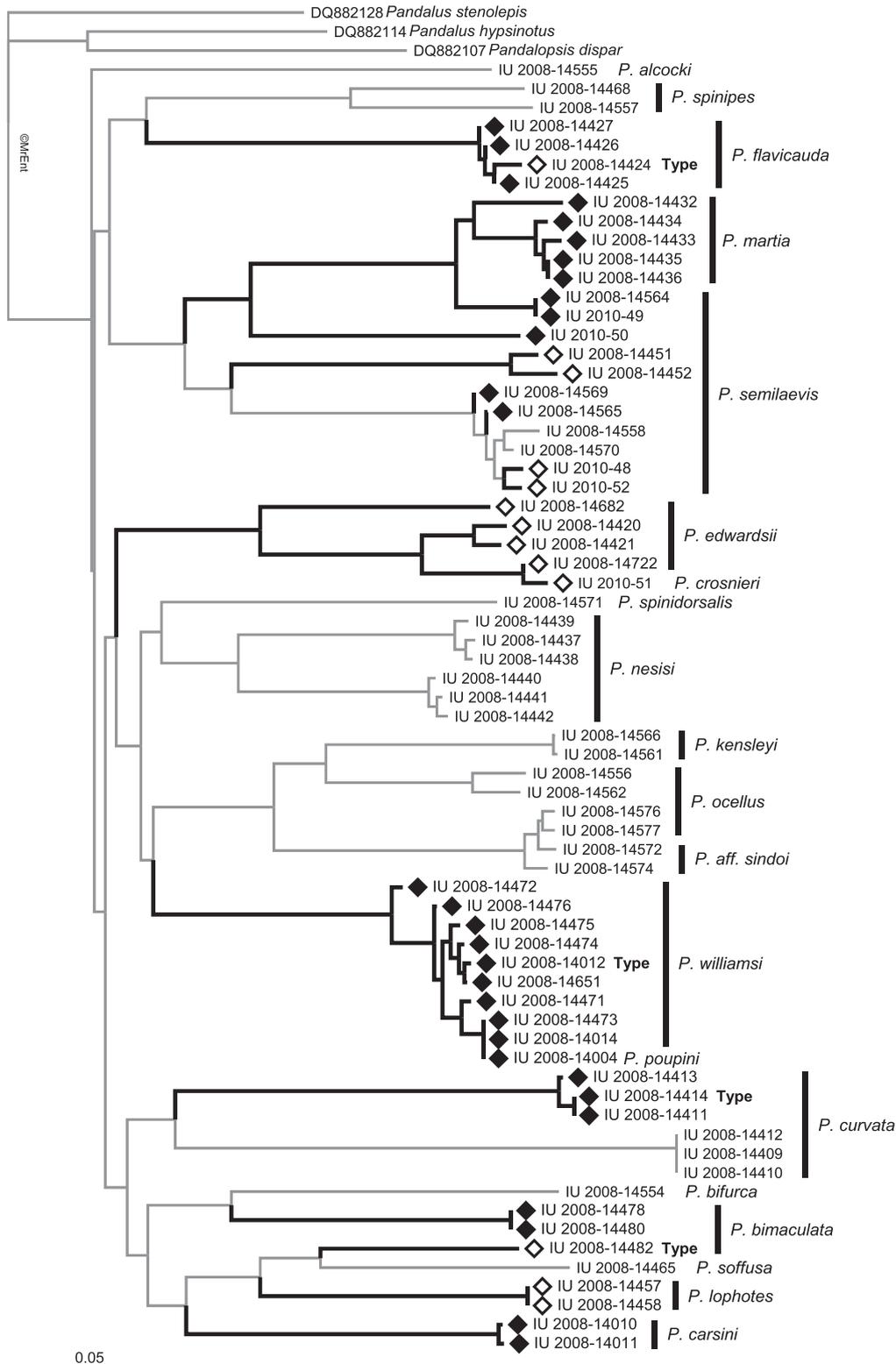
The age of the biological material and the preservation conditions are largely responsible for the degree of DNA damage (Dean and Ballard 2001; Mandrioli *et al.* 2006). Today the crustacean wet collection at the MNHN is stored in 75% ethanol in glass jars at 14°C in a temperature-stable room. Unfortunately, little is known of the post-collection fate of the individual specimens used in our study and this uncertainty is a recurrent problem in museum collection-based studies. Some material, especially from the first Tropical Deep-Sea Benthos campaigns, might have been briefly fixed in formalin before definitive preservation in ethanol. The ethanol concentration might have varied over time and the removal of specimens from the ethanol for study might have further affected the tissue preservation. Some of the material might also have experienced temperature fluctuation. All these factors are known to affect the DNA preservation and the chance of recovering amplicons (Dean and Ballard 2001; Mandrioli *et al.* 2006), but with little and mostly anecdotal information available it is impossible to specifically correlate the amplification success that we observed with the individual specimen preservation histories.

Despite all these possible shortcomings, the MNHN decapod collection remains an effective source of genetic material and it proved useful for establishing a barcoding library for this taxon. For fresh material or for tissue samples preserved under appropriate conditions, the Folmer primers are likely to be highly useful for the amplification of the COI gene in decapods. Indeed, they have been used successfully in several studies for barcoding freshly collected decapods e.g. Radulovici *et al.* (2009), Filipová *et al.* (2011) and Matzen da Silva *et al.* (2011). The exploitation of decapod specimens not originally intended for genetic studies is

**Table 2. Amplification success rates with the different primer combinations for the major decapod groups**

The first primer combinations for both fragments (Columns 3 and 5) were tested on those samples that failed to amplify with the Folmer primers (Column 2), and in turn the second primer combination for fragments (Columns 4 and 6) were used with those that failed with the previous attempts. Folmer, LCO1490/HCO2198; Folmer-intR, LCO1490/COI-IntR1; V1-intR, LCO-V1/COI-IntR1; intF-Folmer, COI-IntF1/HCO2198; intF-Fish, COI-IntF1/Fish-R2

Taxon	Folmer	Fragment 1		Fragment 2	
		Folmer-intR	V1-intR	intF-Folmer	intF-Fish
Anomura	45.8% (n = 655)	39.2% (n = 355)	29.6% (n = 216)	37.2% (n = 355)	23.8% (n = 223)
Astacidea	50.8% (n = 65)	18.8% (n = 32)	11.5% (n = 26)	15.6% (n = 32)	7.4% (n = 27)
Brachyura	42.0% (n = 567)	39.5% (n = 329)	29.6% (n = 199)	48.9% (n = 329)	30.4% (n = 168)
Caridea	37.7% (n = 308)	39.6% (n = 192)	29.3% (n = 116)	19.8% (n = 192)	12.3% (n = 154)
Dendrobranchiata	26.3% (n = 137)	33.7% (n = 101)	17.9% (n = 67)	33.7% (n = 101)	32.8% (n = 67)
Polychelida	43.3% (n = 171)	22.7% (n = 97)	17.3% (n = 75)	36.1% (n = 97)	27.4% (n = 62)
Stenopoidea	66.7% (n = 3)	0% (n = 1)			
Thalassinidea	42.9% (n = 14)	25% (n = 8)	16.7% (n = 6)	12.5% (n = 8)	14.3% (n = 7)



**Fig. 3.** The *Plesionika* shrimps illustrate how the optimised protocol adds more barcodes, more species and more barcode clusters to the barcoding library. The grey branches denote the topology generated by the barcodes recovered with the Folmer primers. The additional barcodes obtained with the internal primers are highlighted with open diamonds (full barcodes) or full diamonds (minibarcodes) and the resulting topology is in black. Name-bearing specimens are indicated with 'Type' after the MNHN accession numbers. The tree was edited in MrEnt 2.2. (Zuccon and Zuccon 2010).

still feasible, but may need *ad hoc* protocols like the one proposed here.

Because we invalidated the hypothesis that failure of amplification of the COI gene with the Folmer primers was linked to a mismatch of the primers with the decapod COI, our working-hypothesis was that, in several cases, DNA was degraded to fragments shorter than the 709-bp amplicon generated with the Folmer primers. The amplification of shorter, overlapping fragments circumvents the DNA shearing and is a widely followed strategy for recovering the full barcoding region from archival material (e.g. Hajibabaei *et al.* 2006). We favour a two-fragment approach, as it is a reasonable compromise between the additional laboratory effort and cost required, and the proportion of additional barcodes that can be retrieved. We were also interested in designing internal primers that would work across a wide set of taxa because it would enhance their utility in multiple contexts. Indeed, the two internal primers that we designed are compatible with all decapod groups, making them suitable even for barcoding samples with limited prior taxonomic information, such as processed food or larval stages.

Targeting fragments 350–400 bp long increases significantly the recovery success from archival specimens. With the two-fragment amplification strategy we were able to double the number of barcoding sequences and increase by 58% the number of species in the barcoding library. A quarter (443 specimens) of the barcodes obtained are actually minibarcodes, because only one of the two fragments was recovered from those specimens. Sequences of only 320–359 bp are clearly less optimal, but it must be remembered that when designing the internal primers we decided to limit their number to two to maximise the taxonomic inclusiveness of our protocol. While 45 species (11%) are represented only by minibarcodes, in all other cases these minibarcodes are additional sequences and at least one full-length barcode was obtained from another conspecific individual. Moreover, simulation studies showed that even 250-bp minibarcodes might provide enough information for a correct species identification (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008).

The strategy of using alternative primer pairs contributed significantly to the amplification success, accounting for 33 full barcodes and 126 minibarcodes. The Folmer primers are located in rather conserved regions in decapods, as in many other animal groups (Folmer *et al.* 1994). However, any large barcode study will inevitably encounter species that fail to amplify due to a mutation near the 3' end of the annealing region and the effects of primer mismatches are even more important in the case of damaged DNA such as those recovered from archival specimens (Van Houdt *et al.* 2010). The use of alternative primers should always be taken into account in the optimisation of barcoding protocols.

The specimens in the MNHN collection represent 68% of all species in the genus *Plesionika*. They have been critical for a taxonomic revision of the group and some new species have been described designating MNHN specimens as types (e.g. Crosnier 1986; Chan and Crosnier 1997; Chan 2004). The barcodes derived from the Folmer primers covered only part of the sampled *Plesionika* taxa. Our two-fragment amplification protocol significantly increases the genetic dataset (43 additional samples and nine species) but also provided a clearer picture of

the cryptic genetic diversity within *Plesionika* (Fig. 3). Our barcoding approach uncovered 28 distinct sequence clusters, which may represent cryptic species, amongst the 20 morphologically defined species-level taxa. Of these, 15 were recovered only using our specific primers. The sequencing of some of the type specimens provides an objective criterion to anchor available names to the correct barcoding cluster.

## Acknowledgements

We are grateful to Bertrand Richer de Forges and Philippe Bouchet, cruise leaders of several deep-sea cruises of the Tropical Deep-Sea Benthos program on board R/V Alis, that generated the deep-sea samples used in this study. All material has been collected under appropriate collection permits and approved ethics guidelines. The MarBOL project in the MNHN, Paris, is a joint effort with funding from (1) the Alfred P. Sloan Foundation; (2) the Consortium National de Recherche en Génomique and the Service de Systématique Moléculaire (UMS 2700 CNRS-MNHN), part of the agreement 2005/67 between the Genoscope and the Muséum National d'Histoire Naturelle on the project Macrophylogeny of Life; (3) the ATM 'Taxonomie moléculaire: DNA Barcode et gestion durable des collections' (MNHN); (4) the Fondation pour la Recherche sur la Biodiversité; and (5) the EDIT program. Sequencing was done as part of the project @SPEED-ID 'Accurate SPECiEs Delimitation and IDentification of eukaryotic biodiversity using DNA markers' proposed by F-BoL, the French Barcode of life initiative.

## References

- Bouchet, P., Héros, V., Lozouet, P., and Maestrati, P. (2008). A quarter-century of deep-sea malacological exploration in the South and West Pacific: Where do we stand? How far to go? In 'Tropical Deep-Sea Benthos 25'. *Mémoires du Muséum national d'Histoire naturelle* **196**, 9–40. (Eds V. Héros, R.H. Cowie and P. Bouchet.) (Paris.)
- Bucklin, A., Steinke, D., and Blanco-Bercial, L. (2011). DNA barcoding of marine metazoa. *Annual Review of Marine Science* **3**, 471–508. doi:10.1146/annurev-marine-120308-080950
- Castro, P. (2010). A new species and new records of palicooid crabs (Crustacea, Decapoda, Brachyura, Palicoidea, Palicidae, Crossotonotidae) from the Indo-West Pacific region. *Zoosystema* **32**, 73–86. doi:10.5252/z2010n1a3
- Chan, T. Y. (2004). The "*Plesionika rostricrescentis* (Bate, 1888)" and "*P. lophotes* Chace, 1985" species groups of *Plesionika* Bate, 1888, with descriptions of five new species (Crustacea: Decapoda: Pandalidae). In 'Tropical Deep-Sea Benthos 23'. (Eds B. Marshall, and B. Richer de Forges.) *Mémoires du Muséum national d'Histoire naturelle* **191**, 293–318, fig. 291–217.
- Chan, T. Y. (in press). A new genus of deep-sea solenocerid shrimp (Crustacea: Decapoda: Penaeoidea) from the Papua New Guinea. *Journal of Crustacean Biology*.
- Chan, T. Y., and Crosnier, A. (1997). Crustacea Decapoda: deep-sea shrimps of the genus *Plesionika* Bate, 1888 (Pandalidae) from French Polynesia, with descriptions of five new species. In 'Résultats des campagnes MUSORSTOM 18'. (Ed. A. Crosnier.) *Mémoires du Muséum national d'Histoire naturelle, Paris (A)* **176**, 187–234, fig. 181–141.
- Crosnier, A. (1986). *Plesionika fenneri*, nouveau nom pour *Plesionika chacei* Crosnier, 1986. *Bulletin du Muséum national d'Histoire naturelle, Paris, 4ème série (A)* **8**, 691.
- De Grave, S., Pentcheff, N. D., Ahyong, S. T., Chan, T.-Y., Crandall, K. A., Dworschak, P. C., Felder, D. L., Feldmann, R. M., Fransen, C. H. J. M., Goulding, L. Y. D., Lemaitre, R., Low, M. E. Y., Martin, J. W., Ng, P. K. L., Schweitzer, C. E., Tan, S. H., Tshudy, D., and Wetzer, R. (2009). A classification of living and fossil genera of decapod crustaceans. *The Raffles Bulletin of Zoology* **21**(Suppl.), 1–109.

- Dean, M. D., and Ballard, J. W. (2001). Factors affecting mitochondrial DNA quality from museum preserved *Drosophila simulans*. *Entomologia Experimentalis et Applicata* **98**, 279–283. doi:10.1046/j.1570-7458.2001.00784.x
- Filipová, L., Grandjean, F., Chucholl, C., Soes, D. M., and Petrussek, A. (2011). Identification of exotic North American crayfish in Europe by DNA barcoding. *Knowledge and Management of Aquatic Ecosystems* **401**, 11. doi:10.1051/kmae/2011025
- Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**, 294–299.
- Hajibabaei, M., Smith, A. M., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., and Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes* **6**, 959–964. doi:10.1111/j.1471-8286.2006.01470.x
- Hall, T. A. (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, P. D. N., Stoeckle, M. Y., Zemlak, T. S., and Francis, C. M. (2004). Identification of birds through DNA barcodes. *PLoS Biology* **2**(10), e312. doi:10.1371/journal.pbio.0020312
- Holthuis, L. B. (1980). F.A.O. species catalogue. I. Shrimps and prawns of the world. An annotated catalogue of species of interest to fisheries. *FAO Fisheries Synopsis* **125** (I), 271 pp.
- Housley, D. J. E., Zalewski, Z. A., Beckett, S. E., and Venta, P. J. (2006). Design factors that influence PCR amplification success of cross-species primers among 1147 mammalian primer pairs. *BMC Genomics* **7**, 253. doi:10.1186/1471-2164-7-253
- Komai, T. (2011). Further records of deep-sea shrimps of the genus *Glyphocrangon* (Crustacea: Decapoda: Caridea: Glyphocrangonidae) from the southwestern Pacific, with descriptions of two new species. *Species Diversity* **16**, 113–135.
- Lee, P. L. M., and Prys-Jones, R. P. (2007). Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. *Molecular Ecology Notes* **8**, 551–560.
- Mandrioli, M., Borsatti, F., and Mola, L. (2006). Factors affecting DNA preservation from museum-collected lepidopteran specimens. *Entomologia Experimentalis et Applicata* **120**, 239–244. doi:10.1111/j.1570-7458.2006.00451.x
- Martínková, N., and Searle, J. B. (2006). Amplification success rate of DNA from museum skin collections: a case study of stoats from 18 museums. *Molecular Ecology Notes* **6**, 1014–1017. doi:10.1111/j.1471-8286.2006.01482.x
- Matzen da Silva, J., Creer, S., dos Santos, A., Costa, A. C., Cunha, M. R., Costa, F. O., and Carvalho, G. R. (2011). Systematic and evolutionary insights derived from mtDNA COI barcode diversity in the Decapoda (Crustacea: Malacostraca). *PLoS ONE* **6**(5), e19449. doi:10.1371/journal.pone.0019449
- Meusnier, I., Singer, G., Landry, J.-F., Hickey, D., Hebert, P., and Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214. doi:10.1186/1471-2164-9-214
- Puillandre, N., Macpherson, E., Lambourdière, J., Cruaud, C., Boisselier-Dubayle, M.-C., and Samadi, S. (2011). Barcoding type specimens helps to identify synonyms and an unnamed new species in *Eumunida* Smith, 1883 (Decapoda: Eumunididae). *Invertebrate Systematics* **25**, 322–333. doi:10.1071/IS11022
- Puillandre, N., Bouchet, P., Boisselier-Dubayle, M.-C., Brisset, J., Buge, B., Castelin, M., Chagnoux, S., Christophe, T., Corbari, L., Lambourdière, J., Lozouet, P., Marani, G., Rivasseau, A., Silva, N., Terryn, Y., Tillier, S., Utge, J., and Samadi, S. (2012). The MarBOL workflow in the MNHN: tracking specimens from the field to the publication. *Molecular Ecology Resources* **12**, 396–402. doi:10.1111/j.1755-0998.2011.03105.x
- Radulovici, A. E., Sainte-Marie, B., and Dufresne, F. (2009). DNA barcoding of marine crustaceans from the Estuary and Gulf of St Lawrence: a regional-scale approach. *Molecular Ecology Resources* **9**(Suppl. 1), 181–187. doi:10.1111/j.1755-0998.2009.02643.x
- Swofford, D. L. (2003). 'PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4.' (Sinauer Associates: Sunderland, MA.)
- Van Houdt, J. K. J., Breman, F. C., Virgilio, M., and De Meyer, M. (2010). Recovering full DNA barcodes from natural history collections of tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular Ecology Resources* **10**, 459–465. doi:10.1111/j.1755-0998.2009.02800.x
- Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R., and Hebert, P. D. N. (2005). DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **360**, 1847–1857. doi:10.1098/rstb.2005.1716
- Zimmermann, J., Hajibabaei, M., Blackburn, D., Hanken, J., Cantin, E., Posfai, J., and Evans, T. C., Jr (2008). DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology* **5**, 18. doi:10.1186/1742-9994-5-18
- Zuccon, A., and Zuccon, D. (2010). MrEnt v.2.2. Program distributed by the authors. Available at <http://www.mrent.org>

Appendix 1. Specimen data for the genus *Plesionika* included in Fig. 1, with the primer combinations used for each specimen

Barcode: F, full barcode; M, minibarcode (&lt;400 bp)

Taxon	MNHN Number	BOLD ID	GenBank Accession No.	Origin	Barcode	Primer combinations
<i>Plesionika</i> aff. <i>sindoi</i>	IU 2008–14572	MDECA459–10	JX681737	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika</i> aff. <i>sindoi</i>	IU 2008–14574	MDECA461–10	JX681736	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika alcocki</i>	IU 2008–14555	MDECA448–10	JX681738	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika bifurca</i>	IU 2008–14554	MDECA447–10	JX681739	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika bimaculata</i>	IU 2008–14478	MDECA756–12	JX681740	New Caledonia	M	LCO-V1, COI-IntR1
<i>Plesionika bimaculata</i>	IU 2008–14480	MDECA770–12	JX681742	New Caledonia	M	LCO-V1, COI-IntR1
<i>Plesionika bimaculata</i>	IU 2008–14482	MDECA740–12	JX681741	Madagascar	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika carsini</i>	IU 2008–14010	MDECA767–12	JX681743	Tuamotu, Maria	M	LCO-V1, COI-IntR1
<i>Plesionika carsini</i>	IU 2008–14011	MDECA750–12	JX681744	Taiwan	M	LCO-V1 COI-IntR1
<i>Plesionika crosnieri</i>	IU 2010–51	MDECA742–12	JX681745	Madagascar	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika curvata</i>	IU 2008–14409	MDECA789–12	JX681749	Australes, Rapa	F	LCO1490, HCO2198
<i>Plesionika curvata</i>	IU 2008–14410	MDECA782–12	JX681750	Australes, Rapa	F	LCO1490, HCO2198
<i>Plesionika curvata</i>	IU 2008–14411	MDECA763–12	JX681751	Australes, Rurutu	M	LCO-V1, COI-IntR1
<i>Plesionika curvata</i>	IU 2008–14412	MDECA786–12	JX681746	Australes, Tubai	F	LCO1490, HCO2198
<i>Plesionika curvata</i>	IU 2008–14413	MDECA751–12	JX681748	Australes, Tubai	M	LCO-V1, COI-IntR1
<i>Plesionika curvata</i>	IU 2008–14414	MDECA774–12	JX681747	Australes, Tubai	M	LCO-V1, COI-IntR1
<i>Plesionika edwardsii</i>	IU 2008–14420	MDECA760–12	JX681752	Guinea	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika edwardsii</i>	IU 2008–14421	MDECA744–12	JX681755	Guinea	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika edwardsii</i>	IU 2008–14682	MDECA743–12	JX681753	Australes, Tubai	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika edwardsii</i>	IU 2008–14722	MDECA757–12	JX681754	La Réunion	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika flavicauda</i>	IU 2008–14424	MDECA775–12	JX681757	New Caledonia	F	LCO-V1 COI-IntR1 and COI-IntF1 Fish-R2
<i>Plesionika flavicauda</i>	IU 2008–14425	MDECA748–12	JX681756	Actaeons	M	LCO-V1, COI-IntR1
<i>Plesionika flavicauda</i>	IU 2008–14426	MDECA764–12	JX681758	Actaeons	M	LCO-V1, COI-IntR1
<i>Plesionika flavicauda</i>	IU 2008–14427	MDECA772–12	JX681759	Tuamotu, Mururoa	M	LCO-V1, COI-IntR1
<i>Plesionika kensleyi</i>	IU 2008–14561	MDECA452–10	JX681760	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika kensleyi</i>	IU 2008–14566	MDECA454–10	JX681761	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika lophotes</i>	IU 2008–14457	MDECA773–12	JX681762	Philippines	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika lophotes</i>	IU 2008–14458	MDECA752–12	JX681763	Philippines	F	LCO-V1, COI-IntR1 and COI-IntF1, Fish-R2
<i>Plesionika martia</i>	IU 2008–14432	MDECA776–12	JX681766	Gulf of Gascogne	M	LCO-V1, COI-IntR1
<i>Plesionika martia</i>	IU 2008–14433	MDECA777–12	JX681765	Congo	M	LCO-V1, COI-IntR1
<i>Plesionika martia</i>	IU 2008–14434	MDECA745–12	JX681764	Congo	M	LCO-V1, COI-IntR1
<i>Plesionika martia</i>	IU 2008–14435	MDECA778–12	JX681768	Guinea	M	LCO-V1, COI-IntR1
<i>Plesionika martia</i>	IU 2008–14436	MDECA768–12	JX681767	Guinea	M	LCO-V1, COI-IntR1
<i>Plesionika nesisi</i>	IU 2008–14437	MDECA783–12	JX681769	Samoa, Lallarokkhe	F	LCO1490, HCO2198
<i>Plesionika nesisi</i>	IU 2008–14438	MDECA784–12	JX681773	Samoa, Lallarokkhe	F	LCO1490, HCO2198
<i>Plesionika nesisi</i>	IU 2008–14439	MDECA781–12	JX681770	Samoa, Lallarokkhe	F	LCO1490, HCO2198
<i>Plesionika nesisi</i>	IU 2008–14440	MDECA787–12	JX681774	Tuamotu, Fangataufa	F	LCO1490, HCO2198
<i>Plesionika nesisi</i>	IU 2008–14441	MDECA790–12	JX681771	Tuamotu, Fangataufa	F	LCO1490, HCO2198
<i>Plesionika nesisi</i>	IU 2008–14442	MDECA785–12	JX681772	Tuamotu, Fangataufa	F	LCO1490, HCO2198
<i>Plesionika ocellus</i>	IU 2008–14556	MDECA449–10	JX681778	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika ocellus</i>	IU 2008–14562	MDECA453–10	JX681777	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika ocellus</i>	IU 2008–14576	MDECA462–10	JX681776	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika ocellus</i>	IU 2008–14577	MDECA463–10	JX681775	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika poupini</i>	IU 2008–14004	MDECA769–12	JX681779	Tuamotu, Makemo	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2008–14451	MDECA746–12	JX681781	Australes, Tubai	F	LCO-V1 COI-IntR1 and COI-IntF1 HCO2198
<i>Plesionika semilaevis</i>	IU 2008–14452	MDECA739–12	JX681784	Australes, Tubai	F	LCO-V1, COI-IntR1 and COI-IntF1, HCO2198

(continued next page)

## Appendix 1. (continued)

Taxon	MNHN Number	BOLD ID	GenBank Accession No.	Origin	Barcode	Primer combinations
<i>Plesionika semilaevis</i>	IU 2008–14558	MDECA451–10	JX681787	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika semilaevis</i>	IU 2008–14564	MDECA762–12	JX681786	Mozambique Channel	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2008–14565	MDECA741–12	JX681782	Mozambique Channel	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2008–14569	MDECA765–12	JX681790	Mozambique Channel	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2008–14570	MDECA457–10	JX681783	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika semilaevis</i>	IU 2010–48	MDECA771–12	JX681788	Madagascar	F	LCO-V1, COI-IntR1 <i>and</i> COI-IntF1, HCO2198
<i>Plesionika semilaevis</i>	IU 2010–49	MDECA749–12	JX681780	Madagascar	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2010–50	MDECA761–12	JX681785	Madagascar	M	LCO-V1, COI-IntR1
<i>Plesionika semilaevis</i>	IU 2010–52	MDECA766–12	JX681789	Madagascar	F	LCO-V1, COI-IntR1 <i>and</i> COI-IntF1, HCO2198
<i>Plesionika spinidorsalis</i>	IU 2008–14571	MDECA458–10	JX681791	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika spinipes</i>	IU 2008–14468	MDECA437–10	JX681793	Coral Sea	F	LCO1490, HCO2198
<i>Plesionika spinipes</i>	IU 2008–14557	MDECA450–10	JX681792	Mozambique Channel	F	LCO1490, HCO2198
<i>Plesionika suffusa</i>	IU 2008–14465	MDECA788–12	JX681794	Norfolk Ridge	F	LCO1490, HCO2198
<i>Plesionika williamsi</i>	IU 2008–14012	MDECA738–12	JX681803	Guadeloupe	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14014	MDECA753–12	JX681797	Marquises, Tahuata	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14471	MDECA758–12	JX681796	Seychelles	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14472	MDECA779–12	JX681799	Guadeloupe	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14473	MDECA754–12	JX681802	Marquises, Eiao	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14474	MDECA759–12	JX681795	Senegal	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14475	MDECA755–12	JX681800	Guinea	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14476	MDECA780–12	JX681801	Guinea	M	LCO-V1, COI-IntR1
<i>Plesionika williamsi</i>	IU 2008–14651	MDECA747–12	JX681798	Congo	M	LCO-V1, COI-IntR1