Syst. Biol. 58(4):445–451, 2009

DOI:10.1093/sysbio/syp033
Advance Access publication on July 6, 2009

Barcoding Bamboozled by Bacteria: Convergence to Metazoan Mitochondrial Primer Targets by Marine Microbes

MARK E. SIDDALL[1*], FRANK M. FONTANELLA[2], SARA C. WATSON[1], SEBASTIAN KVIST[3], AND CHRISTER ERSÉUS[4]

[1] *Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA;*
[2] *The Graduate Center, City University of New York, New York, New York, USA;*
[3] *Richard Gilder Graduate School, American Museum of Natural History, New York, New York, USA;*
[4] *Department of Zoology, University of Gothenburg, Göteborg, Sweden;*
**Correspondence to be sent to: Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York, USA; E-mail: siddall@amnh.org.*

DNA barcoding has become a legitimate tool for assessment of global biodiversity patterns in a manner intended to allow diagnoses of known species to nontaxonomists. Relying on a standardized region (~650 nucleotides) of the cytochrome *c* oxidase subunit I mitochondrial gene (COI), a variety of taxon-based initiatives (www.barcoding.si.edu/major_projects.html), as well as large-scale locality-based barcoding projects (e.g., bscit.berkeley.edu/biocode/), are well underway. If DNA barcoding is to achieve its promise, these projects are essential as they will provide much-needed comparative baseline data for taxonomic groups, or for regional assemblages, upon which future identification efforts will rely. Thorough coverage of clades and guilds is essential to the success of DNA barcoding, but so too is highly accurate taxonomic identification of the source organisms. Although surreptitious amplification of nuclear mitochondrial pseudogenes (numts) may lead the unwary to overestimation of species diversity (Song et al. 2008), it may not necessarily lead to incorrect taxonomic determinations so long as the source organism was accurately identified. Erroneous taxonomic identification of a reference organism is a more serious problem (Vilgalys 2003; Traub et al. 2007), yet one that can be mostly mitigated by expert identification and appropriate deposition of reference specimens in natural history collections (Smith et al. 2008). Even when great care is taken to avoid contamination, inadvertent amplification of surface-adhering or tissue-infecting organisms may be impossible to avoid (e.g., Zhang et al. 1997; Nilsson et al. 2006). As such, underlying broad barcoding efforts is a presumption that the primers used are more likely to amplify the animal of interest than they are likely to amplify (e.g.) an unobserved bacterial or fungal contaminant. The LCO1490 and HCO2198 primers that provided the basis for the DNA barcoding standard were described by Folmer et al. (1994) as being "universal" for 11 invertebrate phyla. Their universality, however, is belied by several groups of animals not amplifying at the predicted optimal annealing temperature for these primers (e.g., Halanych and Janosik 2006). This difficulty has led to the use of modified or degenerate primers (Bely and Wray 2004; Hebert et al. 2004; Ivanova et al. 2007) and to greatly reduced primer annealing temperatures (de Waard et al. 2008). In contrast to universality, little attention has been given to the exclusivity with which barcoding primers will amplify target versus nontarget taxa. Here we demonstrate that primers widely used for DNA barcoding may prove problematic in light of their predicted, and repeatedly demonstrated, ability to amplify nontarget cold-adapted marine gammaproteobacteria.

DNA was isolated, using the E.Z.N.A. Mollusc DNA kit (Omega Bio-Tek, Norcross, GA), from specimens of two marine mollusc species, the shipworm *Xylotrya setacea* and the piddock clam *Penitella penita* collected near Moclips, Washington, DC. DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), from specimens of a marine oligochaete, *Paranais botniensis*, collected on the Island of Öland and near Tjärnö Marine Biological Station in Sweden. Amplification of the barcoding region of eukaryotic mitochondrial COI was attempted first with the published (Folmer et al. 1994) primers LCO1490: 5′-GGTCAACAAATCATAAAGATATTGG-3′ and HCO-2198: 5′-TAAACTTCAGGGTGACCAAAAAATCA-3′ with an annealing temperature of 48 or 45 °C. For the molluscs, and in light of failed reactions in the foregoing, additional amplifications were accomplished substituting the COI-E: 5′-TATACTTCTGGGTGTCCGAAGAA-TCA-3′ primer (Bely and Wray 2004) for HCO2198 and a 52 °C annealing temperature. Amplification reaction mixtures for gene fragments used Ready-To-Go™ PCR Beads (Amersham Pharmacia Biotech, Piscataway, NJ), 0.5–1 μL of each primer, 1–2 μL DNA template, and 21–23 μL RNase-free H_2O (total volume, 25 μL). Amplification reactions were performed either in an Eppendorf Mastercycler or in a PTC-100 (MJ Research, Waltham, MA). Amplification products were purified with the AMPure Purification system PCR Cleaning protocol (Agencourt, Beverly, MA) or the E.Z.N.A. Cycle-Pure kit (Omega Bio-Tek) and sequenced in both directions. Each sequencing reaction mixture included 0.5 μL BigDye (Applied Biosystems, Perkin-Elmer Corporation, CA), 0.5 μL extender buffer, 1 μL of each (1 mM) primer (single primer for each direction), and 3 μL of DNA template. Sequences were purified by ethanol precipitation to remove primers and unincorpo-

rated dyes or by following the CleanSEQ protocol (Agencourt). Products were electrophoresed in an ABI Prism 3730 sequencer (Applied Biosystems). Specimens of the oligochaete *P. botniensis* were sent to Macrogen Inc., Seoul, South Korea, for sequencing. Sequences of complimentary strands were edited and reconciled using EditSeq (DNAStar Inc., Madison, WI) and CodonCode Aligner (CodonCode, Dedham, MA).

Preliminary determinations of taxonomic identity for amplified sequences were accomplished using BLAST against NCBI-nr using the blastn protocol. Comparative COI sequences were then obtained from each of putatively metazoan taxa returned from NCBI-nr that closely matched query sequences submitted; whole-genome–derived COI sequences of gammaproteobacterial species that closely matched query sequences submitted; whole-genome–derived COI sequences of some additional related gammaproteobacterial species; and whole-mitochondrial genome–derived COI sequences for a range of eukaryotic taxa (see Table 1). That is, with the exception of the molluscan query sequences, the oligochaete query sequences, and the putatively metazoan COI sequences returned from NCBI-nr, the remaining comparative sequences in Appendix A were obtained from whole-genome (bacterial and eukaryotic mitochondrial) sequencing projects so as to avoid interpretation difficulties that could arise from actual DNA barcoding amplicons. Moreover, for those taxa with COI sequences directly available from whole bacterial or mitochondrial genomes, the nucleotide sequence of the priming region is known directly.

Forty-nine sequences for COI were aligned with MUSCLE (Edgar 2004) on the European Bioinformatics Institute server. This included whole-mitochondrial genome-acquired COI sequences for 11 eukaryotic species, whole-genome–acquired COI sequences for 23 gammaproteobacterial species, 11 amplification–acquired putatively metazoan COI sequences from NCBI-nr, and the four newly generated query sequences from molluscs and oligochaetes. In addition, a whole-genome-acquired COI sequence for *Bacillus anthracis* was included to better test for an expected Eukaryota/Prokaryota bipartition, and a single metazoan amplification–acquired COI sequence for an amphipod was included for comparison. Aligned sequences (TreeBASE study S2307, matrix M4381) were subject to phylogenetic analysis using maximum likelihood with RaxML (Stamatakis et al. 2005) on the CIPRES server (www.phylo.org) in which stability of groups was determined with 1000 bootstrap replicates using the general time reversible (GTR) model with a gamma-distributed rate parameter—the model class selected by FindModel (Tao et al. 2008) under the Akaike information criterion (Akaike 1974). Because the DNA barcoding method uses a Neighbor-Joining approach (Saitou and Nei 1987), a Neighbor-Joining tree was obtained using PAUP* (Swofford 2002) employing the GTR model with a shape parameter of 0.691245 as estimated by RaxML. In light of nonstationary base compositions (chi-square = 1356, $P < 0.000$), the data also were subject to Neighbor-Joining with a LogDet model in PAUP* (Swofford 2002) but with invariable sites removed.

Highly significant (i.e., *e* value $< e^{-100}$) scores from BLAST queries using sequences derived from the two molluscan isolates (*X. setacea* and *P. penita*), or the oligochaete isolates (*P. botniensis*), overlapped on several putatively eukaryotic taxa as well as on several definitively prokaryotic taxa. Putatively eukaryotic sequences that were found to significantly match all three query species were the ascidian chordate *Cnemidocarpa verrucosa* AJ830012, the amphipod crustacean *Scopelocheirus schellenbergi* AY830439, and the clupeiform fish *Engraulis japonicus* EU266375; none of which had been generated by whole mitochondrial genome sequencing. Prokaryotic COI sequences that were found to significantly match all three query species were the gammaproteobacteria *Shewanella amazonensis* CP000507, *Shewanella baltica* CP000753, *Pseudomonas putida* CP000926, and *Photobacterium profundum* CR378663, each of which had been generated through whole-genome sequencing. Eukaryotic amplification-derived sequences that were found to significantly match two of the query sequences included the limpet mollusc *Patelloida striata* AB161589, four separate isolates from the amphipod crustacean *Eurythenes gryllus* (AY830426, AY830434, AY830436, and AY830437), and 44 distinct isolates from the bivalve mollusc *Scapharca broughtonii* (see Cho et al. 2007, for accession numbers), none of which had been generated by whole mitochondrial genome sequencing. Prokaryotic COI sequences that were found to significantly match two query species were the gammaproteobacteria *Marinobacter aquaeolei* CP000514, *Psychromonas ingrahamii* CP000510, *Vesicomyosocius okutanii* AP009247, *Vibrio vulnificus* BA000038, *Colwellia psychrerythraea* CP000083, *Shewanella denitrificans* CP000302, *Shewanella frigidimarina* CP000447, and *Ruthia magnifica* CP000488, each of which had been generated through whole-genome sequencing.

Neighbor-Joining analyses (Fig. 1) employing the GTR model placed amplification-based sequences of the two mollusc isolates (*X. setacea* and *P. penita*) and the oligochaete COI isolate (*P. botniensis*) among a clade of gammaproteobacterial COI genomic sequences that was dominated by marine Alteromonadales and Vibrionales and which was distinct from the clade of whole–mitochondrial genome eukaryotic COI sequences (the analysis employing LogDet distances differed from the GTR distance tree only in placing the two *P. botniensis* sequences with *Legionella pneumophila* closer to the base of, yet well within, the gammaproteobacterial clade). Maximum likelihood bootstrap resampling yielded 100% support for a Eukaryota/Prokaryota bipartition, 70% support for Metazoa genomic COI, and 100% support for gammaproteobacteria. Of the 12 eukaryotic COI isolates derived from amplification, as opposed to whole mitochondrial genome sequencing, each of eight grouped sister to a clade of gammaproteobacteria, half of which had 100% bootstrap support: the bivalve *S. broughtonii* as sister to *S. frigidimarina*; the brachiopod *Lingula anatina* with *Vibrio parahaemolyticus* (notably,

TABLE 1. Comparison of DNA barcoding target priming sites in eukaryotic and microbial genomes

Phylum/Higher Taxon	Species	LCO1490* GGTCAACAAATCATAAAGATATTGG		HCO2198* TGATTTTTTGGTCACCCTGAAGTTTA .....C..C..A.....A.....A.. COIE**
EUKARYOTA				
NEMATODA	*Ascaris suum*	AAAGTT.T.....G.....	†	..G.....T.....G.....
CHORDATA	*Danio rerio*	TC.....T.....C.....		.....C.....C.....A.....C..
PLATYHELMINTHES	*Echinococcus canadensis*	TAA.TTTGG.....GCG...A..		..G.....T..A..G.....
ARTHROPODA	*Ligia oceanica*	AT..T..C..C.....G.....		..G.....T..A..G..G..
BRACHIOPODA	*Lingula anatina*	AA..GGT.....G.....		.....G...A.....G..
MOLLUSCA	*Lottia digitalis*	C...C....GA..C.....		.....C.....C.....G.....
ANNELIDA	*Lumbricus terrestris*	AC.....T.....C.....		.....C.....A.....A..
MOLLUSCA	*Mytilus edulis*	Ligia oceanica		.....G.....G..G..
ARTHROPODA	*Parhyale hawaiiensis*	TT..T..T.....C.....		.....A.....
PORIFERA	*Xestospongia muta*	TT..T.....		..G.....G..T..G.....
VIRIDPLANTAE	*Oryza sativa Japonica*	TC..C..T..C..C..G.....C..		..G..C..C.....T..A..G..G..
APICOMPLEXA	*Plasmodium falciparum*	CAAATTGT..C.....AC.T.A..		..G.....A..T.....A..
JAKOBIDA	*Reclinomonas americana*	TT.....C.....		..G..C.....T.....A..
PROKARYOTES				
ARCHAEA	*Halorubrum lacusprofundi*	TCA.G..G..C..C..G.....C..		..G..C..C..C..T..A..G..G..
CYANOBACTERIA	*Prochlorococcus marinus*	TTAGCCTTG.....TA.....		..G....A.TC...T..A.CT..G..
FIRMICUTES	*Bacillus anthracis str Ames*	C.A..GT.G.C.....A.G....C		..GA.....T..A.....A..
ALPHA PROTEOBACTERIA	*Rickettsia prowazecki* Madrid E	TT..T..T.....C.....C..		..G.....T.....A..
BETA PROTEOBACTERIA	*Burkholderia pseudomallei*	TCG.C..C..C..C..G.....C..		..G..C..C..C.....C..G..G..
GAMMA PROTEOBACTERIA	*Alteromonas macleodii* Deep	TTA.C..T..C..C..G..C..A..		..G..C..C.....G..
GAMMA PROTEOBACTERIA	*Colwellia psychrerythraea* 34H	ATA.C..T..C.....G..C..C..		..G..C.....G.....
GAMMA PROTEOBACTERIA	*Hahella chejuensis* KCTC 2396	TAA.C..C..C.....C.....		..G..C..C..T....G..A..
GAMMA PROTEOBACTERIA	*Legionella pneumophila* Paris	TTA....C..C..C.....C..A..		..G..C.....T.....G..
GAMMA PROTEOBACTERIA	*Marinobacter aquaeolei* VT8	T.A.C.....C.....C..C..		..G..C..C..G..T..G..G.....
GAMMA PROTEOBACTERIA	*Oceanobacter sp* RED65	ATA....C..C.....C.....		..G..C.....A.....A..
GAMMA PROTEOBACTERIA	*Photobacterium profundum* SS9	TTA.G..T.....		..G..C.....C..
GAMMA PROTEOBACTERIA	*Pseudoalteromonas atlantica* T6	ACA.G.....C..C.....C..A..		..G..C.....G.....G.....
GAMMA PROTEOBACTERIA	*Pseudoalteromonas haloplanktis*	ATA.C..T.....A..		..G.....C.....C.....A..
GAMMA PROTEOBACTERIA	*Pseudomonas putida* W6	T.A.C..C..C..C..G..C..C..		..G..C..C..C.....A..G..G..
GAMMA PROTEOBACTERIA	*Psychromonas ingrahamii* 37	TTAGT..C..C.....		..G..C.....
GAMMA PROTEOBACTERIA	*Ruthia magnifica* str Cm	AAA.G..C.....C..G.....		..G.....C..T.....G.....
GAMMA PROTEOBACTERIA	*Saccharophagus degradans*	TTA....C..C.....C..		..G..C..C..G..T..A..G.....
GAMMA PROTEOBACTERIA	*Shewanella amazonensis* SB2B	TAA....C.....C..G.....A..		..G..C..C.....G..G..
GAMMA PROTEOBACTERIA	*Shewanella baltica* OS185	TAA.C..C..C.....C.....		..G..C..C..C.....C..
GAMMA PROTEOBACTERIA	*Shewanella denitrificans* OS217	TAA.C..C..C.....C..A..		..G..C..C.....C.....C..
GAMMA PROTEOBACTERIA	*Shewanella frigidimarina* NCIMB	TCA.C..T..C.....		..G..C.....T..C.....
GAMMA PROTEOBACTERIA	*Vesicomyosocius okutanii* HA	AAA....C.....		..G.....T.....G.....
GAMMA PROTEOBACTERIA	*Vibrio parahaemolyticus* RIMD	AT..C..C..C.....C..C..		..G..C.....A.....G..
GAMMA PROTEOBACTERIA	*Vibrio shilonii* AK1	AC..C..T..C.....C..		..G..C.....C..T..A.....G..
GAMMA PROTEOBACTERIA	*Vibrio vulnificus* YJ016	AT..C..C.....C.....C.....		..G..C..C..C.....

* Folmer et al., (1994).
** Bely and Wray (2004).
† nucleotides matching the Folmer et al. (1994) primers denoted ".".

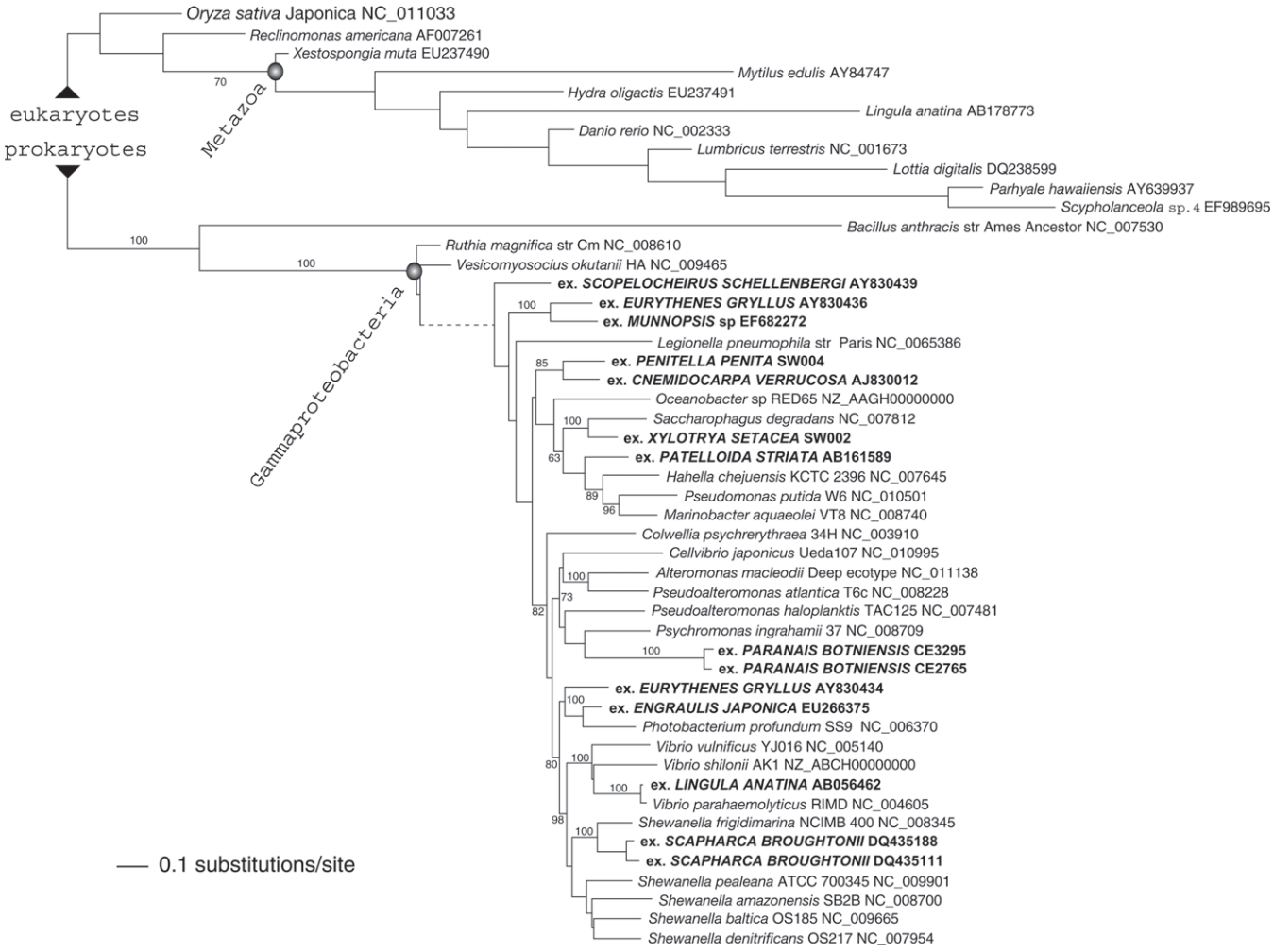

FIGURE 1. Neighbor-Joining tree, using a GTR + Γ model, of eukaryotic and prokaryotic COI DNA sequences. A variety of metazoan COI sequences from public databases (in capitals and denoted "ex.") and newly generated sequences from a shipworm, *Xylotrya setacea*, a piddock clam, *Penitella penita*, and a marine oligochaete, *Paranais botniensis*, group among whole-genome–acquired COI sequences for gammaproteobacterial species instead of grouping among eukaryotic whole-mitochondrial genome–acquired COI sequences. The interrupted branch indicates where sequences from *Paranais* and genomic *Legionella* isolates placed under the LogDet model. Numbers at nodes are bootstrap proportions from maximum likelihood (GTR + Γ) analyses. Gene or genome accession identifiers are appended to each taxon name.

not with the mitochondrial genomic COI sequence for *L. anatina*); the fish *E. japonica* as sister to *P. profundum*, and the shipworm *X. setacea* with *Saccharophagus degradans*. There were only two clades, of two taxa each, that did not include a gammaproteobacterium (though each was well nested among marine gammaproteobacteria): the COI sequence of the piddock *P. penita* was supported (85%) as sister to an isolate from the ascidian chordate *C. verrucosa* and an amphipod crustacean (*E. gryllus*) sequence was supported (100%) as sister to an isopod crustacean, *Munnopsis* sp. (but not to the other sequence of *E. gryllus).*

The average genetic (*p*) distance between the amplification-derived eukaryotic COI isolates and the gammaproteobacterial COI sequences derived from whole-genome sequencing was 28%. The average genetic (*p*) distance between the amplification-derived eukaryotic COI isolates and the eukaryotic COI sequences derived from whole mitochondrial genomes was 47%, which was identical to the average genetic (*p*) distance between the gammaproteobacterial COI sequences derived from whole-genome sequencing and eukaryotic COI sequences derived from whole mitochondrial genomes. The smallest pairwise genetic (*p*) distance was 1.2% between a brachiopod *L. anatina* isolate and *V. parahaemalyticus* genomic COI. This contrasts with the 49% difference between that *L. anatina* isolate and the *L. anatina* mitochondrial genome COI sequence.

Table 1 illustrates homologous priming site sequences obtained from whole prokaryotic genome data and from whole mitochondrial eukaryotic genome data that are targeted by LCO1490 and HCO2198 (and COI-E) primers, respectively. The first 10 nucleotide positions of the LCO1490 priming site showed marked variability across a wide range of eukaryotes and prokaryotes.

Among the 15 positions closest to the 3′ end of that priming target, most of the variability was found in third codon positions. Each of *P. profundum, V. okutanii, Pseudoalteromonas haloplanktis, S. frigidimarina*, and *P. ingrahamii* had longer stretches of contiguous nucleotides matching the LCO1490 primer than most of the metazoan taxa. With the exception of *B. anthracis* and *Prochlorococcus marinus*, the HCO2198 priming site showed variation only in third codon positions. All eukaryotic nucleotide variants at these HCO2198 positions also were found among prokaryotes. The gammaproteobacterium *P. ingrahmii* exhibited a closer match to the HCO2198 primer than all but three of the metazoan taxa. Several gammaproteobacterial species also exhibited more contiguously matched nucleotides for HCO2198 than the majority of animals.

Unequivocally, the primers and annealing temperatures typically employed for DNA barcoding of invertebrate metazoan taxa efficiently amplify the COI locus of certain marine gammaproteobacteria. Consequently, many COI sequences in public databases like GenBank are erroneously described as being from various marine metazoan taxa (Fig. 1). Already this has the potential to confound DNA barcoding initiatives because most of the marine bacterial contaminants found here are also included in the full Barcode of Life Database (BOLD Systems, though clearly noted to be "unvalidated") under an erroneous metazoan name. Indeed, identification attempts through BOLD reveal additional contaminants in the database that appear to have resulted from marine locality-based barcoding surveys not yet in public databases. These include the shipworm *Bankia carinata* GMBL0400-06, the shrimp *Hymenodora* EKDF0020-07 and EKDF0021-07, as well as bivalve molluscs CCMLS010-07 and CCMLS011-07. In most cases, one would expect that accidental amplification of a bacterial contaminant would be obvious from an examination of results obtained from a BLAST search (either through NCBI or through BOLD). Alas, because these marine gammaproteobacteria appear so amenable to accidental amplification, already there is a sufficient number of contaminant sequences with metazoan names to mislead one to believe that an animal had been properly amplified and sequenced. Particularly vexing is a published attempt to understand the population genetics of the commercially important marine bivalve mollusc *S. broughtonii* (see Cho et al. 2007) which generated 44 bacterial COI contaminants, such that our querying putative molluscan isolates from *X. setacea* and *P. penita* returned more apparently molluscan matches than bacterial matches. Ultimately, the population genetic analysis by Cho et al. (2007), although analytically sound, was clearly an analysis of a *S. frigidimarina*–like bacterium associated with *S. broughtonii* as opposed to being one of the clam itself. Obviously, the utility of that study in guiding the management, conservation, and assessment of a commercially important shellfish resource will have to be reevaluated.

The marine bacteria that have so clearly been accidentally amplified and sequenced in this and preceding work share several common features. They are marine gammaproteobacterial species allied with the Alteromonadales and Vibrionales that are cold adapted and have a propensity for associations with surfaces of marine animals. The Alteromonadales, which has received attention for marine bioremediation of crude oil spills (Deppe et al. 2005; Gerdes et al. 2005), was established for a group of gram-negative marine heterotrophic microbes that are phenotypically similar to pseudomonads but which have a consistently lower G+C content (Baumann et al. 1972). Gammaproteobacteria that were found to significantly match our BLAST queries, and among which the various marine animal-isolated bacterial contaminants in Figure 1 are nested, include representatives of each alteromonad subclade (I, II, III, IV, and V, plus Shewanellaceae) detailed by Ivanova et al. (2007). In addition to being well known for psychrophilic habitat profiles (growing optimally at less than 20 °C), many alteromonads are known to form surface associations with marine animals (Mikhailov et al. 2006) ranging from molluscs (Ivanova et al. 1996) to sponges (Ivanova et al. 2000, 2002), ascidians (Holmström et al. 1998), and fish (Simidu et al. 1990) wherein some cause dermal lesions (McGarey et al. 1991).

Like metazoan mitochondrial DNA sequences (Martin 1995; Perna and Kocher 1995), a variety of organisms show a tendency toward reduced G + C richness. Microbial genomic nucleotide compositions are under strong thermoadaptive pressure (Khachane et al. 2005), and the thermodynamic stability of guanine–cytosine bonds versus adenosine–thymidine bonds predicts that psychrophiles and thermophiles would have markedly different genomic base compositions. This prediction is borne out in broad surveys of microbes across the range of optimal growth temperatures (Khachane et al. 2005) in which psychrophilic microbes are expected to favor A + T richness over mesophilic G + *C* compositions. In coding regions, this drift in base composition is least constrained in third codon positions to the extent that most such changes have no effect on translated amino acids (Crick 1966), and third position differences constitute most of the variation seen at the COI priming sites (Table 1). Several marine species that group closest to the metazoan-derived contaminants exhibit among the lowest known genomic G + C base compositions (Mikhailov et al. 2006), such as *Pseudoalteromonas denitrificans* (37%), *Pseudoalteromonas atlanticus* (41%), *Pseudoaltermonas haloplanktis* (42%), and *Colwellia* species (40%). Given a melting temperature (*T*m) of 81.5 + 0.41(%GC) − 675/*N* − % mismatch and a predicted annealing temperature of 0.3 × *T*m(primer) + 0.7*T*m (product) − 14.9, at least four primer template mismatches per primer would be tolerated at a 45 °C annealing temperature when using LCO1490 and HCO2198 and standard conditions. These conditions alone would predict COI amplification for several gammaproteobacteria (Table 1). Only *B. anthracis* fails to match on the critical 3′ dinucleotides at the COI priming sites (Table 1). Each of *Oceanobacter* sp., *P. profundum*,

Psuedoaltermonas halplanktis, *P. ingrahmii*, and *V. okutanii* has fewer than five mismatched nucleotides in the 15 most-3′ positions of the two Folmer et al. (1994) priming sites combined. Half of these grouped closest to one of the contaminated animal isolates (Fig. 1).

The dual features of having converged on base compositions typical of metazoan mitchondria and having an affinity for the surfaces of animals have conspired to allow the homologous COI priming sites of marine gammaproteobacterial species to be readily available for amplification by what were presumed to be exclusively metazoan universal DNA barcoding primers. Moreover, because the Folmer et al. (1994) primers are such a poor match to many eukaryotic groups (Table 1), associated marine surface bacteria may well be preferentially amplified under the low annealing temperature profiles typically used. Indeed, the 48 and 45 °C annealing temperatures employed for the *P. botniensis* isolates are well in keeping with barcoding standards (de Waard et al. 2008), as well as several published papers shown here to have generated contaminant COI sequences (e.g., Webb et al. 2006; Nakano and Ozawa 2007). In one of these, wherein a reference sequence was required for an invertebrate whose gut contents were being assessed, the amplification profile for COI included a 42 °C annealing temperature (Blankenship and Yayanos 2005). Unfortunately, as seen in Figure 1, the procedure resulted in a marine bacterial contaminant being used for the amphipod *Scophelocheirus schellenbergii* reference sequence.

Stringent annealing temperatures should not be expected to remediate the problem of marine bacterial contaminant amplification. The poor match of barcoding primers to metazoan genomic targets necessitates liberal amplification regimes. Our use of the COI-E primer in lieu of HCO2198 for two molluscan taxa employed a 52 °C annealing profile that nevertheless preferentially amplified two different gammaproteobacterial species (Fig. 1). Examination of the overlapping variation in marine metazoan and marine microbial priming sites (Table 1) should disabuse anyone of the notion that temperature profiles or group-specific primers can be expected to easily ameliorate this problem in a marine barcoding census (terrestrial and freshwater barcoding initiatives are, of course, unlikely to amplify marine gammaproteobacteria). If the coamplification or preferential amplification of alteromonad and vibrionid marine bacteria cannot be avoided, DNA barcoding of animals in a marine environment needs only to add a step to what is currently just a similarity matching protocol. A simple blastn query against the growing database of whole bacterial genomes should precede any barcode identification attempt so as to filter out nontarget contaminants. In fact, a thorough barcoding survey of marine and other psychrophilic bacteria may be in order so as to provide a more effective filter against clandestine amplification of marine animal-associated bacterial contaminants. It should not go unnoticed that the sow's ear of accidental bacterial barcoding has already elucidated a silk purse in the form of heretofore undescribed animal-associated marine microbial diversity.

In addition to approaching DNA barcoding of marine organisms with a degree of circumspection, there clearly is need for some corrections to comparative databases upon which the community relies for species determinations. Public databases, like those maintained by NCBI, only permit the submitter of record to invalidate or reclassify an accessioned sequence, thus putting accuracy at the mercy of the submitter's good will. BOLD, on the other hand, could take a more proactive posture by denoting suspicious and demonstrably spurious contaminant sequences by the prefixed qualifier "ex." indicating (e.g., Fig. 1) that a sequence was merely "isolated from" the named taxon.

Acknowledgments

We thank the National Science Foundation (DEB 0640463 to M.E.S.), the Richard Lounsbery Foundation, and the Swedish Research Council (grant #621-2007-5313 to C.E.). This paper was improved by thoughtful suggestions from Tripp Macdonald, Alejandro Oceguera, Anna Phillips, Dirk Steinke, and an anonymous reviewer.

References

Baumann L., Baumann P., Mandel M., Allen R.D. 1972. Taxonomy of aerobic marine eubacteria. J. Bacteriol. 110:402–429.

Bely A. E., Wray G.A. 2004. Molecular phylogeny of naidid worms (Annelida: Clitellata) based on cytochrome oxidase I. Mol. Phylogenet. Evol. 30:50–63.

Blankenship L.E., Yayanos A.A. 2005. Universal primers and PCR of gut contents to study marine invertebrate diets. Mol. Ecol. 14: 891–899.

Cho E.S., Jung C.G., Sohn S.G., Kim C. W., and Han S. J. 2007. Population genetic structure of the ark shell *Scapharca broughtonii* Schrenck from Korea, China, and Russia based on COI gene sequences. J. Mar. Biotechnol. 9:203–216.

Crick F.H. 1966. Codon-anticodon pairing: the wobble hypothesis. J. Mol. Biol. 19:548–555.

Deppe U., Richnow, H.H., Michaelis W., Antranikian G. 2005. Degradation of crude oil by an arctic microbial consortium. Extremophiles 9:461–470.

De Waard J.R., Ivanova, N.V., Hajibabaei M., Hebert P.D.N. 2008. Assembling DNA barcodes. Analytical protocols. Methods Mol. Biol. 410:275–294.

Edgar R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Folmer O., Black M., Hoen W., Lutz R., Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metozoan invertebrates. Mol. Mar. Biol. Biotechnol. 3:294–299.

Gerdes B., Brinkmeyer R., Dieckmann G., Helmke E. 2005. Influence of crude oil on changes of bacterial communities in Arctic sea-ice. FEMS Microbiol. Ecol. 53:129–139.

Halanych K.M., Janosik A.M. 2006. A review of molecular markers used for annelid phylogenetics. Integr. Comp. Biol. 46:533–543.

Hebert P.D.N., Stoeckle M.Y., Zemlak S.T., Francis C.M. 2004. Identification of birds through DNA barcodes. PLoS Biol. 2:1657–1663.

Holmström C., James S., Neilan B.A., White D.C., Kjelleberg S. 1998. *Pseudoalteromonas tunicata* sp. nov., a bacterium that produces antifouling agents. Int. J. Syst. Bacteriol. 48:1205–1212.

Ivanova E.P., Chun J., Romanenko L.A., Matte M.H., Mikhailov V.V., Huq A., Colwell R.R. 2000. Reclassification of *Alteromonas distincta* Romanenko et al., 1995 as *Pseudoalteromonas distincta* comb. nov. Int. J. Syst. Evol. Microbiol. 50:141–144.

Ivanova E.P., Kiprianova E.A., Mikhailov V.V., Levanova G.F., Garagulya A.D., Gorshkova N.M., Yumoto N., Yoshikawa S. 1996.

Characterization and identification of marine *Alteromonas nigrifaciens* strains and emendation of the description. Int. J. Syst. Bacteriol. 46:223–228.
Ivanova E.P., Shevchenko L.S., Sawabe T., Lysenko A.M., Svetashev V.I., Gorshkova N.M., Satomi M., Christen R., Mikhailov V.V. 2002. *Pseudoalteromonas maricaloris* sp. nov., isolated from an Australian sponge, and reclassification of [*Pseudoalteromonas aurantia*] NCIMB 2033 as *Pseudoalteromonas flavipulchra* sp. nov. Int. J. Syst. Evol. Microbiol. 52:263–271.
Ivanova N.V., Zemlak T. S., Hanner R.H., Hebert P.D.N. 2007. Universal primer cocktails for fish DNA barcoding. Mol. Ecol. Notes 7:544–548.
Khachane A.N., Timmis K.N., dos Santos M., Vitor A.P. 2005. Uracil content of 16S rRNA of thermophilic and psychrophilic prokaryotes correlates inversely with their optimal growth temperatures. Nucleic Acids Res. 33:4016–4022.
Martin A.P. 1995. Metabolic rate and directional nucleotide substitution in animal mitochondrial DNA. Mol. Biol. Evol. 12:1124–1131.
McGarey D.J., Milanesi L., Foley D.P., Reyes B. Jr., Frye L.C., Lim D.V. 1991. The role of motile aeromonads in the fish disease, ulcerative disease syndrome (UDS). Experientia 47:441–444.
Mikhailov V.V., Romanenko L.A., Ivanova E.P. 2006. The genus *Alteromonas* and related proteobacteria. Prokaryotes 6:597–645.
Nakano T., Ozawa T. 2007. Worldwide phylogeography of limpets of the order Patellogastropoda: molecular, morphological and palaeontological evidence. J. Mollusc. Stud. 73:79–99.
Nilsson R.H., Ryberg M., Kristiansson E., Abarenkov K., Larsson K.H., Kõljalg U. 2006. Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 1:e59.
Perna N.T., Kocher T.D. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41:353–358.
Saitou N., Nei M. 1987. The Neighbor-Joining method—a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.
Simidu U., Kita-Tsukamoto K., Yasumoto T., Yotsu M. 1990. Taxonomy of four bacterial strains that produce tetrodotoxin. Int. J. Syst. Bacteriol. 40:331–336.
Smith M.A., Rodriguez J.J., Whitfield J.B., Deans A.R., Janzen D.H., Hallwachs W., Hebert P.D. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. PNAS 105: 12359–12364.
Song H., Buhay J.E., Whiting M.F., Crandall K.A. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. PNAS 105:13486–13491.
Stamatakis A., Ott M., Ludwig T. 2005. RAxML-OMP: an efficient program for phylogenetic inference on SMPs. Lect. Notes Comput. Sci. 3606:288–302.
Swofford D. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b. Computer software and manual. Sunderland (MA): Sinauer Associates.
Tao N., Richardson R., Bruno W., Kuiken C. 2008. Find Model. Available from: http://www.hiv.lanl.gov/content/sequence/ findmodel/findmodel.html (date last accessed June 22, 2009).
Traub R.J., Hobbs R.P., Adams P.J., Behnke J.M., Harris P.D., Thompson R.C.A. 2007. A case of mistaken identity—reappraisal of the species of canid and felid hookworms (*Ancylostoma*) present in Australia and India. Parasitology 134:113–119.
Vilgalys R. 2003. Taxonomic misidentification in public DNA databases. New Phytol. 160:4–5.
Webb K.E., Barnes D.K.A., Clark M.S., Bowden D.A. 2006. DNA barcoding: a molecular tool to identify Antarctic marine larvae. Deep Sea Res. Part II 53:1053–1060.
Zhang W., Wendel J.F., Clark L.G. 1997. Bamboozled again! Inadvertent isolation of fungal rDNA sequences from bamboos (Poaceae: Bambusoideae). Mol. Phylogenet. Evol. 8:205–217.

Received 28 September 2008; reviews returned 21 January 2009; accepted 5 April 2009
Associate Editor: Karl Kjer

APPENDIX 1

Accession numbers for taxa used in this study follow here:

Alteromonas macleodii	NC 011138	
Bacillus anthracis	NC 007530	Anthrax
Cellvibrio japonicus	NC 010995	
ex. *Cnemidocarpa verrucosa*	AJ830012	Ascidian
Colwellia psychrerythraea	NC 003910	
Danio rerio	NC 002333	Zebrafish
ex. *Engraulis japonica*	EU266375	Japanese anchovy
ex. *Eurythenes gryllus*	AY830434	Amphipod
ex. *Eurythenes gryllus*	AY830436	Amphipod
Hahella chejuensis	NC 007645	
Hydra oligactis	EU 237491	Brown hydra
Legionella pneumophila	NC 0065368	Legionnaire's
ex. *Lingula anatina*	AB056462	Brachiopod
Lingula anatina	AB 178773	Brachiopod
Lottia digitalis	DQ 238599	Finger limpet
Lumbricus terrestris	NC 001673	Earthworm
Marinobacter aquaeolei	NC 008740	
ex. *Munnopsis* sp.	EF682272	Isopod
Mytilus edulis	AY 84747	Blue mussel
Oceanobacter sp.	NZ AAQH 00000000	
Orzya sativa Japonica	NC 011033	Rice
ex. *Paranais botniensis*	FJ868210	Oligochaete
ex. *Paranais botniensis*	FJ868211	Oligochaete
Parhyale hawaiensis	AY 639937	Amphipod
ex. *Patelloida striata*	AB161589	Striate limpet
ex. *Penitella penita*	FJ868209	Piddock clam
Photobacterium profundum	NC 006370	
Pseudoalteromonas atlantica	NC 008228	
Pseudoalteromonas haloplanktis	NC 007481	
Pseudomonas putida	NC 010501	
Psychromonas ingrahamii	NC 008709	
Reclinomonas americana	AF 007261	Loricate jakobid
Ruthia magnifica	NC 008610	
Saccharophagus degradans	NC 007912	
ex. *Scapharca broughtonii*	DQ435188	Ark clam
ex. *Scapharca broughtonii*	DQ435111	Ark clam
ex. *Scopelocheirus shcellenbergi*	AY830439	Amphipod
Scypholanceola sp.	EF 989695	Amphipod
Shewanella amazonensis	NC 008700	
Shewanella baltica	NC 009665	
Shewanella dentrificans	NC 007954	
Shewanella frigidimarina	NC 008345	
Shewanella pealeana	NC 009901	
Vesicomyosocius okutanii	NC 009465	
Vibrio parahaemolyticus	NC 004605	
Vibrio shilonii	NZ ABCH 00000000	
Vibrio vulnificus	NC 005140	
Xestospongia muta	EU 237490	
ex. *Xylotra setacea*	FJ868208	Shipworm