

Redesign of PCR primers for mitochondrial cytochrome *c* oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys

J. GELLER,* C. MEYER,† M. PARKER† and H. HAWK*¹

*Moss Landing Marine Laboratories, 8272 Moss Landing Road, Moss Landing CA 95309, USA, †Department of Invertebrate Zoology, Smithsonian Institution, National Museum of Natural History, Washington DC 20013-7012, USA

Abstract

DNA barcoding is a powerful tool for species detection, identification and discovery. Metazoan DNA barcoding is primarily based upon a specific region of the cytochrome *c* oxidase subunit I gene that is PCR amplified by primers HCO2198 and LCO1490 ('Folmer primers') designed by Folmer *et al.* (*Molecular Marine Biology and Biotechnology*, 3, 1994, 294). Analysis of sequences published since 1994 has revealed mismatches in the Folmer primers to many metazoans. These sequences also show that an extremely high level of degeneracy would be necessary in updated Folmer primers to maintain broad taxonomic utility. In primers jgHCO2198 and jgLCO1490, we replaced most fully degenerated sites with inosine nucleotides that complement all four natural nucleotides and modified other sites to better match major marine invertebrate groups. The modified primers were used to amplify and sequence cytochrome *c* oxidase subunit I from 9105 specimens from Moorea, French Polynesia and San Francisco Bay, California, USA representing 23 phyla, 42 classes and 121 orders. The new primers, jgHCO2198 and jgLCO1490, are well suited for routine DNA barcoding, all-taxon surveys and metazoan metagenomics.

Keywords: biotic surveys, cytochrome *c* oxidase subunit I, DNA barcoding, Moorea, universal primers

Received 20 March 2013; revision received 22 May 2013; accepted 5 June 2013

Introduction

The mitochondrial cytochrome *c* oxidase subunit I gene (COI) has been used extensively for studies of population genetics, phylogeography, speciation and systematics. For many species and genera, genetic variation at this locus is sufficient to study processes that occur over relatively short and recent time intervals. Despite this variation, some regions of the gene are sufficiently conserved to design primers for the polymerase chain reaction (PCR) that match a broad spectrum of organisms (Hebert *et al.* 2003; Kress & Erickson 2012). Consequently, for many studies, COI occupies a 'sweet spot' of variation allowing for meaningful population and interspecies studies while conserved enough for practicality. Two primers (LCO1490 and HCO2198), commonly referred to as the 'Folmer primers', have been used extensively (Folmer *et al.* 1994). Query of citation

databases at the Web of Science on 21 February 2013 revealed 2967 citations of the Folmer *et al.* (1994) paper.

DNA barcoding, the use of diagnostic nucleotide variation to identify species, is a further development that has used the COI gene as a primary tool (Savolainen *et al.* 2005; Stoeckle & Hebert 2008; Ward *et al.* 2009; Bucklin *et al.* 2011). Where interspecific variation does not overlap intraspecific variation, species can be reliably identified by DNA sequences. Studies have shown that this condition often, but not always, applies to COI. DNA barcoding, therefore, includes a degree of uncertainty (Meyer & Paulay 2005). The method both relies upon and is improved by large data sets, and formal protocols for DNA barcoding have been proposed to promote uniformity of data quality (Ratnasingham & Hebert 2007; Kress & Erickson 2012). For animal taxa, the region of COI flanked by the Folmer primers has been designated as a DNA barcode region (barcodeoflife.org), and records with sufficient metadata are given the keyword BARCODE by GenBank.

Despite the popularity and success of the Folmer primers described above, they are not truly 'universal' in applicability. Our own experience and informal

Correspondence: Jonathan Geller, Fax: 1-831-632-4403; E-mail: geller@mml.calstate.edu

¹Present address: Département de Biologie, Université Laval, 1045 Avenue de la Médecine, Québec QC G1V 0A6, Canada

conversations with colleagues indicate that the Folmer primers often fail or perform poorly, producing faint products despite attempts at optimization. Failures of the Folmer primers are at least in part due to mismatches with the target annealing position for many taxa. Relatively few full-length COI gene sequences were available in 1994, when the primers were published. Present-day analysis of full-length COI sequences often reveals mismatches with the primers. For this reason, a common strategy has been to obtain a few sequences using Folmer primers under nonstringent annealing temperatures, with additives, or with reamplification of weak products, then to design new primers specific for the particular study (e.g. lepidopterans: Hebert *et al.* 2004; fishes: Ward *et al.* 2005; bryozoans: Mackie *et al.* 2006). This primer customization is tolerable for focal taxon studies, but workflows for major biodiversity surveys, routine DNA barcoding or identification of unknowns preclude frequent primer redesign.

More universal COI primers would be useful for biodiversity and barcoding studies. The existing body of data and the literature for the fragment flanked by the Folmer primers places a positional constraint on primer design. A few alternative primers have been suggested. Meyer (2003) redesigned the Folmer primers, using sequences in the Folmer *et al.* (1994) paper to make them degenerate in the 3' region (dgLCO1490 and dgHCO2198). Meusnier *et al.* (2008) proposed 'mini barcode' primers that amplify an internal fragment of the standard barcode region from a broader swath of taxa at the cost of sequence length. A 21 February 2013 survey of the BOLDSystems public primer database, the repository for DNA barcoding primers maintained by the Barcode of Life Database, revealed minimally 418 different primers targeting the COI gene for various taxa. An alternative to the Folmer primers is thus highly desirable.

Fortunately, mitochondrial genomics has flourished in recent years, and over 3000 complete or nearly complete mitochondrial genomes were known in 2011 from 28 phyla (<http://mi.caspar.it/mitozoa>) (Lupi *et al.* 2010). This allows more comprehensive alignments of the COI gene and identification of mismatches in the region targeted by the Folmer primers. These alignments are the basis for the design of new or improved primers for the COI region. The molecular interactions involved in primer annealing are complex and involve nucleotide complementarity to the target, primer homoduplex and heteroduplex formation, potential for secondary structures in incompletely denatured template, ionic strength of the PCR buffer and nucleotide-neighbour effects in the target region. Ideal primer design would include an accurate model for these effects, and several computer

programs exist that include algorithms for some of these factors [e.g. PRIMER3 (Rozen & Skaletsky 2000), PRIMER PREMIER 6 (Premier Biosoft, Palo Alto, CA, USA), AMPLICON (Jarman 2004)]. However, these computer programs usually are meant to design primers for single sequences or for groups of similar sequences. Design of primers for alignments of highly divergent sequences is a more difficult task.

In this study, we describe changes to the standard Folmer primers that were meant to correct mismatches for many marine invertebrate species. In alignments of COI, variation in 3rd codon positions is extensive, and no primer of 20–30 bp can be suggested that is 100% conserved across animal phyla. However, degeneracy in the PCR primer can accommodate this variation in the priming region. Degeneracy is created during primer synthesis by mixing nucleotides at the variable sites, thereby creating a pool of primers containing all variants. This has the downside of diluting the effective primers: only a small proportion of the primer mix will be an exact match to any template, and many primers in the pool will poorly match the target sequence. Further, sequence variation in the primer pool makes it difficult to predict interprimer interactions or the potential for mispriming. A different approach is to use inosine nucleotides (dITP) at variable positions. Inosine nucleotides form pair bonds with all natural nucleotides, thus increasing potential target sequences (PTS) without increasing degeneracy.

In this study, we suggest a more universal version of the Folmer primers using degenerate positions and internal inosines and show its applicability to biotic surveys in Moorea, French Polynesia and San Francisco Bay, CA. We conclude that the redesigned primers are broadly applicable and complement the standard Folmer primers in DNA barcoding applications.

Methods

Cytochrome *c* oxidase subunit I sequences for representative marine invertebrate taxa were acquired from GenBank using available complete mitochondrial genomes in 2009. Sequences were aligned in Geneious (Biomatters, New Zealand), and the consensus of nucleotides at the positions that correspond to LCO1490 and HCO2198 was determined. Positions with fourfold degeneracy were replaced with dITP. Positions with twofold degeneracy were synthesized with mixed nucleotides to create a primer pool. Resulting primers were named jgLCO1490 and jgHCO2198 to make the relationship to the original Folmer primers explicit.

All morphologically discernable invertebrate species were collected from settling plates placed in San Francisco Bay, California quarterly in 2010 for 4 months.

Tissues were subsampled, preserved and stored in 90% ethanol for about 2 months and extracted with a DNeasy Tissue kit (Qiagen, Germantown, MD, USA). Templates from (mostly) invertebrates from Moorea, French Polynesia, were prepared from unpreserved tissues using the Qiagen Biosprint or Autogen (Holliston, MA, USA) apparatus and reagents. For estimation of PCR success in this study, we did not include templates that had been previously tested with other primers. For comparison to other primers, we included a subset of templates that had previously failed with both Folmer and dgLCO1490/dgHCO2198 using published PCR protocols (Folmer *et al.* 1994; Meyer 2003).

PCR conditions for jgLCO1490/jgHCO2198 were developed independently for the Moorea and San Francisco Bay samples. PCR with the San Francisco samples was prepared with 25 μL of Promega Green GoTaq 2X master mix, augmented with MgCl_2 to a final concentration of 3.5 mM, 0.2 μM each primer and 1 μL of genomic DNA in a final volume of 50 μL . PCR conditions were 94 °C for 2 min followed by 30 cycles of 94 °C-1 min, 48 °C-1 min and 72 °C-1 min. PCR products were examined on 1.2% agarose gels stained with ethidium bromide. Reactions producing strong, single bands of the expected size were set aside. Remaining PCR mixes were replaced in the thermocycler for five additional cycles and checked again by agarose gel electrophoresis. A subset of PCR products was shipped to Elim Biopharmaceuticals, Inc. (Hayward, CA, USA) for purification and sequencing.

PCR for Moorean specimens was prepared with 2 μL 10 \times PCR buffer (Bioline, Taunton, MA, USA) 0.2 μL (1 unit) of Biolase *Taq* polymerase (Bioline), 2 mM MgCl_2 , 0.3 μM of each primer, 0.5 μM dNTP and 1 μL of genomic DNA in 20 μL reactions. PCR conditions were an initial 5 min at 95 °C; then 35 cycles of 30 s at 95 °C, 30 s at 48 °C; 45 s at 72 °C and a final 5 min at 72 °C. PCR products were examined on a 1% agarose gel stained with ethidium bromide. Successful products were treated with the ExoSAP kit (Affymetrix, Santa Clara, CA, USA) and cycle sequenced following standard protocols.

Results

The primer LCO1490 contained mismatches to invertebrate COI in the 3rd and 6th nucleotide from the important 3' terminus, as well as in the 5' end (Table 1). The fully degenerated consensus of all LCO1490 sequences in Table 1 would require 3 869 835 264 oligonucleotides, as variation exists at all but four sites. Through use of internal dITP and selected degenerate sites, the primer pool was reduced to 16 oligonucleotides while maintaining the same range of PTS. The primer HCO2198 also has

variation in the 3rd and 6th positions internal to the 3' end. Otherwise, this primer is well matched to most invertebrate COI sequences (Table 1). A consensus of the HCO2198 priming sites in the taxa included predicts up to 4096 PTS. Our redesigned primer is 32-fold degenerate and complements all these PTS.

Using the primers jgLCO1490 and jgHCO2198 for PCR, sequences were obtained from 9105 specimens (1419 from San Francisco Bay and 7686 from Moorea) from 23 phyla, 42 classes and 121 orders of mostly marine invertebrates. Exact or near exact (99–100% identical) matches to COI sequences in GenBank demonstrate that the primers correctly amplified COI [e.g. Cirripedia: *Balanus improvisus* (GenBank accessions FJ845843), Bryozoa: *Bugula neritina* (AY633485), Chordata: *Botryllus schlosseri* (JN083241), Gastropoda: *Ilyanassa obsoleta* (GQ129488)]. Table 2 presents the number of sequences from mostly invertebrate taxa that were successfully amplified and sequenced in the Moorea Biocode Project, and Table 3 lists invertebrate genera amplified and sequenced from the San Francisco fouling community.

Primer performance

PCR success, meaning a single PCR product of the expected size, for all samples attempted was 72.5% (of 8612 PCR) and 71.1% (of 4043 attempts) for Moorean and San Franciscan samples, respectively. In the Moorea project, among phyla with 50 or more specimens, success varied from 85% for Sipuncula to 56% for Platyhelminthes. Phoronida, Echiura and Chaetognatha had 100% success albeit with far fewer specimens. Entoprocta (one success of seven templates) and Nematoda (4 of 14) showed poor results. The only total failures were seen with Ciliophora and Nematomorpha, for which attempts were made with two different templates each.

For San Francisco Bay phyla with 50 or more specimens, success varied from 50% for Cnidaria to 80% for Bryozoa. Success was about 70% for all Arthropoda (mainly barnacles and peracarids), but varied more within other phyla. Among Mollusca, for example, Gastropoda were amplified with 85% success, while Bivalvia registered only 54% success. For cnidarians, Hydrozoa (68%) amplified more consistently than Anthozoa (54%). In contrast to results from Moorea, six of eight entoproct specimens (75%) from San Francisco Bay were successfully amplified. Another way to judge success is sequence recovery for each species attempted. In Moorea, a majority of specimens were not identified to the species level, but this analysis was possible for San Francisco Bay where taxonomists identified 160 distinct morphospecies among the collections. We obtained at least one COI sequence from 146 (92%) of these morphospecies.

Contaminant and nonspecific amplification

From San Francisco Bay, 19 species had at least one PCR attempt (2% of all PCR) that produced two discrete PCR products, one of the expected sizes and one smaller. All but two of these species produced, in other reactions, single products that were sequenced and verified. Smear PCR products were obtained in 6% of all PCR attempts and were counted as failures. Many of these appeared to be results of over cycling or excess template, indicating that further PCR optimization was possible. From San Francisco Bay, 25 sequences (which were not counted in COI totals) from six species of ascidians, two species of bryozoans and one sponge were bacterial in origin. Twelve bacterial sequences were from *Ciona savignyi*, for which eight other templates produced a correct sequence. Sixteen sequences of 1419 (1%), mostly from bryozoans and tunicates, could not be identified to phylum

in BLAST searches of GenBank. Fourteen sequences (1%) were obvious mismatches with the morphological identification. As examples, one bryozoan specimen yielded a caprellid amphipod sequence, two specimens identified as solitary tunicates had sequences that matched a compound tunicate sequence and five compound tunicate specimens produced solitary tunicate sequences.

Performance of Folmer primers and revised primers

In an earlier phase, the Moorea Biocode Project used the Folmer primers in 2095 PCR attempts, with a success rate of 44%, lower than the 72.5% observed with the jgLCO1490/jgHCO2198 primers. A total of 1422 of 3412 templates that failed with both the Folmer primers and the dgLCO1490/dgHCO2198 primer set were amplified with the jgLCO1490/jgHCO2198 and sequenced, a 42% recovery of sequence.

Table 1 Consensus sequences from alignments in 2010 of cytochrome *c* oxidase subunit I from various taxa corresponding to LCO1490 and HCO2198 of Folmer *et al.* (1994), consensus of consensus sequences and a new primer set, jgLCO1490 and jgHCO2198. Mitochondrial genomes and GenBank accession numbers used for alignments: Nematoda: *Ancylostoma duodenale* (NC_003415), *Ascaris suum* (NC_001327), *Brugia malayi* (NC_004298), *Caenorhabditis elegans* (NC_001328), *Dirofilaria immitis* (NC_005305), *Necator americanus* (NC_003416), *Onchocerca volvulus* (NC_001861), *Steinernema carpocapsae* (NC_005941), *Trichinella spiralis* (NC_002681). Bivalvia: *Argopecten irradians* (NC_009687), *Chlamys farreri* (NC_012138), *Crassostrea gigas* (NC_001276), *Crassostrea virginica* (NC_007175), *Hiattella arctica* (NC_008451), *Lucinella divaricata* (NC_013275), *Mytilus edulis* (NC_006161), *Mytilus galloprovincialis* (NC_006886), *Mytilus trossulus* (NC_007687), *Venustaconcha ellipsiformis* (NC_013659). Crustacea: *Artemia franciscana* (NC_001620), *Charybdis japonica* (NC_013246), *Eriocheir hepuensis* (NC_011598), *Eriocheir japonica* (NC_011597), *Exopalaemon carinicauda* (NC_012566), *Farfantepenaeus californiensis* (NC_012738), *Gandalfus yunohana* (NC_013713), *Litopenaeus stylirostris* (NC_012060), *Macrobrachium lanchesteri* (NC_012217), *Oratosquilla oratoria* (NC_014342), *Pagurus longicarpus* (NC_003058), *Paracyclopsina nana* (NC_012455), *Panulirus stimpsoni* (NC_014339), *Scylla olivacea* (NC_012569), *Scylla serrata* (NC_012565), *Scylla paramamosain* (NC_012572), *Scylla tranquebarica* (NC_012567), *Triops cancrivorus* (NC_004465), *Xenograpsus testudinatus* (NC_013480). Entoprocta: *Loxocorone allax* (NC_010431), *Loxosomella aloxiata* (NC_010432). Annelida: *Clymenella torquata* (NC_006321), *Nephtys* sp. (NC_010559), *Orbinia latreillii* (NC_007933), *Perionyx excavatus* (NC_009631), *Pista cristata* (NC_011011), *Platynereis dumerilii* (NC_000931), *Urechis unicinctus* (NC_012768), *Terebellides stroemi* (NC_011014), *Whitmania pigra* (NC_013569). Bryozoa: *Bugula neritina* (NC_010197.1), *Flustrellidra hispida* (NC_008192.1), *Watersipora subtorquata* (NC_011820.2). Nemertea: *Cephalothrix simula* (NC_012821), *Cephalothrix* sp. (NC_014869), *Lineus viridis* (NC_012889), *Paranemertes cf. peregrina* (NC_014865). Urochordata: *Aplidium conicum* (NC_013584), *Ciona intestinalis* (NC_004447.2), *Ciona savignyi* (NC_004570.1), *Clavelina lepadiformis* (NC_012887), *Diplosoma listerianum* (NC_013556), *Doliolum nationalis* (NC_006627), *Halocynthia roretzi* (NC_002177.1), *Herdmania momus* (NC_013561.1), *Microcosmus sulcatus* (NC_013752), *Phallusia fumigata* (NC_009834), *Phallusia mammillata* (NC_009833), *Styela plicata* (NC_013565.1). Platyhelminthes: *Benedenia hoshinai* (NC_014591), *Clonorchis sinensis* (NC_012147), *Echinococcus canadensis* (NC_011121), *Echinococcus ortleppi* (NC_011122), *Gyrodactylus salaris* (NC_008815), *Opisthorchis felinus* (NC_011127), *Spirometra erinaceieuropaei* (NC_011037), *Symsagittifera roscoffensis* (NC_014578), *Taenia multiceps* (NC_012894), *Taenia pisiformis* (NC_013844)

LCO1490	GGTCAACAAATCATAAAGATATTGG	HCO2198	TAAACTTCAGGGTGACCAAAAAATCA
Nematoda	VDDSWGTDAAAYCAYAARRMWATYGG		TAHACYTCWGGRTGHCCRAARAAYCA
Bivalvia	DDDSNWVHWMHCAYHDWGAYRTHGG		TANACYTCHGGRTGVCCRAARAAYCA
Crustacea	WYTCHWSDAAYCAYAARGAYATTGG		TANACYTCNGGRTGNCCRAARAAYCA
Decapoda	TYTCHACWAAYCAYAARGAYATTGG		TANACTTCDGGRTGNCCRAARAAYCA
Entoprocta	TTTCAACAAATCATAAAGATATTGG		TAMACTTCWGGRTGACCAAAAAAYCA
Annelida	WYTCWACHAAHCAYAAAGAYATTGG		TADACYTCDGGRTGNCCRAARAAYCA
Bryozoa	TATCWACWAAYCACAARGACATTGG		TAWACTTCKGGGTGTCCAAAAAYCA
Nemertea	WTTWCWAATCATAARGATATTGG		TAMACYTCAGGRTGWCCAAAAAYCA
Urochordata	TDTCDACNAAYCATAARGAYATYRG		TANRCYTCNGGRTGNCYRAARARYCA
Platyhelminthes	TNACTNYNGAHCAYAAAGSGTATHGG		TANACYTCNGGRTGNCCRAARAAYCA
Consensus	NNDSNYNNDMHCAYHDDVVHRTHRG		TANACYTCNGGRTGNCYRAARAAYCA
jgLCO1490	TITCIACIAAYCAYAARGAYATTGG	jgHCO2198	TAIACYTCIGGRTGICRAARAAYCA

Table 2 Phyla, Class and Order of specimens successfully amplified and sequenced within the Moorea Biocode Project. We report sequences from all PCR that used the jgLCO1490/jgHCO2198 regardless of prior success or failure with other primers. Resolution of lower levels of identification was uneven; thus, numbers of sequences at higher levels may exceed the sum of those identified to lower levels

Phylum	Class	Order	Number of sequences
Totals			
23	42	102	7686
Porifera			283
	Demospongiae		52
		Hadromerida	6
		Verongida	22
		Chondrosida	2
		Dendroceratida	4
		Dictyoceratida	18
		Agelasida	6
		Haplosclerida	4
	Calcarea	Clathrinida	8
Cnidaria			546
	Anthozoa		256
		Actiniaria	84
		Alcyonacea	8
		Zoanthidea	15
		Scleractinia	133
		Corallimorpharia	10
	Scyphozoa		4
		Coronatae	1
	Cubozoa		8
		Carybdeida	6
	Staurozoa		1
		Stauromedusae	1
	Hydrozoa		247
		Narcomedusae	2
		Hydroida	2
		Anthoathecata	90
		Leptothecata	136
		Trachymedusae	1
		Siphonophorae	3
Platyhelminthes			105
	Turbellaria		12
		Tricladida	2
		Polycladida	49
		Seriata	4
	Rhabditophora		53
	Acoela		2
Annelida			1129
	Polychaeta		744
		Amphinomida	64
		Phyllodocida	356
		Spionida	64
		Terebellida	53
		Eunicida	59
		Sabellida	16
	Clitellata		3
Nemertea			69
	Anopla		18
		Palaeonemertea	2
		Enopla	2

Table 2 (Continued)

Phylum	Class	Order	Number of sequences
Totals			
23	42	102	7686
Sipuncula			7
	Phascolosomatidea		4
		Phascolosomatida	2
		Aspidosiphonida	2
	Sipunculidea		3
		Golfingiida	3
Echiura			3
	Echiuroidea		2
		Bonellida	2
Arthropoda			2805
	Diplopoda		4
		Polydesmida	4
		Spirostreptida	6
	Ostracoda		29
		Podocopida	14
		Myodocopida	2
	Arachnida		16
		Aranaea	8
		Oribatida	2
		Pseudoscorpiones	4
		Halcarida	2
	Malacostraca		2346
		Stomatopoda	45
		Mysidacea	8
		Tanaidacea	33
		Decapoda	2119
		Isopoda	46
		Amphipoda	83
		Cumacea	2
	Maxillopoda		30
		Pedunculata	8
		Laurida	4
	Insecta		131
		Hemiptera	2
		Blattaria	8
		Hymenoptera	20
		Diptera	16
		Psocoptera	6
		Coleoptera	8
		Lepidoptera	71
	Pycnogonida		12
	Copepoda		16
		Cyclopoida	4
		Harpacticoida	8
		Poecilostomatidoida	4
Mollusca			1776
	Cephalopoda		12
		Octopoda	12
	Bivalvia		212
		Lucinoida	8
		Veneroida	124
		Mytiloida	20
		Pteroida	9

Table 2 (Continued)

Phylum	Class	Order	Number of sequences
Totals			
23	42	102	7686
		Arcoidea	5
		Pectinoidea	22
		(Euheterodonta)	1
		Limoida	2
		Carditoidea	1
	Gastropoda		1318
		Notaspidea	2
		Neogastropoda	264
		Littorinomorpha	157
		Caenogastropoda	196
		Stylommatophora	151
		Nudibranchia	159
		Anaspidea	40
		Cephalaspidea	89
		Archaeopulmonata	10
		Systellommatophora	6
		Saccoglossa	44
		Cycloneritimorpha	8
		Neritoida	24
	Polyplacophora		2
Brachiopoda			5
	Rhynchonellata		1
		Terebratulida	1
Bryozoa			246
	Gymnolaemata		206
		Cheilostomatida	206
	Stenolaemata		18
		Cyclostomatida	18
Phoronida			6
Entoprocta			1
Nematoda			4
Chaetognatha			1
	Sagittoidea		1
		Aphragmophora	1
Echinodermata			209
	Echinoidea		32
		Camarodonta	4
		Cidaroida	2
		Spatangoida	14
	Holothuroidea		58
		Aspidochirotida	42
		Apodida	7
	Ophiuroidea		103
		Ophiurida	52
	Asteroidea		10
		Valvatida	10
Hemichordata			17
	Enteropneusta		17
Chordata			443
	Leptocardii		9
	Thaliacea		4
		Salpida	4
	Asciacea		419
		Phlebobranchia	62

Table 2 (Continued)

Phylum	Class	Order	Number of sequences
Totals			
23	42	102	7686
		Stolidobranchia	101
		Aplousobranchia	238
Other Metazoa			16
Chlorophyta			5
Rhodophyta			6
Ascomycota			2
	Lecanoromycetes		2
		Lecanorales	2
Phaeophyta			2
Cyanobacteria			4

Discussion

The novel primers jgLCO1490 and jgHCO2198, revisions of the Folmer primers, were used to amplify a fragment of the COI gene from a phylogenetically broad sample of marine invertebrates. As expected, the revised primers amplified templates where Folmer primers had failed, and overall PCR success rate, defined as producing a discrete band of the expected size, was about 70% for two large-scale biodiversity survey projects.

The primer performance reported here should be interpreted in the context of the conditions under which PCR was conducted. PCR success depends on many factors that will vary across projects. Among these factors are template quality, taxon-specific effects and opportunity for PCR optimization. Our projects were unselective towards template quality, as our objective was to obtain sequences from all specimens collected and identified to at least the phylum level (although generally much better). Consequently, attempts at extraction and PCR were made without respect to apparent tissue and DNA quality. For example, Moorean tissues were extracted soon after collecting without prior preservation, but some were visibly moribund prior to extraction. All San Francisco Bay samples were preserved in 90% ethanol, and condition of specimens varied widely: some specimens were minuscule, insufficiently covered with ethanol, or had all ethanol evaporated. Our reported PCR success rate clearly would have been higher with more stringent template selection. The important results, therefore, are the breadth of taxa that were successfully amplified and the generally high rate of success that was achieved without template selectivity or PCR optimization.

Amplification and sequencing of nontarget templates was rare. Bacterial sequences were found only from bryozoans and tunicates. It is intriguing that host-specific bacterial symbionts are known from bryozoans in

fouling communities (Lim-Fong *et al.* 2008), but we can make no further comment about the source of bacterial DNA in our samples. Tunicates feed on bacteria, and the sequences we obtained could come from gut contents, but this remains speculation. About 1% of sequences from San Francisco Bay were obviously not from the identified specimen. These results could be due to mislabelling, mixed tissues (from commensal relationships, overgrowth or in guts) or mixed templates (laboratory error). In any of these cases, careful harvesting of tissues and laboratory procedures can further minimize anomalous results.

Amplification success was seen with most phyla, and we are reluctant to discourage further experiments with taxa that failed. For example, we had no success with Nematomorpha, but Looney *et al.* (2012) report sequences from PCR that had used the standard Folmer primers; we expect that our primers should have worked. Therefore, when amplification failed in the present study, we cannot exclude template quality as a factor, or that PCR optimization might have yielded positive results. The few contrasting results from Moorea and San Francisco Bay (e.g. entoprocts discussed above) suggest caution in generalizing that particular taxa amplify poorly with these primers. Species level effects may bias these results within higher taxonomic groups, with some species being especially easy or difficult to amplify. The difficulty of species level identification in Moorea, the source of our larger data set, makes this bias hard to quantify.

It is possible that these primers might yet be further improved. Aside from inosine containing nucleotides, other nonselective bases are available, such as *N*-nitroindole, which might have different performance although at a higher cost of synthesis (\$150/internal base at 100 nM synthesis scale vs. \$10 per internal base for dITP at Integrated DNA Technologies (IDT), Coralville, IA, USA). The disadvantage of nonselective bases is that

Table 3 Phylum, Class, Order and Families of specimens from settling plates in San Francisco Bay successfully amplified and sequenced with the primer combination of jgLCO1490 and jgHCO2198

Phylum	Class	Order	Family	N
Totals				
10	13	31	56	
Porifera	Demospongiae	Halichondrida	Halichondriidae	1
Cnidaria	Anthozoa	Actiniaria	Diadumenidae	1
			Metridiidae	1
	Scyphozoa	Semaeostomeae	Ulmaridae	2
	Hydrozoa	Leptothecata	Campanulariidae	3
		Anthoathecata	Pandeidae	1
			Tubulariidae	1
		Unknown hydrozoan		1
Platyhelminthes	Turbellaria		Unknown turbellarian	1
Annelida	Polychaeta	Phyllodocida	Nereididae	1
			Polynoidae	1
			Syllidae	3
		Terebellida	Cirratulidae	2
			Terrellidae	1
		Eunicida	Dorvilleidae	1
		Sabellida	Sabellidae	2
			Serpulidae	1
		(Scolecida)	Orbinidae	1
		Unknown polychaetes		2
Nemertea	Anopla	Palaeonemertea	Cephalothricidae	1
Arthropoda	Malacostraca	Amphipoda	Amphithoidae	1
			Caprellidae	3
			Corophiidae	1
			Gammaridae	1
			Ischyroceridae	1
			Leucothoidae	1
		Tanaidacea	Unknown tanaid	1
	Maxillopoda	Sessilia	Balanidae	3
			Unknown barnacle	1
Mollusca	Bivalvia	(Euheterodonta)	Hiatellidae	1
		Myoida	Myidae	1
		Ostreoida	Ostreidae	1
		Pectinoida	Pectinidae	1
		Mytiloida	Mytilidae	4
		Unknown Bivalvia		1
	Gastropoda	Neogastropoda	Columbellidae	1
			Muricidae	1
			Nassariidae	1
		Littorinimorpha	Calyptraedae	1
			Littorinidae	1
		Nudibranchia	Unknown nudibranchs	2
		Cephalaspidea	Haminoeidae	1
			Unknown gastropod	1
	Gymnolaemata	Cheilostomatida	Bugulidae	6
			Candidae	2
			Cryptosulidae	2
			Electidae	4
			Hippothoidae	1
			Lepraliellidae	1
			Microporellidae	1
			Schizoporellidae	2
			Smittinidae	1
			Watersiporidae	3

Table 3 (Continued)

Phylum	Class	Order	Family	<i>N</i>
Totals				
10	13	31	56	
		Ctenostomatida	Nolellidae	1
			Vesiculariidae	1
	Unknown bryozoans			3
Entoprocta		Coloniales	Barentsiidae	1
Chordata	Ascideacea	Phlebobranchia	Asciidiidae	3
			Cionidae	2
			Corellidae	1
		Stolidobranchia	Molgulidae	1
			Styelidae	5
		Aplousobranchia	Clavelinidae	1
			Didemnidae	2
Rhodophyta	Florideophyceae	Ceramiales	Rhodomelaceae	1

N, number of species in each family.

they contribute no specificity and add cost (compared with \$0.35 per conventional base at IDT) to primer synthesis. The primers we designed did not accommodate every polymorphism detected (Table 1). Any such primer will contain a great many degenerate or nonselective bases and inadvertently match many PTS that are not COI, including numts (nuclear copies of mitochondrial genes) and pseudogenes. However, the primers we designed produced very few double-banded PCR products in which the product of incorrect size might represent such artefacts. We do not have a method to determine the optimal trade-off between taxonomic breadth and loss of specificity for COI, but this could be further explored. Improvement of these primers may be possible by, for example, selection of different nucleotides at positions where a choice was made (rather than opting for degeneracy), using nonequimolar mixes at degenerate positions, or by experimenting with the number and position of inosine bases.

Other research groups may find these primers a useful complement to existing primers. For example, the Cnidarian and Poriferan Tree of Life projects (A. G. Collins, personal communication), and the Smithsonian's Laboratory of Analytical Biology DNA barcoding program (A. Driskell, personal communication) have adopted these primers for routine COI barcoding. Projects with fewer species might initially explore the Folmer primers or the dgLCO1490/HCO2198 primers, with their lower costs, and adopt the jgLCO1490/jgHCO2198 primers if these fail. Biodiversity survey projects, however, handling diverse samples in a high-throughput workflow, require a high success rate without repeating experiments. It is our hope that these primers will be useful for such studies, for general population genetic and phylogeographic studies, and as a tool to obtain

preliminary sequences when custom primers are an option. Because of their success across a broad spectrum of metazoan phyla, we also envision their use in creating amplicon libraries from environmental samples such as planktonic communities or from gut contents for food web analyses, and such studies are in progress in our laboratories.

Acknowledgements

This work was supported by the California Department of Fish and Wildlife (San Francisco) and the Gordon and Betty Moore Foundation (Moorea). The Smithsonian Environmental Research Center Marine Invasions Laboratory made specimen collections in San Francisco Bay, and we particularly thank Gail Ashton, Christopher Brown, Tracy Campbell, Linda McCann and Greg Ruiz. Michelle Marraffini, Kristin Meagher and Gillian Rhett provided additional laboratory and computer assistance at MLML. We also acknowledge a multitude of Moorea Biocode Project participants for their field and collections efforts. Comments by two reviewers significantly improved our manuscript.

References

- Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA barcoding of marine metazoa. *Annual Review of Marine Science*, **3**, 471–508.
- Folmer O, Black M, Hoeh W, Lutz R, 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.
- 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**, 313–321.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, USA*, **101**, 14812–14817.

- Jarman SN (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. *Bioinformatics*, **20**, 1644–1645.
- Kress WJ, Erickson DL (eds) (2012) DNA barcodes: methods and protocols. *Methods in Molecular Biology*, **858**, Springer, Berlin, Germany.
- Lim-Fong GE, Regali LA, Haygood MG (2008) Evolutionary relationships of “*Candidatus endobugula*” bacterial symbionts and their *Bugula* bryozoan hosts. *Applied and Environmental Microbiology*, **74**, 3605–3609.
- Looney C, Hanelt B, Zack RS (2012) New records of nematode parasites (Nematomorpha: Gordiida) of ground beetles (Coleoptera: Carabidae) and camel crickets (Orthoptera: Rhaphidophoridae) in Washington state. *Journal of Parasitology*, **98**, 554–559.
- Lupi R, de Meo PD, Picardi E *et al.* (2010) MitoZoa: a curated mitochondrial genome database of metazoans for comparative genomics studies. *Mitochondrion*, **10**, 192–199.
- Mackie JA, Keough MJ, Christidis L (2006) Invasion patterns inferred from cytochrome oxidase I sequences in three bryozoans, *Bugula neritina*, *Watersipora subtorquata*, and *Watersipora arcuata*. *Marine Biology*, **149**, 285–295.
- Meusnier I, Singer GA, Landry J-F *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics*, **9**, 214.
- Meyer CP (2003) Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. *Biological Journal of the Linnean Society*, **79**, 401–459.
- Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS biology*, **3**, e422.
- Ratnasingham S, Hebert PD (2007) BOLD: the Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, **7**, 355–364.
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (eds Krawetz S, Misener S), pp. 365–386. Humana Press, Totowa, New Jersey.
- Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R (2005) Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **360**, 1805–1811.
- Stoeckle MY, Hebert PD (2008) Barcode of life. *Scientific American*, **299**, 82–86, 88.
- Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1847–1857.
- Ward RD, Hanner R, Hebert PD (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.

J.G. designed primers jgLCO1490 and jgHCO2198, developed PCR conditions, supervised the San Francisco Bay project, performed data analysis, and wrote the manuscript. C.M. organized sampling in Moorea, supervised labwork at the Smithsonian Institution, performed data analysis, and edited the manuscript. M.P. and H.H. performed most labwork at the Smithsonian Institution and Moss Landing Marine Laboratories, respectively. H.H. also contributed to data analysis.

Data Accessibility

The data contained in this study are the taxonomic names of specimens from Moorea, French Polynesia and San Francisco Bay, California that were successfully amplified by the primer pair jgHCO2198 and jgLCO1490. These data are contained in Tables 2 and 3. Accession numbers for sequences used to generate these novel primers are contained in the caption to Table 1.