Received: 25 September 2019 | Revised: 8 January 2020 | Accepted: 8 January 2020

DOI: 10.1002/edn3.72

ORIGINAL ARTICLE

Environmental DNA
Open Access
Dedicated to the study and use of environmental DNA for basic and applied sciences

WILEY

Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates

Elena Valsecchi[1,2] | Jonas Bylemans[3] | Simon J. Goodman[4] | Roberto Lombardi[1] | Ian Carr[5] | Laura Castellano[6] | Andrea Galimberti[7] | Paolo Galli[1,2]

[1]Department of Environmental and Earth Sciences, University of Milano-Bicocca, Milan, Italy

[2]MaRHE Center, Magoodhoo Island, Faafu Atoll, Republic of Maldives

[3]Fondazione Edmund Mach, S. Michele all'Adige (TN), Italy

[4]School of Biology, University of Leeds, Leeds, UK

[5]Leeds Institute of Molecular Medicine, St James's University Hospital, Leeds, UK

[6]Acquario di Genova, Costa Edutainment SPA, Genoa, Italy

[7]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy

Correspondence
Elena Valsecchi, Department of Environmental and Earth Sciences, University of Milano-Bicocca, Milan, Italy.
Email: elena.valsecchi@unimib.it

Abstract

Metabarcoding studies using environmental DNA (eDNA) and high-throughput sequencing (HTS) are rapidly becoming an important tool for assessing and monitoring marine biodiversity, detecting invasive species, and supporting basic ecological research. Several barcode loci targeting teleost fish and elasmobranchs have previously been developed, but to date primer sets focusing on other marine megafauna, such as marine mammals, have received less attention. Similarly, there have been few attempts to identify potentially "universal" barcode loci which may be informative across multiple marine vertebrate orders. Here we describe the design and validation of two new sets of primers targeting hypervariable regions of the vertebrate mitochondrial 12S and 16S rRNA genes, which have conserved priming sites across virtually all cetaceans, pinnipeds, elasmobranchs, boney fish, sea turtles, and birds, and amplify fragments with consistently high levels of taxonomically diagnostic sequence variation. "In silico" validation using the OBITOOLS software showed our new barcode loci outperformed most existing vertebrate barcode loci for taxon detection and resolution. We also evaluated sequence diversity and taxonomic resolution of the new barcode loci in 680 complete marine mammal mitochondrial genomes demonstrating that they are effective at resolving amplicons for most taxa to the species level. Finally, we evaluated the performance of the primer sets with eDNA samples from aquarium communities with known species composition. These new primers will potentially allow surveys of complete marine vertebrate communities in single HTS metabarcoding assessments, simplifying workflows, reducing costs, and increasing accessibility to a wider range of investigators.

KEYWORDS

12S, 16S, cetaceans, fish, pinnipeds, sea turtles

1 | INTRODUCTION

The use of DNA fragments extracted from environmental sources (e.g., soil and water samples) is becoming a well-established tool for monitoring biodiversity (Deiner et al., 2017; Jarman, Berry, & Bunce, 2018). Within the marine environment, such eDNA surveys have been used to assess the diet of marine species (Deagle, Chiaradia, McInnes, & Jarman, 2010; McInnes et al., 2017; Peters et al., 2015), monitor the species diversity of marine communities (Port et al., 2016; Sigsgaard et al., 2017), determine the presence/absence of invasive species (Borrell, Miralles, Do Huu, Mohammed-Geba, & Garcia-Vazquez, 2017), and obtain estimates of population genetic diversity (Sigsgaard et al., 2016). Community biodiversity surveys from eDNA (i.e., eDNA metabarcoding) rely on primers targeting specific taxonomic groups and high-throughput sequencing (HTS) to amplify and sequence barcoding regions from all species of interest (Creer et al., 2016). While DNA metabarcoding primers have been developed to target several individual marine taxonomic groups (e.g., elasmobranchs, teleost fish, cephalopods, and crustaceans) (e.g., Bylemans, Gleeson, Hardy, & Furlan, 2018; Jarman, Redd, & Gales, 2006; Komai, Gotoh, Sado, & Miya, 2019; Miya et al., 2015), to date no primer sets have been specifically designed to maximize the recovery and identification of marine mammals.

The few marine eDNA studies focussing on marine mammals used targeted species-specific assays (Baker, Steel, Nieukirk, & Klinck, 2018; Foote et al., 2012) or used universal fish-specific primers as a proxy to assess the total vertebrate biodiversity (e.g., Andruszkiewicz et al., 2017). All these approaches present at least one drawback when aiming to detect the presence of cetacean and pinniped species within an eDNA sample. Primer sets designed for one group, that is, "fish-specific" primers, might amplify eDNA from other taxa, but unquantified primer mismatch risk reduced detection rates and the introduction of biases in HTS results (e.g., Elbrecht, Hebert, & Steinke, 2018). Where primers are designed for the species of interest, amplicon target size may also be a consideration. The instability and short life span of eDNA molecules (e.g., Thomsen et al., 2012) favors the use of short and informative (hypervariable) DNA regions, such as the mitochondrial 12S, 16S, and cytochrome oxidase I (*COI*) genes, with amplicons of typically around 100–200 bp although larger mtDNA fragments have been shown to be successfully amplified from eDNA samples (Deiner et al., 2017).

There is therefore a need to develop marine mammal-specific primers suitable for metabarcoding analysis from marine eDNA samples. Mitochondrial 12S and 16S regions provide suitable targets since their sequence variation provides good taxonomic resolution for macro-eukaryotes, while also maintaining conserved sites across regions for siting primers (Deagle, Jarman, Coissac, Pompanon, & Taberlet, 2014). The design criteria for such primers should be a) primer sites which are conserved among marine mammal groups, while amplifying hypervariable DNA fragments for taxonomic resolution; b) where possible, identify marine mammal-specific priming sites in order to reduce cross-amplification with human DNA, thus lessening contamination risks from investigators, swimmers, or other biological residues left by humans; and c) for each primer set, evaluate predicted binding efficiency and amplicon sequence diversity for other marine vertebrates such as fish and sea turtles (when necessary allowing for a single degenerate base per primer). This final point would give a more accurate understanding of primer specificity and suitability for use with other vertebrate groups. Primer sets have been proved to have reliable affinity across multiple vertebrate orders and could support more efficient and cost-effective eDNA biodiversity surveys.

In this paper, we report on the development of novel "universal" marine vertebrate eDNA primers meeting the above criteria, their "in silico" validation against a large marine vertebrate NCBI-GenBank dataset, and their successful initial application and validation with a HTS analysis of environmental water samples collected from a public marine aquarium. Finally, we test for inter- and intraspecific variation over large mitogenomic datasets available for marine mammal species, presenting ready-to-use guidelines for the accurate selection of primer set of choice in specific marine mammal studies.

2 | MATERIALS AND METHODS

2.1 | Initial design of primer sets

Seventy-one complete mitochondrial genome sequences, representative of most marine vertebrate groups (fish, sea turtles, birds, and marine mammals), were retrieved from GenBank and used for initial primer development (Appendix S1). The selected sequences represented 30 marine vertebrate families, including most marine mammal families (all three Pinniped families, both Sirenian and 13 Cetacean families). The selection comprised all Cetacean species occurring in the Mediterranean Sea. In addition, four human mitochondrial genomes representative of the four main human haplogroups (i.e., haplotypes 16, 31, 33, and 52 in Ingman, Kaessmann, Pääbo, & Gyllensten, 2000) were included (Appendix S1) in order to design primers with reduced amplification efficiency for human DNA. All sequences (n = 75) were aligned with the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters, and the complete ribosomal 12S and 16S genes were isolated. Potential sites for metabarcoding primers were identified by manually searching for suitable locations within alignments. Gene regions were considered suitable for designing metabarcoding primers if they encompassed a short (80–230 bp) highly variable fragment, required for species discrimination, and were flanked by highly conserved sites for situating primers. Where possible (i.e., when enough intramammal variation was found in proximity of the priming sites), we also tried to design, for each candidate locus, alternative primers which minimized the probability of amplifying human targets, by ensuring mismatches between the primers and human templates. Such variants could be preferentially used in studies specifically targeting marine mammals.

2.2 | Primer evaluation and validation

Three approaches were used to assess primer performance. Firstly, primers were evaluated in silico in two steps: (a) predicted primer binding and amplicon sequence diversity were assessed using the ecoPCR scripts within the OBITOOLS software package (Boyer et al., 2016; Ficetola et al., 2010); and (b) 680 complete marine mammal mitogenome sequences deposited in GenBank were used to quantify sequence diversity for the primer target regions within marine mammal Families, Genera, and species, to provide recommendations on taxonomic resolution utility of primer sets for specific taxa. Secondly, the performance of the primers was evaluated in vitro using tissue-derived DNA extracts with varying levels of degradation. Finally, eDNA samples, obtained from tanks of known species composition at the Aquarium of Genoa (Italy) as a proxy for "real-world" environmental samples, were used to assess the metabarcoding performance of the primers.

2.2.1 | In silico primer evaluation

An in silico approach was used to assess the universality of the newly designed primers against all standard nucleotide sequences in the NCBI-GenBank data repository (accessed April 2019) for three taxonomic groups: (a) vertebrates (excluding cetaceans), (b) cetaceans only, and (c) invertebrates. As a performance benchmark, the newly designed primers were compared against 12S-V5 (Riaz et al., 2011), one of the most commonly used metabarcoding primers targeting vertebrates.

The ecoPCR script was used to simulate an in silico PCR and extract the amplifiable barcoding regions for each primer pair while allowing for a maximum of three base-pair (bp) mismatches between the primers and template DNA. Barcode regions shorter than 50 bp and longer than 400 bp were not considered. Subsequently, the OBIGREP command was used to extract sequences that were reliably assigned to a species-level taxonomy. Ambiguous species-level identifications (i.e., sequences with "sp." and "aff.," in the definition of the sequence) and nuclear pseudogenes were excluded from the final sequence database. The OBIUNIQ command was then used to remove duplicate records for each species. Sequences were classified according to their higher taxonomy (i.e., vertebrates [excluding Cetacean species], cetaceans, and invertebrates) before summarizing the data into a tabular format for further analyses using R version 3.5.2 (R Development Core Team 2010). Finally, the taxonomic resolution of the different primers was assessed by splitting the data into their higher taxonomy and running the ECOTAXSPECIFICITY script with three different thresholds for barcode similarity (i.e., sequences were considered different if they have 1, 3, or 5 bp differences). The data obtained from the in silico analyses were imported into R, and the packages tidyverse (Wickham, 2016) and gridExtra (Auguie, 2016) were used to construct summary figures to evaluate the taxonomic coverage, the specificity, and taxonomic resolution for each primer pair.

Finally, we downloaded 680 complete marine mammal mitogenomes from GenBank and evaluated levels of polymorphism within Families, Genera, and species at the two proposed loci. Complete 12S and 16S genes were extracted from the retrieved sequences and aligned for each type of taxonomic comparison, and the number of variable sites recorded within the two loci amplicons was reported.

2.2.2 | In vitro primer evaluation

Tissue-derived DNA extracts were used to assess the performance of the newly designed primer pairs in vitro and optimize amplification conditions. DNA extracts of diverse marine vertebrate groups (cetaceans, pinnipeds, sea turtles, and fish) were used as templates for PCR amplification (see Table 4). The 13 DNA templates were purposely selected to have different levels of degradation, being extracted with different techniques and spanning 1–31 years since extraction, in order to evaluate the ability of the primer to amplify low-quality DNA. High-quality DNA extracts were obtained from fresh samples (i.e., muscle, skin, or blood) using the Qiagen DNeasy Blood & Tissue extraction kit following the manufacturer's protocols. Low-quality DNA extracts consisted of phenol–chloroform-extracted DNA which was over 25 years old or DNA extracts obtained from boiling tissue samples in a buffer solution (Valsecchi, 1998). For each primer pair and each of the DNA extracts a (single, duplicate, triplicate), PCR was performed in reaction mixes consisting of 0.025 u/µl of GoTaq G2 Flexi DNA polymerase (Promega); 1X Green GoTaq Flexi Reaction Buffer (Promega); 1.25–2 mM $MgCl_2$ (Promega); 0.2 mM dNTPs (Promega); and 0.25–0.75 µM of each primer and dH_2O to reach a final volume of 20 µl. Thermal cycling conditions followed a touch-down PCR protocol with annealing temperatures depending on primer pairs: 10/10/18 cycles at 54/55/56°C for MarVer1 and 8/10/10/10 cycles at 54/55/56/57°C for MarVer3. After an initial denaturation step of 4 min at 94°C, each of the 38 cycles consisted of a 30 s at 95°C, 30 s at the primer specific annealing temperatures as described above, followed by 40 s at 72°C. The final extension consisted of 5 min at 72°C. To confirm the amplification of the desired amplicon, PCR products were visually assessed using gel electrophoresis and Sanger sequenced (GENEWIZ, UK; data not shown).

2.2.3 | Evaluation of primer performance with environmental samples

Environmental DNA (eDNA) samples derived from water collected from six tanks of the Aquarium of Genoa, Italy, in June 2018, were also employed to further validate the performance of the two primer sets. The tanks contained from 1 to 14 known vertebrate species and were named after their main hosted species or typology: (a) "Manatee," (b) "Dolphin," (c) "Shark," (d) "Seal," (e) "Penguin," and (f) "Rocky shore"—a multispecies tank hosting fish and invertebrates typical of Mediterranean rocky shores. Two tanks (Dolphin and Seal) were single species, the Penguin included two penguin species (*Spheniscus*

demersus and *Pygoscelis papua*), and the Manatee included two teleost species beside the manatees, while the Shark and the Rocky shore tanks included a combination of cartilaginous and bony fish.

For each tank, a total of 13.5 L of water was collected from the water surface using a sterile graduated 2,000 ml glass cylinder, while wearing sterile gloves, and stored within a 15-L sterile container in order to homogenize the water sample and avoid stochastic variability due to sampling of small volumes. For each tank, 3 × 1.5 L and 3 × 3 L replicates (total six replicates per tank) were then aliquoted from the larger sample. To capture eDNA, immediately after aliquoting, each of the six replicates was filtered using individual 0.45 μm pore size nitrocellulose filters, using a BioSart® 100 filtration system (Sartorius). After filtering, membranes were placed on ice for transport to the University of Milano-Bicocca and subsequently stored at −20°C. Two weeks later, eDNA was extracted from the filter membranes using a DNeasy PowerSoil Kit® (Qiagen), following the manufacturer's protocol.

For each of the two novel primer sets, PCR performance with the eDNA extractions was initially evaluated using the same PCR conditions as the "in vitro" validation. After confirming amplification of amplicons in the expected size range with eDNA templates, indexed forward and reverse sequencing primers were created for each primer set, comprising (5′–3′): an 8 bp Illumina barcode tag—4 random nucleotides—amplification primer sequence, and sourced from Sigma, UK. Eight forward primer indexes were combined with 12 reverse primer indexes, to allow pooling of amplicons from up to 96 uniquely identifiable samples within a single sequencing library (Taberlet, Bonin, Zinger, & Coissac, 2018). Trial amplifications with the Illumina barcode-tagged primers suggested their yield, and specificity was unchanged, and so the previously optimized PCR conditions were used to generate amplicons for MiSeq sequencing. For each locus, eDNA was amplified in triplicate in 40 μl final PCR volume; 5 μl of each replicate was used to check for successful amplification via agarose gel electrophoresis, and the remainder combined to yield a single pool for each sample.

Each sample amplicon pool was first assessed for fragment size distribution using an Agilent TapeStation and cleaned with AMPure beads (Beckman Coulter), following the manufacturer's protocol, to remove primer dimers. The cleaned samples were quantified with a Qubit fluorometer, and then for each metabarcoding locus, separate Illumina NEBNext Ultra DNA libraries were generated, with the pooled samples in equimolar ratios. Prior to sequencing, each library was further assessed for fragment size distribution and DNA quantity by Agilent TapeStation and Qubit fluorometer. The library for locus MarVer1 (see Results) was sequenced in a 150 bp paired-end lane, and locus MarVer3 (see Results) in a 250 bp paired-end lane, using an Illumina MiSeq sequencer at the University of Leeds Genomics Facility, St James's Hospital.

2.2.4 | Bioinformatics for environmental HTS data

Paired reads were first screened for the presence of the expected primer and index sequence combinations to exclude off-target amplicons. The reads were then combined to generate the insert sequence, and the sequence of the random nucleotide region was noted, such that only one instance of an insert per sample with the sample random nucleotide fingerprint was saved to a sample-specific file (i.e., to avoid PCR duplicates and chimeric sequences). The insert data were aggregated to create a count matrix containing the occurrence of each unique sequence in each sample. The taxonomic origin of each insert was determined by blasting their sequence against a local instance of the GenBank NT database (Nucleotide [Internet]). The level of homology of insert to the hit sequence was noted, as was the species name of the hit sequence. The taxonomic hierarchy for each unique insert was generated by searching a local instance of the ITIS database (ITIS [Internet]) with the annotated GenBank species name. The count matrix and taxonomic hierarchy for all annotated unique sequences were then aggregated into values for equivalent molecular operational taxonomic units (MOTUs), by combining all inserts with a set homology (≥98%) to the GenBank hit at a specified taxonomic level (i.e., "order," "family," "genus," or "species"), using bespoke software (available on request). Summaries and visualizations of read counts for different taxonomic levels were generated using the R package "Phyloseq" (McMurdie & Holmes, 2013).

3 | RESULTS

3.1 | Description of primer sets

From the initial evaluation of aligned marine vertebrate mitochondrial sequences, three hypervariable regions flanked by conserved motifs were identified, two within the 12S gene, which we term MarVer1 and MarVer2, and one in the 16S gene, which we term MarVer3. PCR primer pairs were designed for all three MarVer loci. Here we focus on MarVer1 and MarVer3 (Table 1), which yield the largest thus more informative amplicons (ca 202 bp and ca 245 bp, respectively), for all levels of further validation. Primers were designed to amplify the target regions in any of the 71 taxa selected representative marine vertebrates, and to allow for variable sites between different vertebrate groups within the primer sequence, single degenerate bases were introduced for MarVer1R and MarVer3F. Appendix S2 shows variability at the priming sites across the eight marine vertebrate categories. Amplicon variability across the 71 taxa (plus the four human sequences) recorded in the regions targeted by the proposed markers is shown in Appendix S3. Further detail on the in silico evaluation of Marver2 is provided in Appendix S6.

3.1.1 | MarVer1

MarVer1 primer set (abbr. MV1) targets a hypervariable region of the 12S gene, amplifying a fragment of about 199–212 bp (Table 2). It partially overlaps with loci Tele02 (Taberlet et al., 2018), Tele03 (as named by Taberlet et al., 2018, corresponding to MiFish-U in Miya et al., 2015), and Elas01 (as named by Taberlet et al., 2018,

TABLE 1 Sequences of the two described primer sets

Locus	Primer ID	Primer sequence (5′–3′)	Size	Region	Average amplicon size
MarVer1	MarVer1F	CGTGCCAGCCACCGCG	16 bp	12S	*ca.* 202 bp
	MarVer1R	GGGTATCTAATCCYAGTTTG	20 bp		
MarVer3	MarVer3F	AGACGAGAAGACCCTRTG	18 bp	16S	*ca.* 245 bp
	MarVer3R	GGATTGCGCTGTTATCCC	18 bp		

corresponding to MiFish-E in Miya et al., 2015), targeting bony and cartilaginous fishes (Figure 1). The forward primer MarVer1F differs from the forward primers of previously described loci, in that by shifting 5–12 bp at the 5′ end, it skips variable sites distinguishing bony from cartilaginous fishes, while gaining, at the 3′ end, bases which are conserved across all surveyed marine vertebrates.

3.1.2 | MarVer3

MarVer3 (abbr. MV3) amplifies a variable region of the 16S gene, producing amplicons ranging in size from 232 to 274 bp in the 71 marine vertebrate taxa tested (Table 2). MarVer3 is partially covered by locus Mamm02 (Taberlet et al., 2018, see Figure 1), but targets a fragment twice as long: for example, in Odontocetes, MarVer3 amplifies a 233 bp fragment versus a 115 bp amplicon for Mamm02. The reverse primer (MarVer3R) was the only one of the presented oligonucleotides to be truly "universal," as it was found to be fully conserved across all tested marine vertebrates (Table 2). The MarVer3 amplicon sequence resolved 69 of the 71 tested marine vertebrate species. The unresolved species fall into the Delphinidae family: *Sousa chinensis* and *Tursiops truncatus* sharing one amplicon sequence and *Tursiops aduncus* and *Delphinus capensis* sharing another.

3.2 | Primer evaluation and validation

3.2.1 | In silico primer evaluation

For MarVer1 and MarVer3 primer pairs and higher taxonomic groups, the total number of unique taxa for which the in silico amplification recovered target sequences is given in Figure 2. The results show that the MarVer3 primer pair amplified DNA from the most vertebrate and cetacean taxa. However, allowing for up to three mismatches per priming site, this primer pair also successfully amplifies the DNA of invertebrate species thus indicating that it has a low overall specificity to the intended taxonomic targets (Figure 2). The MarVer1 primer pair amplified DNA from slightly fewer target taxa than the commonly used 12S-V5 primers.

The proportion of sequences amplified for each higher taxonomic group is a function of the bp mismatches between the primers, and template DNA is shown in Figure 3. The results of the commonly used 12S-V5 primer pair show that a high number of vertebrate sequences have very few bp mismatches while the recovered nonvertebrate sequences generally have ≥5 bp mismatches at both primer binding regions (Figure 3). The results of both the MarVer1 and MarVer3 primers show that even with a low number of bp mismatches (i.e., ≥2 mismatches for both the forward and reverse primers) a significant proportion of the amplified sequences belong to nonvertebrate taxa.

The taxonomic resolution power of the different primer pairs was evaluated using both the Cetacean and Vertebrate sequences, and the results are summarized in Figure 4. Overall, the commonly used 12S-V5 metabarcoding primers have a lower resolution capacity compared to our newly designed primers (for both the Cetacean and Vertebrate taxa), with our primers assigning > 25% more sequences to the correct species-level taxonomy (Figure 4). For the newly designed primers, no obvious differences are observed in their taxonomic assignment power, with MarVer1 and MarVer3 generally assigning a similar percentage of the sequences to the correct family-, genus- and species-level taxonomy (Figure 4).

3.2.2 | Primer set resolution for marine mammal taxonomic detection

Table 3 shows the results of the comparison performed on 680 GenBank complete marine mammal mitogenome sequences (GenBank accession numbers shown in Appendix S4), in order to evaluate levels of polymorphism within Families, Genera, and species.

Family level

Both targeted regions contained high genetic variability within the seven analyzed marine mammal Families (pinnipeds [2], Mysticetes [1], and Odontocetes [4]). The DNA fragment amplified by MarVer3 primer set (16S region) proved to be the most diverse, highlighting 59 variable sites within the Phocoidea and over 40 in the Otariidae, Ziphiidae, and Delphinidae Families (Table 3).

Genus level

Nine Genera, each including 2–5 species, were assessed for within-genus variability, including pairwise congeneric-species comparisons (Table 3). MarVer3 consistently revealed the highest levels of polymorphism: 23 of the 31 intergeneric pairwise comparisons show differences of at least 1 bp, and in 20 comparisons, MarVer3 performed better (i.e., showed a higher number of variable positions) than MarVer1. Conversely, MarVer1 showed a higher number of variable sites than MarEvr3 in 6 out of the 31 congeneric-species

TABLE 2 Characteristics of the amplicons produced by each of the two MarVer primer sets for the 71 marine vertebrate species used for primers' design. The overall amplicon size is the mean value of the amplicon sizes recorded in the eight marine vertebrate groups analyzed

	CODO	CMYS	PINN	SIRE	STUR	SBIR	TELE	ELAS	Overall
MarVer01									
Approximate amplicon size (bp)	199	199	197	199	211	209	197	212	*ca.* 202
n variable sites	77	31	33	18	36	45	80	40	171
Resolved over tested species	36/36	12/12	3/3	2/2	6/6	2/2	7/7	3/3	71/71
Percentage of variable sites (%)	38.7	15.6	16.8	9.0	17.1	21.5	40.6	18.9	84.7
Nucleotide *p*-distance	0.084	0.042	0.097	0.084	0.059	0.157	0.176	0.090	0.196
MarVer03									
Approximate amplicon size (bp)	233	232	240	235	240	246	274	265	*ca.* 245
n variable sites	79	41	53	36	50	31	142	37	251
Resolved over tested species	34/36	12/12	3/3	2/2	6/6	2/2	7/7	3/3	69/71
Percentage of variable sites (%)	33.9	17.7	22.1	15.3	20.8	12.6	51.8	14.0	102.4
Nucleotide *p*-distance	0.052	0.040	0.098	0.081	0.054	0.058	0.151	0.046	0.167

	name	region	amplicon size	targeted group	reference
1	Tele02, Tele03, Elas02, Mi-Fish-E, Mfish-U	12S	167 bp	Fishes	Taberlet et al. 2018, Miya et al. 2015
2	12S-5V, Vert1	12S	97 bp	Vertebrates	Riaz et al. 2011, Taberlet et al. 2018
3	12S	12S	430 bp	Animalia	Yang et al. 2014
4	Mamm01, Tele01	12S	68 bp	Mammals, Fishes	Taberlet et al. 2018, Valentini et al. 2016
5	16S	16S	500 bp	Animalia	Yang et al. 2014
6	16S	16S	105 bp	Vertebrates	Riaz et al. 2011
7	Mamm02	16S	74 bp	Mammals	Taberlet et al. 2018, Giguet-Covex et al. 2014

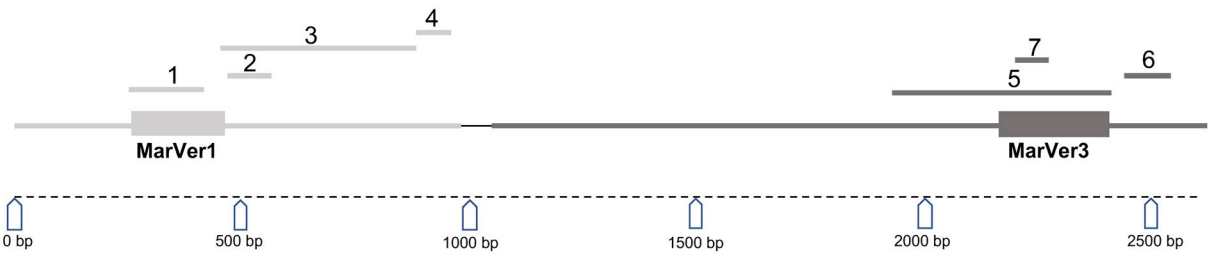

FIGURE 1 Map of the regions amplified by the newly presented primer sets within the 12S (light gray) and 16S (dark gray) genes. The positions of some of most commonly used barcode markers used for detecting vertebrate groups are shown for comparison. The size marker at the bottom refers to the 12S and 16S mtDNA fragment from position 72 to position 2,690 in the stripe dolphin (*Stenella coeruleoalba*) complete mitogenome (GenBank accession number NC_012053)

pairwise comparisons. Within all three *Tursiops* spp. pairwise comparisons, MarVer1 was the locus showing the highest variability. In four comparisons, MarVer1 and MarVer3 exhibited the same number of variable sites in the 31 comparisons. Only in one of these congeneric comparisons (*Neophocena asiaeorientalis* [*n* = 4] versus *Neophocena phocaenoides* [*n* = 8]), both MarVer1 and MarVer3 showed no variability.

Species level

Genetic variability was investigated also below the nominal species level. This could be tested only on the few species for which large enough sample sizes were available from GenBank to evaluate intraspecific variation. We assessed 14 marine mammal species for which mitogenomic data were available for a number of individuals ranging from 2 (*Megaptera novaeangliae* and *Dugong*

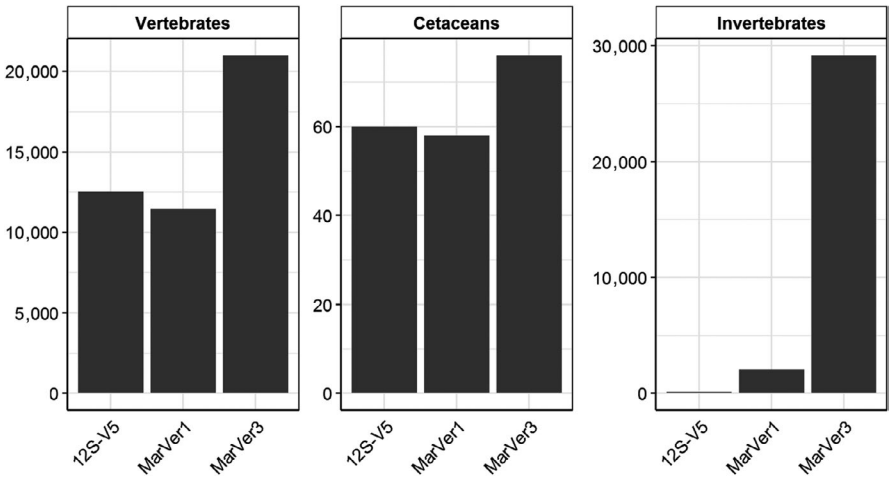

FIGURE 2 The total number of unique taxa (*y*-axis) recovered by the different primer pairs (*x*-axis). The results are shown for all three higher taxonomic groups (i.e., vertebrates—excluding cetacean species, cetaceans, and invertebrates) considered during the analyses

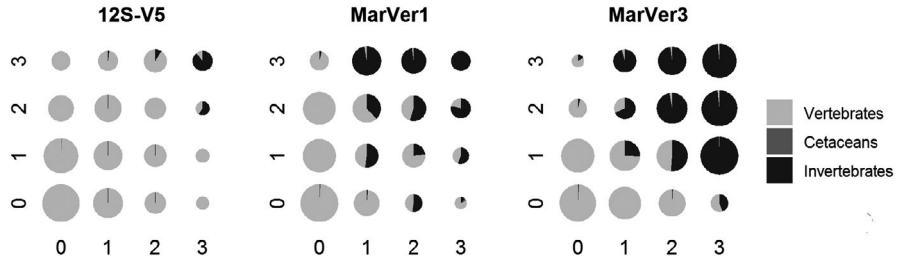

FIGURE 3 The proportion of sequences amplified for each higher taxonomic group as a function of the bp mismatches between the template DNA and the forward (*y*-axis) and reverse (*x*-axis) primers. The size of the pie charts is proportional (on a log scale) to the total number of sequences recovered for a given number of bp mismatches between the forward and reverse primer

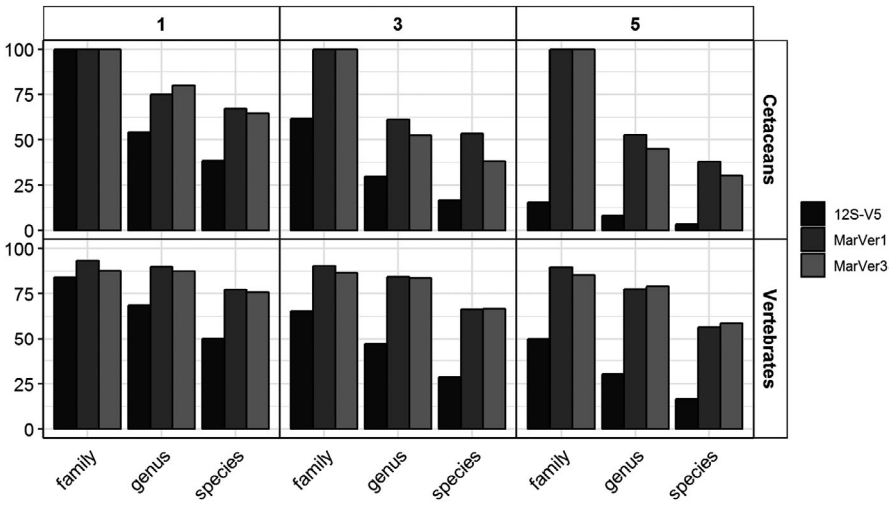

FIGURE 4 The percentage of sequences correctly identified (*y*-axis) to the family-, genus- and species-level taxonomy (*x*-axis) for the different primer pairs. Results are shown for all sequences belonging to the cetaceans and vertebrates (horizontal panels) and using different threshold values for barcode similarity (i.e., sequences were considered different if they have 1, 3, or 5 bp differences) (vertical panels)

dugong) to 152 (*Balaenoptera physalus*) individuals (Table 3). The 14 species were representative of seven marine mammal Families: Dugongidae (1 species), Otariidae (2 species), Phocidae (1 species), Balaenopteridae (2 species), Delphinidae (4 species), Ziphiidae (3 species), and Phocoenidae (1 species). All sequences were retrieved from GenBank (see Appendix S4), with the exception of 22 unpublished *Pusa caspica* sequences (provided by SG). In only one of the 14 species (*Eumetopias jubatus*, fam. Otariidae) were none of the two loci polymorphic (11 individuals compared). In the other six cases (potentially uninformative), singletons were found at one or both loci. In the remaining seven instances, some level of informative variability was found either in both (three cases) or in one of the two loci (four cases). In some occurrences, MarVer1 was found to be the most informative locus (e.g., 12 variable sites in *B. physalus*, $n = 152$), in others MarVer3 (e.g., six variable sites in *Stenella longirostris*, $n = 104$).

3.2.3 | In vitro primer evaluation

PCR amplicons of the expected size were generated by the two primer sets in all the 13 DNA extracts (Table 4). Sanger sequencing test performed on PCR products confirmed the amplification of both the correct 12S/16S fragments targeted by MarVer1 and MarVer3 loci and the correct species from which the tissue DNA samples originated.

3.2.4 | Application to environmental samples

Amplicons of the expected size were obtained also from all 36 water samples collected at the Genoa Aquarium with the two primer sets MarVer1 and MarVer3; thus, we proceeded with HTS metabarcoding evaluation. After sequence quality filtering and demultiplexing, annotated read counts per sample ranged from 682 to 52,478 for MarVer1 and 1,025 to 43,003 for MarVer3, with combined reads per tank of 10,750 to 158,950 for MarVer1, and 13,251 to 232,015 for MarVer3 (see Table 5, Figure 5, Appendix S5).

The percentage of resident vertebrate species with amplicons annotated to the species level within each tank ranged between 0% (Tank 6, Rocky shore) and 100% (Tank 2, Dolphin and Tank 5, Seal), mean 50.4% for MarVer1; and between 66.6% (Tank 1, Manatee) and 100% (Tanks 4, 5, 6), mean 89.7%, for MarVer3 (see Appendix S5). Overall, amplicons for 9 and 22 out of 27 resident taxa were annotated to the species level using MarVer1 and MarVer3 respectively, (see Table 5, Figure 5, Appendix S5).

Amplicons for the aquarium's four frequently used vertebrate feed species (*Culpea harengus*, *Mallotous villosus*, *Merluccius productus*, and *Scomber scombrus*) were detected in Tanks 2 to 6 (in the Manatee tank, feed consists of lettuce) using both loci, with the exception of MarVer1 failing to detect *Merluccius productus*. Squid (unspecified species) is also supplied to the Dolphin Tank, but no cephalopod amplicons was recovered.

Amplicons for resident species were also detected in tanks other than their host tanks at low levels (e.g., with MarVer3, dolphin and seal amplicons were detected in the manatee and shark, and penguin and rocky shore tanks, respectively), suggesting possible transfer of eDNA between tanks in the aquarium, for example, via the equipment used by staff members. Similarly, human amplicons were detected in all tanks, consistent with the practice of aquarium staff entering the water for maintenance.

Both MarVer1 and MarVer3 identified amplicons (partially overlapping between loci and tanks) that were not directly attributable to resident species or food sources (category B in Appendix S5). These comprised six recurrent species, two of which were previously (but no longer) used as feed (*Sardina pilchardus*, *Engraulis encrasicolus*), and four species present in the Ligurian Sea (e.g., *Auxis rochei*, *Auxis thazard*, *Belone belone*, and *Coris julis*) from which the water used to fill the tanks is drawn, after being filtered and UV irradiated. All of these unexpected species detections were at low abundance, with read counts greater than 100 to a maximum of 947 in at least 1 tank (range 0.3% to 3.7% of total reads with MarVer3 and MarVer1, respectively). Very low abundance amplicons (<100 reads per tanks) for at least 20 other Mediterranean resident teleost fish species were also observed, but were not considered further as definitive detections.

Amplicons from invertebrate species (category C in Appendix S5) were detected in two tanks by the MarVer3 locus at low abundances (read count < 100), which would normally be discounted as a detection. These cases refer to an Anthozoa species, *Seriatopora hystrix*, in the manatee tank, and a Sipunculid worm, of the family Phascolosomatidae, in the seal tank (Appendix S5). Neither taxa were in the tank in which their traces were found. No invertebrate amplicons were recovered in the "rocky shore" tank which contains some Anthozoa (e.g., *Anemonia viridis*), some unidentified sponges growing spontaneously and hydrozoans, or from the shark tank where *Aiptasia* spp grows spontaneously. The other tanks (with the exception of dolphin and manatee) may also contain other spontaneously growing small invertebrates, such as copepods, amphipods, and hydrozoans, but none were detected.

4 | DISCUSSION

4.1 | Comparison with previously described barcode primer sets

This study was conceived to identify cetacean specific barcode loci complementing the many primer sets already available for fish species (e.g., Miya et al., 2015; Sato, Miya, Fukunaga, Sado, & Iwasaki, 2018), for use in eDNA biodiversity monitoring studies of Mediterranean marine vertebrates. However, in the primer design process, we realized that with minor adaptations, it was possible to cover the whole range of marine vertebrates in a single HTS run, potentially dramatically reducing costs for eDNA HTS biodiversity monitoring of pelagic vertebrates.

Most existing 12S/16S barcode primer sets (e.g., Bylemans et al., 2018; Miya et al., 2015; Taberlet et al., 2018) were designed for particular vertebrate groups and thus contained conserved sequence elements specific for their target taxonomic group. Most of these primer sets are partially overlapping, at least at one end, with the ones presented in this study although they are never identical (Figure 1). Our proposed primer sets were also found to be different from the universal 12S and 16S primers combinations described by Yang et al. (2014)—although MarVer3 sits within Yang et al.'s 16S target fragment—however, their amplicon sizes were too large (ca 430 bp and ca 500 bp for, respectively, the 12S and the 16S primer sets) to be easily employed in eDNA studies using current short-read HTS technology.

The 12S-V5 primer set (Riaz et al., 2011, renamed Vert01 by Taberlet et al., 2018) is the only one previously described as being specific for vertebrates. It is located adjacent to MarVer1 site (forward 12S-V5 primer partially overlaps with reverse MarVer1 sequence, see Figure 1). Within our alignments, the 12S-V5 site was not as variable as any of the three loci candidates identified in this paper.

TABLE 3 Levels of polymorphism found at the two proposed loci in a series of marine mammal Families, Genera and species for which complete mitogenomic data were available from GenBank (accession numbers listed in Appendix S4, with the exception of 22 *Pusa capsica* entries for which data are unpublished). Numbers in the two columns on the right indicate the number of variable sites detected in the sequence targeted by the two primers sets in each given comparison: Clear boxes indicate polymorphism thus usefulness of the primer set to resolve taxa case, while green boxes show lack of diagnostic polymorphism and light green boxes indicate lack of resolution at 98% homology threshold for MOTUs assignment

Taxonomic level	Group	Taxa (sample size)	Comparisons between	MV1	MV3
Family level	PINN	Phocidae (10 Genera)	*Cystophora*, *Erignathus*, *Halichoerus*, *Hydrurga*, *Lobodon*, *Mirounga*, *Monachus*, *Phoca*, *Pusa*, *Leptonychotes*	39	59
		Otariidae (6 Genera)	*Arctocephalus*, *Callorhinus*, *Eumetopias*, *Neophoca*, *Phocarctos*, *Zalophus*	31	42
	CMYS	Balaenopteridae (2 Genera)	*Balaenoptera*, *Megaptera*	23	25
	CODO	Ziphiidae (5 Genera)	*Ziphius*, *Mesoplodon*, *Indopacetus*, *Hyperodon*, *Berardius*	37	43
		Monodontidae (2 Genera)	*Delphinapterus*, *Monodon*	6	13
		Delphinidae (14 Genera)	*Cephalorhynchus*, *Sousa*, *Tursiops*, *Stenella*, *Delphinus*, *Lagenodelphis*, *Legenorhynchus*, *Grampus*, *Peponocephala*, *Feresa*, *Pseudorca*, *Orcinus*, *Globicephala*, *Orcaella*	41	42
		Phocoenidae (2 Genera)	*Neophocaena*, *Phocoena*	13	10
Genus level	PINN	*Phoca* (4 species, 5 individuals)	*Phoca vitulina* (1) versus *Phoca larga* (2)	0	2
			Phoca vitulina (1) versus *Phoca groenlandica* (1)	3	11
			Phoca vitulina (1) versus *Phoca fasciata* (1)	10	10
			Phoca larga (2) versus *Phoca groenlandica* (1)	3	11
			Phoca larga (2) versus *Phoca fasciata* (1)	11	10
			Phoca groenlandica (1) versus *Phoca fasciata* (1)	7	8
		Pusa (3 species, 3 individuals)	*Pusa caspica* (1) versus *Pusa hispida* (1)	1	4
			Pusa caspica (1) versus Pusa sibirica (1)	1	5
			Pusa hispida (1) versus *Pusa sibirica* (1)	0	1
		Arctocephalus (3 species, 48 individuals)	*Arctocephalus forsteri* (46) versus *Arctocephalus pusillus* (1)	8	4
			Arctocephalus forsteri (46) versus *Arctocephalus townsendi* (1)	4	5
			Arctocephalus pusillus (1) versus *Arctocephalus townsendi* (1)	8	4
	CODO	*Tursiops* (3 species, 23 individuals)	*Tursiops truncatus* (16) versus *Tursiops aduncus* (5)	4	1
			Tursiops truncatus (16) versus *Tursiops australis* (2)	3	1
			Tursiops aduncus (5) versus *Tursiops australis* (2)	3	2
		Delphinus (2 species, 2 individuals)	*Delphinus delphis* (1) versus *Delphinus capensis* (1)	1	2
		Stenella (3 species, 176 individuals)	*Stenella coeruleoalba* (2) versus *Stenella attenuata* (70)	3	3
			Stenella coeruleoalba (2) versus *Stenella longirostris* (104)	1	4
			Stenella attenuata (70) versus *Stenella longirostris* (104)	2	2
		Globipephala (2 species, 2 individuals)	*Globicephala melas* (1) versus *Globicephala macrorhynchus* (1)	1	2

(Continues)

TABLE 3 (Continued)

Taxonomic level	Group	Taxa (sample size)	Comparisons between	MV1	MV3
		Mesoplodon (5 species, 26 individuals)	*Mesoplodon europaeus* (8) versus *Mesoplodon densirostris* (12)	8	11
			Mesoplodon europaeus (8) versus *Mesoplodon grayi* (2)	10	13
			Mesoplodon europaeus (8) versus *Mesoplodon ginkgodens* (2)	7	11
			Mesoplodon europaeus (8) versus *Mesoplodon stejnegeri* (2)	12	12
			Mesoplodon densirostris (12) versus *Mesoplodon grayi* (2)	6	9
			Mesoplodon densirostris (12) versus *Mesoplodon ginkgodens* (2)	8	15
			Mesoplodon densirostris (12) versus *Mesoplodon stejnegeri* (2)	7	7
			Mesoplodon grayi (2) versus *Mesoplodon ginkgodens* (2)	10	12
			Mesoplodon grayi (2) versus *Mesoplodon stejnegeri* (2)	5	11
			Mesoplodon ginkgodens (2) versus *Mesoplodon stejnegeri* (2)	6	17
		Neophocena (2 species, 12 individuals)	*Neophocena asiaeorientalis* (4) versus *Neophocena phocaenoides* (8)	0	0
Species level	SIRE	*Dugong dugong*	2 individuals	1	2
	PINN	*Arctocephalus forsteri*	46 individuals	2	5[c]
		Eumetopias jubatus	11 individuals	0	0
		Pusa caspica	23 individuals	2[a]	2
	CMYS	*Balaenoptera physalus*	152 individuals	12[b]	8[b]
		Magaptera novaeangliae	2 individuals	1	1
	CODO	*Orcinus orca*	151 individuals	3	0
		Stenella longirostris	104 individuals	4[a]	6
		Stenella attenuata	70 individuals	3[b]	3
		Tursiops truncatus	16 individuals	1[a]	2[b]
		Ziphius cavirostris	20 individuals	1[a]	4
		Mesoplodon densirostris	12 individuals	1	4
		Mesoplodon europaeus	8 individuals	1	0
		Phocoena phocoena	5 individuals	1	0

[a]All mutations present in only one individual, probably not indicative of a populations/subspecies diagnostic site.

[b]Some of these are mutations present in only one individual.

[c]Further seven variable sites were found in a single individual (NZFS8, in Emami-Khoyi et al. 2016) and were not included in this list as they might be an artefact.

TABLE 4 List of tissue DNA extracts used for wet lab primer tests. "PC" indicates phenol–chloroform extracts, and "CK" indicates commercial kit extracts, while "BE" refers to boil extracts as described in Valsecchi (1998). The sign "✓" specifies successful amplification

Group	Species	Extraction type	Tissue	Time since ext (years)	MarVer1	MarVer3
CODO	*Stenella coeruleoalba*	Standard (PC)	Skin	27	✓	✓
	Stenella frontalis	Standard (PC)	Muscle	26	✓	✓
	Pontoporia blainvillei	Boil extraction	Skin	25	✓	✓
CMYS	*Megaptera novaeangliae*	Standard (PC)	Skin	31	✓	✓
	Eubalaena australis	Boil extraction	Skin	24	✓	✓
PINN	*Puca caspica*	Standard (CK)	Blood	1	✓	✓
STUR	*Lepidochelys kempii*	Standard (CK)	Blood	1	✓	✓
	Chelonia midas	Standard (CK)	Muscle	1	✓	✓
	Lepidochelys olivacea	Standard (CK)	Blood	1	✓	✓
	Caretta caretta	Standard (CK)	Blood	1	✓	✓
	Caretta caretta	Standard (CK)	Muscle	1	✓	✓
TELE	*Thunnus albacares*	Standard (CK)	Muscle	1	✓	✓
	Merlangius merlangus	Standard (CK)	Muscle	1	✓	✓

For example, looking at one of the most variable Families considered in this study, the Ziphiidae (Odontocetes), our two loci MarVer1 and MarVer3 recovered 37 and 43 variable sites, respectively (Table 3), while 12S-V5 highlighted only 13 within the same pool of sequences. This was not attributable to differences in amplicon size since the MarVer2 candidate region includes more than twice (30) the number of variable sites found in the 12S-V5 target sequence (13) within the Zophiidae family, while having a similar size (respectively, 96 bp and 106 bp). Such performance differences were also highlighted in the in silico simulation (Figure 4). Moreover, within both forward and reverse 12S-V5 primer sites, polymorphisms were present within the Ziphiidae family, suggesting that the 12S-V5 primers may not be conserved across all vertebrate classes. However, 12S-V5 performed slightly better than MarVer1 based on the total number of unique taxa for which DNA was amplified in the in silico simulation (Figure 2). While the lower predicted number of taxa recovered by these primers may indicate a failure to amplify DNA from some species, the completeness of the reference database used will also influence the results given that all sequence records were considered and not only the full mitochondrial genomes. Within the partial 12S sequences deposited on GenBank, the fragment including locus 12S-V5 might be over-represented, as most previous studies relied on this marker.

Besides being highly conserved among vertebrates, the two MarVer primer sets were shown to be potentially nonexclusive to vertebrates when 2–4 bp primer/template mismatches were allowed (Figure 3). With reduced specificity, these primer pairs could potentially amplify unwanted nonvertebrate taxa. Given the high number of vertebrate taxa recovered by the MarVer3 primers in silico (Figure 2), and the observation that the majority of the nonvertebrate taxa recovered have ≥3 bp mismatches in the primer binding regions (Figure 3), this primer pair should still be valuable if stringent thermal cycling conditions are used during PCR amplification. This was supported by the eDNA sequencing of aquarium samples where there was minimal recovery of invertebrate amplicons from tanks known to contain invertebrate species. This suggests the use of MarVer3 in Vertebrate biodiversity surveys would not be limited by potential homology with some invertebrate sequences.

4.2 | Performance of MarVer1 and MarVer3 primer sets with environmental samples

We evaluated the performance of MarVer1 and MarVer3 primer sets with water samples collected at the Genoa aquarium, from tanks with known community compositions. Amplicons annotated to species level were recovered for 81.5% and 37% of the 27 resident taxa for MarVer3 and MarVer1, respectively. For MarVer3, the five "undetected" species included *Diplodus cervinus* (Zebra seabream), *Pterygoplichthys gibbiceps* (Armored catfish), two ray species (*Dasiatis americana* and *Teaniura grabata*), and *Pristis zjisron* (Longcomb sawfish). In the case of *D. cervinus*, amplicons assigned to other nonresident *Diplodus* species were observed, so eDNA from the species may have been present, but annotated as a congeneric. For the four other "undetected" species, there were no other incompletely (above species level) assigned reads at genus, family, or order level which could be attributed to these taxa (for *Pristis zjisron*, no matching reference sequence for the MarVer3 region was available in GenBank). These four cases therefore appear to be genuine nondetections. These are all bottom-dwelling species, whereas our water samples were collected at the surface, and therefore, potentially we did not capture eDNA from these species.

For MarVer1, 9 out of 13 resident species with GenBank reference sequences covering the MarVer1 region were detected successfully. Amplicons correctly assigned to the species level were not observed for the two penguin species, Blackspot seabream (*Pagellus*

TABLE 5 Species composition of the six tanks of the Aquarium of Genoa from which water samples were collected for extracting eDNA. The last four columns on the right show NGS outcomes for primer sets MarVer1 and MarVer3: "GB" columns indicate whether reference sequences are present in GenBank for that specific species and locus; "■" indicate successful species detection (half-full and empty square indicates a number of reads 0.001> nr> 0.005 or nr < 0.001, respectively); △ denotes that a congenereric was identified in place of the resident vertebrate species; and "?" indicates possible detection of the resident species at a higher taxonomic group (this case together with those instances for which reference sequences were available but the species remained undetected are discussed in the text)

					MarVer1		MarVer3	
Tank	Name	Hosted species (*n.* individuals)	Common name	Vertebrate group	GB	Detection	GB	Detection
1	Manatee	*Trichechus manatus* (4)	Matatee	SIRE	Yes	■	Yes	■
		Piaractus brachypomus (3)	Pirapitinga	TELE	Yes	■	Yes	■
		Pterygoplichthys gibbiceps (2)	Armored catfish	TELE	No	X	Yes	X
2	Dolphin	*Tursiops truncatus (10)*	Bottlenose dolphin	CODO	Yes	■	Yes	■
3	Shark	*Ginglymostoma cirratum* (3)	Nurse shark	ELAS	Yes	■	Yes	■
		Stegostsoma fasciatum (1)	Zebra shark	ELAS	Yes	■	Yes	■
		Carcharhinus plumbeus (4)	Grey shark	ELAS	Yes	■	Yes	■
		Dasiatis americana (1)	Southern stingray	ELAS	No	X	No	X
		Pristis zjisron (2)	Longcomb sawfish	ELAS	Yes	X	Yes	X
		Teaniura grabata (2)	Round stingray	ELAS	No	X	Yes	X
		Epinepehelus itajara (1)	Atlantic goliath grouper	TELE	No	△	Yes	◩
		Epinephelus costae (1)	Goldblotch grouper	TELE	No	△	Yes	■
		Epinephelus marginatus (2)	Dusky grouper	TELE	No	△	Yes	■
		Seriola dumerili (9)	Greater amberjack	TELE	Yes	■	Yes	◩
		Dicentrarchus labrax (7)	European seabass	TELE	Yes	■	Yes	■
		Diplodus sargus (4)	White seabream	TELE	No	X	Yes	■
		Diplodus cervinus (1)	Zebra sea bream	TELE	No	X	Yes	△
		Pagellus bogaraveo (1)	Blackspot seabream	TELE	Yes	X	Yes	□
4	Penguin	*Pygoscelis papua* (12)	Gentoo penguin	SBIR	Yes	X	Yes	■
		Spheniscus magellanicus (24)	Magellanic penguin	SBIR	Yes	△	Yes	■
5	Seal	*Phoca vitulina* (6)	Harbour seal	PINN	Yes	■	Yes	■
6	Rocky Shore	*Diplodus vulgaris* (21)	Common two-banded seabream	TELE	No	X	Yes	■
		Diplodus sargus (3)	White seabream	TELE	No	X	Yes	■
		Sciaena umbra (3)	Brown meagre	TELE	No	?	Yes	■
		Sarpa salpa (2)	Salema	TELE	No	X	Yes	■
		Serranus scriba (1)	Painted comber	TELE	No	X	Yes	◩
		Scyliorhinus stellaris (2)	Nursehound	ELAS	No	X	Yes	■

(Continues)

TABLE 5 (Continued)

Tank	Name	Hosted species (*n*. individuals)	Common name	Vertebrate group	MarVer1		MarVer3	
					GB	Detection	GB	Detection
	Feed species							
2,3,4,5,6		*Culpea harengus*	Atlantic herring	TELE	Yes	■	Yes	■
2,3,4,5,6		*Mallotous villosus*	Capelin	TELE	Yes	■	Yes	■
3.6		*Merluccius productus*	North Pacific hake	TELE	No	X	Yes	■
6		*Scomber scombrus*	Atlantic mackerel	TELE	Yes	◪	Yes	■
2			Squid		Yes		Yes	

bogaraveo), and longcomb sawfish, despite reference sequences being available. For the Magellanic penguin, amplicons assigned to the congeneric *S. demersus* were observed, but no candidate incompletely or misattributed amplicons for the Gentoo penguin (*P. papua*) or two fish species were recorded.

For 14 species (51.6% cases), no GenBank reference sequences covering the MarVer1 region were available. Reads assigned to the nonresident grouper *Epinepehelus lanceolatus* were observed, indicating that eDNA from the three resident groupers may have been detected but attributed to a congeneric for which a reference was available. For the remaining 11 species, no other candidate amplicons attributable to related taxa were recorded.

Our demultiplexing and annotation pipeline required an amplicon sequence homology of at least 98% with other MOTUs and with GenBank reference sequences. Therefore, the lower assignment rate for MarVer1 compared to MarVer3 likely reflects the lower GenBank coverage for the 12S region encompassed by MarVer1. In this case, reducing stringency (e.g., to 95% homology) may increase annotation rates, allowing successful attribution of amplicons to genus and/or family level, but with a requirement to check homology level for individual MOTUs before accepting species-level assignments. For both primer sets, annotation success rates would be expected to increase as taxonomic coverage of GenBank reference sequences improves over time. Similarly, while MarVer3 was predicted to potentially recover invertebrate amplicons when allowing for low levels of degeneracy at priming sites, few were observed. Potentially, this may also be accounted for by low coverage with reference sequences and the level of stringency applied in the annotation pipeline. The annotation of the few observed invertebrate amplicons from the aquarium samples should also be treated cautiously for the same reasons.

4.3 | Optimizing locus choice for different eDNA and taxon detection applications

The two loci described in this paper provide investigators with flexible options to target different barcode markers depending on priorities for their study objectives, tailored to requirements for taxonomic breadth, variation and resolution at different taxonomic levels and amplicon size where eDNA degradation is a concern (e.g., Speller et al., 2016).

The 12S-based MarVer1 offers the advantage of smaller amplicon sizes (approximately 202 bp), which may be a consideration for applications requiring work with more degraded eDNA templates (Nichols et al., 2018; Wei, Nakajima, & Tobino, 2018).

The in silico analysis predicts that overall, the 16S-based MarVer3, with the largest amplicon product size of approximately 245 bp, has the highest taxonomic coverage across all vertebrates, and taxon resolution from species level upwards. Our trial eDNA HTS assay with aquarium samples demonstrated the locus performed well with environmental samples despite its larger amplicon. However, for marine mammals, there was variation across groups as to which

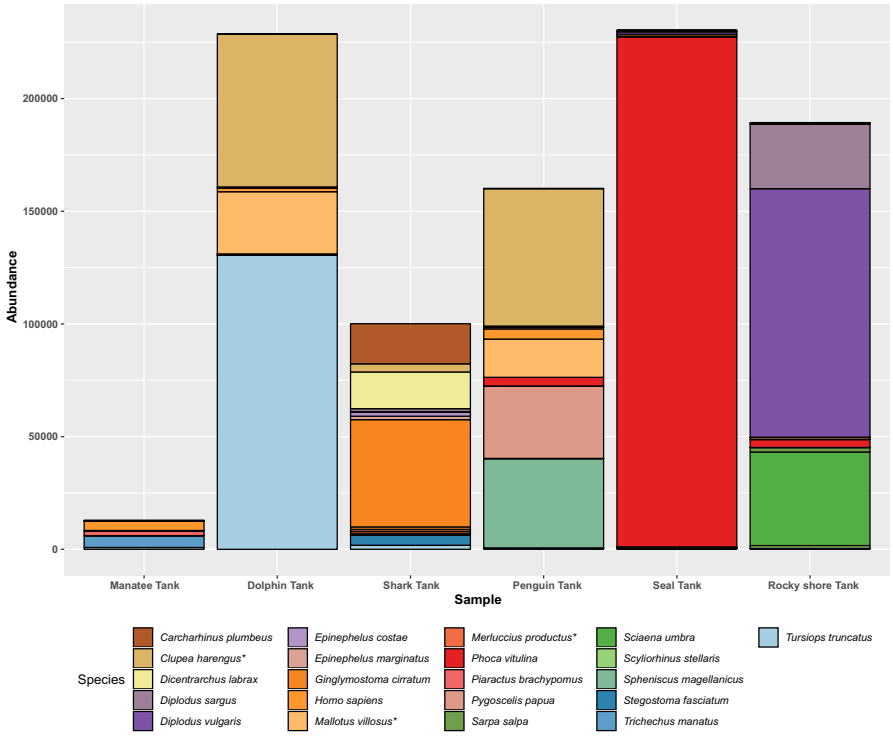

FIGURE 5 Taxon "abundance" (read counts) for amplicons generated with the MarVer3 primer set for environmental DNA extracted from water samples collected from 6 tanks at the Genoa Aquarium. Read counts are combined across the 6 replicate samples assayed for each tank, and sequence demultiplexing and amplicon annotation against Genbank references were carried out using a threshold of 98% sequence similarity. Taxa presented in the figure were filtered to exclude those with read counts less than 0.005*median read count across tanks. Taxa names with * are feed species, all others are resident

locus provided the best resolution. For example, in congeneric comparisons of *Arctocephalus* species MarVer1 performed better, while MarVer3 yielded the best resolution among *Mesoplodon* taxa.

At the intraspecies level for marine mammals, MarVer1 amplicons were typically more variable than MarVer3, which showed varying levels of polymorphism, ranging from no variability among 151 killer whale (*Orcinus orca*) samples to high levels of diversity in 104 spinner dolphins (*Stenella longirostris*), Table 3. Overall, this pattern is consistent with lower rates of evolution in 16S compared to 12S genes and suggests 12S genes may be more informative for resolving intraspecific differences (see below).

We provide guidelines in Table 3 for the choice of the most suitable marker to be employed where specific marine mammal taxa are of interest, while for metabarcoding studies aiming at the detection of all marine vertebrates, we would recommend using the combination of the two loci to maximize taxonomic coverage and to ameliorate potential gaps in reference sequence databases.

Our initial search for hypervariable regions flanked by conserved sequences highlighted another candidate within the 12S gene (located between MarVer1 and MarVer3 loci, thus named MarVer2; Appendix S6). The candidate was not evaluated in HTS screening as, on the basis of its restricted size (98 bp), it was not predicted to provide full taxonomic resolution among Mediterranean delphinids, which was a primary aim of our original study design. Primer details and an initial in silico evaluation of its performance however are presented in Appendix S6, since the in silico analysis indicates it has good potential for some eDNA applications such as the development of species presence/absence assays (data not shown) for a wide range of vertebrates pending appropriate validation. This locus was also the only one among the three MarVer sites to present a suitable string of nucleotides adjacent to the priming site presenting cetacean specific base combinations. Thus, a further primer set, named "Ceto2" (Appendix S6), was designed to preferentially amplify Cetacean DNA by minimizing base-pair mismatches for cetacean species while maximizing base-pair mismatches for other vertebrate groups, including humans, thus reducing contamination risk.

Resolution of intraspecific phylogeographic variation is likely to be best attempted with either more variable or longer target amplicons (e.g., d-loop region; Kunal, Kumar, Menezes, & Meena, 2013). However, the large Cetacean sequence dataset we evaluated allowed us to test the potential of our loci to identify phylogeographically informative variation, which could be used for simple haplotype clade determination with eDNA for some species (Adams et al., 2019). For instance, within the MarVer1 amplicon in the 151 killer whale mitogenomes (Morin et al., 2010, Morin et al., 2015 and Filatova et al., 2018), some variable sites were private either to the Transient clade or to the AntB and AntC clades identified by the larger dataset.

The preliminary investigation of sequence variation in other large marine vertebrate groups (tuna and sea turtle species) suggested our loci also have potential to be informative for species identification in those taxa. While not assessed directly in this study, the MarVer loci may also prove to be useful as barcode markers for terrestrial vertebrates given taxonomic conservation of the priming sites.

Finally, the high levels of diagnostic variation seen within MarVer loci amplicons offer potential for designing additional species-specific nested internal primers (Stoeckle, Das, & Charlop-Powers, 2018). These might have utility for species-focused, non-sequencing-based

detection applications, such as quantitative PCR (qPCR) or digital droplet PCR (ddPCR), or simple agarose gel-based amplicon visualization when there is limited access to laboratory facilities or funding.

5 | CONCLUSIONS

This paper presents four novel primer sets targeting 12S and 16S vertebrate mitogenome regions, with a particular focus on marine mammals. Using a combination of "in silico" validation, and application to eDNA samples from aquarium communities with known species composition, we show the loci to have high potential for metabarcoding and eDNA studies targeting marine vertebrates. These primer sets have broader taxonomic coverage and resolution compared to previously developed 12S and 16S primer sets, potentially allowing surveys of complete marine vertebrate communities (including fish, sea turtles, bird, and mammals) in single HTS runs, simplifying workflows, reducing costs, and increasing accessibility to a wider range of investigators. They may be applied in any context focusing on resolving vertebrate taxonomic identity, from biodiversity surveys and forensics (e.g., CITES surveillance or surveys of commercially targeted fish species), through to behavioral ecology studies and supporting conservation of rare or endangered marine vertebrate species.

ACKNOWLEDGMENTS

We thank Fulvio Maffucci for providing marine turtles' DNA samples, Giudo Gnone of the Aquarium of Genoa for allowing and supporting collection of controlled environmental eDNA samples, Antonia Bruno for advises on filtering procedures of environmental samples, and Anna Sandionigi for bioinformatic advice.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

AUTHOR CONTRIBUTIONS

E.V. designed primer sets, planned the testing approach, compiled the marine mammal guideline to the primers' use, and wrote the manuscript; J.B. performed the in silico validation of the novel primer sets; R.L. carried out samples collection, wet lab tests, and eDNA amplifications and analyzed controlled environment HTS data; S.G. contributed to the design and implementation of the wet lab primer validation, provided support and facilities for HTS analysis at UoL, and contributed to the drafting and editing of the manuscript; I.C. provided HTS services and bioinformatics support; L.C. allowed and supported collection of water samples from tanks of the Aquarium of Genoa structure; A.G. provided useful comments on manuscript; P.G. enthusiastically supported and hosted the research study at UnMB.

ETHICAL APPROVAL

All applicable international, national, and/or institutional guidelines for the care and use of animals were followed by the authors.

SAMPLING AND FIELD STUDIES

All necessary permits for sampling and observational field studies have been obtained by the authors from the competent authorities and are mentioned in the acknowledgements, if applicable.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are included in this published article (and its Supporting Information files) and are available from the corresponding author on reasonable request.

ORCID

Elena Valsecchi https://orcid.org/0000-0003-3869-6413
Jonas Bylemans https://orcid.org/0000-0001-6263-0874
Simon J. Goodman https://orcid.org/0000-0003-4118-8575
Ian Carr https://orcid.org/0000-0001-9544-1068
Andrea Galimberti https://orcid.org/0000-0003-3140-3024
Paolo Galli https://orcid.org/0000-0002-6065-8192

REFERENCES

Adams, C. I. M., Knapp, M., Gemmell, N. J., Jeunen, G.-J., Bunce, M., Lamare, M. D., & Taylor, H. R. (2019). Beyond biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes (Basel)*, *10*, 192. https://doi.org/10.3390/genes10030192

Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*, *12*, 1–20. https://doi.org/10.1371/journal.pone.0176343

Auguie, B. (2016). gridExtra: Miscellaneous functions for "grid" graphics. R package version, 2(1), 242.

Baker, C. S., Steel, D., Nieukirk, S., & Klinck, H. (2018). Environmental DNA (eDNA) from the wake of the whales: Droplet digital PCR for detection and species identification. *Frontiers in Marine Science*, *5*, 1–11. https://doi.org/10.3389/fmars.2018.00133

Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle - Rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS ONE*, *12*, 1–17. https://doi.org/10.1371/journal.pone.0183347

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*, 176–182. https://doi.org/10.1111/1755-0998.12428

Bylemans, J., Gleeson, D. M., Hardy, C. M., & Furlan, E. (2018). Toward an ecoregion scale evaluation of eDNA metabarcoding primers: A case study for the freshwater fish biodiversity of the Murray-Darling Basin (Australia). *Ecology and Evolution*, *8*, 8697–8712. https://doi.org/10.1002/ece3.4387

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., … Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, *7*, 1008–1018. https://doi.org/10.1111/2041-210X.12574

Deagle, B. E., Chiaradia, A., McInnes, J., & Jarman, S. N. (2010). Pyrosequencing faecal DNA to determine diet of little penguins: Is what goes in what comes out? *Conservation Genetics*, *11*, 2039–2048. https://doi.org/10.1007/s10592-010-0096-6

Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, *10*, 2–5. https://doi.org/10.1098/rsbl.2014.0562

Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, *8*, 1888–1898. https://doi.org/10.1111/2041-210X.12836

Elbrecht, V., Hebert, P. D. N., & Steinke, D. (2018). Slippage of degenerate primers can cause variation in amplicon length. *Scientific Reports*, *8*, 1–5. https://doi.org/10.1038/s41598-018-29364-z

Emami-Khoyi, A., Hartley, D. A., Paterson, A. M., Boren, L. J., Cruickshank, R. H., Ross, J. G., … Else, T. A. (2016). Identifying prey items from New Zealand fur seal (Arctocephalus forsteri) faeces using massive parallel sequencing. *Conservation Genetics Resources*, *8*(3), 343–352. https://doi.org/10.1007/s12686-016-0560-9

Ficetola, G., Coissac, E., Zundel, S., Riaz, T., Shehzad, W., Bessière, J., … Pompanon, F. (2010). An in silico approach for the evaluation of DNA barcodes. *BMC Genomics*, *11*, 434. https://doi.org/10.1186/1471-2164-11-434

Filatova, O. A., Borisova, E. A., Meschersky, I. G., Logacheva, M. D., Kuzkina, N. V., Shpak, O. V., … Hoyt, E. (2018). Colonizing the wild west: Low diversity of complete mitochondrial genomes in Western North Pacific killer whales suggests a founder effect. *Journal of Heredity*, *109*, 735–743. https://doi.org/10.1093/jhered/esy037

Foote, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L. A., … Gilbert, M. T. P. (2012). Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE*, *7*, 2–7. https://doi.org/10.1371/journal.pone.0041781

Giguet-Covex, C., Pansu, J., Arnaud, F., Rey, P. J., Griggo, C., Gielly, L., … Poulenard, J. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature communications*, *5*(1), 1–7.

Ingman, M., Kaessmann, H., Pääbo, S., & Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modem humans. *Nature*, *408*, 708–713. https://doi.org/10.1038/35047064

ITIS (2018). The Integrated Taxonomic Information System on-line database. Retrieved from http://www.itis.gov. Data retrieved 2018-12-18.

Jarman, S. N., Berry, O., & Bunce, M. (2018). The value of environmental DNA biobanking for long-term biomonitoring. *Nature Ecology & Evolution*, *2*, 1192–1193. https://doi.org/10.1038/s41559-018-0614-3

Jarman, S. N., Redd, K. S., & Gales, N. J. (2006). Group-specific primers for amplifying DNA sequences that identify Amphipoda, Cephalopoda, Echinodermata, Gastropoda, Isopoda, Ostracoda and Thoracica. *Molecular Ecology Notes*, *6*, 268–271. https://doi.org/10.1111/j.1471-8286.2005.01172.x

Komai, T., Gotoh, R. O., Sado, T., & Miya, M. (2019). Development of a new set of PCR primers for eDNA metabarcoding decapod crustaceans. *Metabarcoding and Metagenomics*, *3*, 1–19. https://doi.org/10.3897/mbmg.3.33835

Kunal, S. P., Kumar, G., Menezes, M. R., & Meena, R. M. (2013). Mitochondrial DNA analysis reveals three stocks of yellowfin tuna *Thunnus albacares* (Bonnaterre, 1788) in Indian waters. *Conservation Genetics*, *14*, 205–213. https://doi.org/10.1007/s10592-013-0445-3

McInnes, J. C., Alderman, R., Deagle, B. E., Lea, M.-A., Raymond, B., & Jarman, S. N. (2017). Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. *Methods in Ecology and Evolution*, *8*, 192–202. https://doi.org/10.1111/2041-210X.12677

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, *8*, e61217. https://doi.org/10.1371/journal.pone.0061217

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., … Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. *Royal Society Open Science*, *2*, 150088. https://doi.org/10.1098/rsos.150088

Morin, P. A., Archer, F. I., Foote, A. D., Vilstrup, J., Allen, E. E., Wade, P., … Harkins, T. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research*, *20*, 908–916, https://doi.org/10.1101/gr.102954.109.908

Morin, P. A., Parsons, K. M., Archer, F. I., Ávila-Arcos, M. C., Barrett-Lennard, L. G., Dalla Rosa, L., … Foote, A. D. (2015). Geographic and temporal dynamics of a global radiation and diversification in the killer whale. *Molecular Ecology*, *24*, 3964–3979. https://doi.org/10.1111/mec.13284

Nichols, R. V., Vollmers, C., Newsom, L. A., Wang, Y., Heintzman, P. D., Leighton, M. K., … Shapiro, B. (2018). Minimizing polymerase biases in metabarcoding. *Molecular Ecology Resources*, *18*, 927–939. https://doi.org/10.1111/1755-0998.12895

Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] - [data retrieved 2018–12-18]. Available from: https://www.ncbi.nlm.nih.gov/nucleotide/

Peters, K. J., Ophelkeller, K., Bott, N. J., Deagle, B. E., Jarman, S. N., & Goldsworthy, S. D. (2015). Fine-scale diet of the Australian sea lion (*Neophoca cinerea*) using DNA-based analysis of faeces. *Marine Ecology*, *36*, 347–367. https://doi.org/10.1111/maec.12145

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., … Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*, 527–541. https://doi.org/10.1111/mec.13481

R Development Core Team (2010). R: A language and environment for statistical computing. Computer programme, Retrieved from http://www.R-project.org/

Riaz, T., Shehzad, W., Viari, A., Pompanon, F., Taberlet, P., & Coissac, E. (2011). EcoPrimers: Inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research*, *39*, 1–11. https://doi.org/10.1093/nar/gkr732

Sato, Y., Miya, M., Fukunaga, T., Sado, T., & Iwasaki, W. (2018). MitoFish and mifish pipeline: A mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. *Molecular Biology and Evolution*, *35*, 1553–1555. https://doi.org/10.1093/molbev/msy074

Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., … Thomsen, P. F. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, *1*, 4. https://doi.org/10.1038/s41559-016-0004

Sigsgaard, E. E., Nielsen, I. B., Carl, H., Krag, M. A., Knudsen, S. W., Xing, Y., … Thomsen, P. F. (2017). Seawater environmental DNA reflects seasonality of a coastal fish community. *Marine Biology*, *164*, 128. https://doi.org/10.1007/s00227-017-3147-4

Speller, C., van den Hurk, Y., Charpentier, A., Rodrigues, A., Gardeisen, A., Wilkens, B., … Hofreiter, M. (2016). Barcoding the largest animals on earth: Ongoing challenges and molecular solutions in the taxonomic identification of ancient cetaceans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*, https://doi.org/10.1098/rstb.2015.0332

Stoeckle, M. Y., Das, M. M., & Charlop-Powers, Z. (2018). Gofish: A versatile nested PCR strategy for environmental DNA assays for marine vertebrates. *PLoS ONE*, *13*, 1–17. https://doi.org/10.1371/journal.pone.0198717

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). *Environmental DNA for biodiversity research and monitoring*. Oxford, UK: Oxford University Press.

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a Diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, *7*, 1–9. https://doi.org/10.1371/journal.pone.0041732

Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., … Dejean, T. (2016). Next-generation monitoring of aquatic

biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, *25*, 929–942. https://doi.org/10.1111/mec.13428

Valsecchi, E. (1998). Tissue boiling: A short-cut in DNA extraction for large-scale population screenings. *Molecular Ecology*, *7*, 1243–1245. https://doi.org/10.1046/j.1365-294x.1998.00379.x

Wei, N., Nakajima, F., & Tobino, T. (2018). Effects of treated sample weight and DNA marker length on sediment eDNA based detection of a benthic invertebrate. *Ecological Indicators*, *93*, 267–273. https://doi.org/10.1016/j.ecolind.2018.04.063

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. useR. Springer.

Yang, L. I., Tan, Z., Wang, D., Xue, L., Guan, M.-X., Huang, T., & Li, R. (2014). Species identification through mitochondrial rRNA genetic analysis. *Scientific Reports*, *4*, 1–11. https://doi.org/10.1038/srep04089

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Valsecchi E, Bylemans J, Goodman SJ, et al. Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. *Environmental DNA*. 2020;2:460–476. https://doi.org/10.1002/edn3.72