Research

Author for correspondence:
Kim M. Parsons
e-mail: kim.parsons@noaa.gov

Water, water everywhere: environmental DNA can unlock population structure in elusive marine species

Kim M. Parsons[1], Meredith Everett[3], Marilyn Dahlheim[2] and Linda Park[4]

[1]Under Contract to NOAA, National Marine Fisheries Service, Marine Mammal Laboratory, and [2]Marine Mammal Laboratory, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA
[3]Under Contract to NOAA, National Marine Fisheries Service, and [4]Northwest Fisheries Science Center, 2725 Montlake Boulevard E, Seattle, WA 98112, USA

KMP, 0000-0003-4580-3651

Determining management units for natural populations is critical for effective conservation and management. However, collecting the requisite tissue samples for population genetic analyses remains the primary limiting factor for a number of marine species. The harbour porpoise (*Phocoena phocoena*), one of the smallest cetaceans in the Northern Hemisphere, is a primary example. These elusive, highly mobile small animals confound traditional approaches of collecting tissue samples for genetic analyses, yet their nearshore habitat makes them highly vulnerable to fisheries by-catch and the effects of habitat degradation. By exploiting the naturally shed cellular material in seawater and the power of next-generation sequencing, we develop a novel approach for generating population-specific mitochondrial sequence data from environmental DNA (eDNA) using surface seawater samples. Indications of significant genetic differentiation within a currently recognized management stock highlights the need for dedicated eDNA sampling throughout the population's range in southeast Alaska. This indirect sampling tactic for characterizing stock structure of small and endangered marine mammals has the potential to revolutionize population assessment for otherwise inaccessible marine taxa.

1. Introduction

As apex predators, marine mammals play a critical role as indicators of the health of the marine ecosystem [1,2]. This is particularly true for cetaceans (whales, dolphins and

THE ROYAL SOCIETY PUBLISHING

porpoises) that inhabit coastal waters and are susceptible to the direct impacts of numerous anthropogenic activities. Population genetics provides a means of generating key parameters for characterizing populations and facilitating conservation and management objectives by enabling the long-term monitoring of local populations. For species suffering mortalities from anthropogenic sources, determination of population structure (i.e. the size and range of the relevant management unit) is critical to understanding conservation status and developing appropriate mitigation strategies.

Accurate characterization and quantification of the genetic diversity within and among populations hinges upon the collection of an adequate number of samples representing the full genetic diversity across the geographical range of a population. While remote collection of tissue samples using biopsy darts has become a routine and highly successful method for many large cetacean species, efforts to collect significant numbers of genetic samples from some of the smaller, more enigmatic dolphin and porpoise species are still often restricted to beach-cast carcasses and fisheries by-catch. Such opportunistic samples provide insight into population structure and geographical breaks in genetic diversity, but lack of certainty about sample provenance and the infrequency of sample discovery present challenges for addressing questions pertaining to human-mediated environmental impacts on relatively local scales.

With the sole exception of the vaquita (*Phocoena sinus*), the harbour porpoise (*Phocoena phocoena*) is the smallest cetacean in the North Pacific. Harbour porpoise are broadly distributed through cold and temperate coastal waters along the continental shelf. The coastal habitat and prey preferences of this small, elusive marine mammal contribute to their vulnerability to anthropogenic impacts. In particular, their shallow coastal distribution makes them highly vulnerable to incidental capture during net-fishing operations [3–5]. The nature and magnitude of incidental takes are currently unknown but may be significant in some fisheries for Pacific salmon (*Oncorhynchus* spp.) and Pacific herring (*Clupea pallasi*) [6]. Line-transect abundance estimates from the inland waters of southeast Alaska indicated an overall decline in porpoise abundance in the mid-2000s. Contrasting trends between the northern and southern regions of this study area suggest significant population structuring within the currently recognized southeast Alaska stock [6]. Concern for localized impact on undefined harbour porpoise stocks underscores the need for population genetic analyses to better define population structure; however, sample sizes generated by both strandings and fisheries by-catch are severely limited in key geographical areas, and efforts to supplement these sources with remotely collected tissue biopsies have proved extraordinarily challenging for these small, elusive cetaceans. Because development of effective conservation and management strategies of living resources relies on our understanding of population dynamics and an accurate definition of genetic populations or stocks, generating the requisite genetic data despite obstacles to direct collection of tissue samples is a high priority [7].

To overcome these sampling challenges, we explored the use of seawater environmental DNA (eDNA) to obtain population-level genetic data. Exploiting the cellular material that is continually shed into aquatic environments by inhabiting organisms is increasingly being used to detect rare and invasive organisms [8,9] and to survey marine biodiversity [10–12]. Technological advances have led to rapid growth in this field and eDNA has been used in at least one case to generate population-level estimates of genetic diversity [13]. Using targeted seawater sampling, and taxon-specific genetic markers, we demonstrate the use of eDNA as a novel approach for characterizing the population genetic structure of harbour porpoise. Targeted amplicon sequencing and quantitative real-time polymerase chain reaction (PCR) provide insight into relative quantities of porpoise DNA in surface seawater samples, and generate mitochondrial sequence data that can be incorporated into a traditional framework for examining genetic diversity among harbour porpoise in the inland waters of southeast Alaska. This novel approach has the potential to unlock population structure in this and other elusive marine taxa where adequate sample sizes from direct sampling of individuals is limited.

2. Methods

2.1. Sample collection

Seawater samples for eDNA were collected in regions of historically high porpoise density [6] throughout the inland waters of southeast Alaska during two, two-week ship-based surveys in June and September 2016. When porpoise were sighted, an 8 m Zodiac was launched to enable close approaches and sample collection. Samples were collected from surface waters in the fluke prints within the wake of diving harbour porpoise using handheld, sterile 96 oz (approx. 3 l) wide-mouth collapsible Nalgene©

polyethylene canteens. Before the start of the field season, all collection canteens were sterilized with a 10% bleach solution and triple rinsed in distilled water to ensure they were free of any exogenous DNA. Samples were kept cool and dark until further processing in an enclosed bow box on the skiff. During sampling, the air and sea-surface temperatures (approx. 10°C) in Alaska were sufficiently low such that external cooling of the collected samples with ice packs was deemed unnecessary. Negative eDNA controls consisted of seawater samples collected in regions where no porpoise were observed.

Seawater eDNA samples were vacuum-filtered within 10 h of collection (mean = 3.42 h). The entire sample volume (2.3–3.1 l) was vacuum-filtered through a sterile, 0.45 μM nitrocellulose filter using a single-use filtration cup (Sterlitech, Auburn, WA) attached to a PVC vacuum filtration manifold. Post-filtration, filters were stored in 5 ml of Longmire's lysis buffer [14] at ambient temperature for storage and transport. The shipboard filtration area, as well as all equipment, including the filter manifold, forceps and gloves were treated with DNA AWAY™ (Molecular BioProducts, San Diego, CA) before the start of filtration to prevent contamination between samples or from other research operations taking place in the laboratory. All collection canteens were sterilized before and after every use, inside and out, by soaking for at least 10 min in a 10% bleach solution, followed by triple rinses with fresh tap water, and air drying.

2.2. DNA extraction

DNA extractions were carried out in a sterile biosafety cabinet in a laboratory dedicated to pre-PCR procedures. Rigorous controls for preventing and monitoring cross-sample contamination and contamination from exogenous DNA were employed throughout sample processing. Prior to each extraction set, the ventilation hood and all equipment were cleaned with DNA AWAY™ and irradiated with UV light for 30 min before and after each extraction. No harbour porpoise tissue extractions were performed concurrently during the period of eDNA extractions. Genomic DNA was extracted from each sample using a standard phenol–chloroform protocol [15]. Filters in Longmire's lysis buffer were incubated at 65°C for one hour on a shaking platform. eDNA from each sample was extracted in two paired extractions, one using 1 ml of Longmire's buffer, and the other using 1 ml of Longmire's buffer plus the eDNA filter. The extraction containing the filter was placed in a Qiagen Tissuelyser II (Qiagen, Germantown, MD) and shaken at an oscillation frequency of 25 Hz for 2 min 20 s with a single sterile, stainless steel 0.5 mm bead. After disruption, 1 ml of Longmire's lysis buffer was pipetted off and carried forward into the phenol–chloroform extractions, and the eDNA filter was frozen for archival purposes. Two chloroform extractions were performed as detailed in Sambrook & Russell [16] and Renshaw et al. [17]. After the second wash, 700 μl of the resulting aqueous layer was transferred to a clean 1.5 ml tube, 30 μl of 5 M NaCl and 700 μl of isopropanol were added to each sample and DNA was allowed to precipitate at −20°C overnight. After precipitation, samples were centrifuged for 20 min at 13 000 r.p.m. in a microcentrifuge to pellet DNA, ethanol removed and DNA resuspended in 100 μl of sterile, laboratory-grade water (Sigma-Aldrich, St Louis, MO). The two isolates from each sample were pooled and re-precipitated, and concentrated by adding 20 μl of 3 M sodium acetate and 400 μl of cold ethanol to each sample, inverting and precipitating overnight at −20°C to remove co-precipitated PCR inhibitors and residual salts. Samples were centrifuged for 20 min at 13 000 r.p.m. in a microcentrifuge, ethanol removed and DNA eluted in a final volume of 50 μl of sterile laboratory-grade water.

2.3. Harbour porpoise eDNA quantitative polymerase chain reaction

Quantitative PCR (qPCR) was used to detect and quantify harbour porpoise DNA in eDNA samples. Cytochrome b gene sequences for cetacean species known to co-occur with harbour porpoise in the inland waters of southeast Alaska were obtained from GenBank and aligned to identify species-specific regions for qPCR probe and primer design. A TaqMan© probe specific to harbour porpoise (Ppho_Cytb-PROBE CTATGACTATTAGTAGTAAGAGCACCC) was designed for the target region and modified with 6-FAM reporter dye at the 5′ end and a non-fluorescent quencher at the 3′ end. Primers Ppho_Cytb-F (GTTCTTCATTTGTCTTTATATCCATATTG) and Ppho_Cytb-R (GCACCTCAAAATGATATTTGTCCT) were used to amplify a 160 bp region of the cytochrome oxidase b gene. Confirmation of species specificity was tested against the two most common odontocetes known to co-occur in the survey region, killer whale (*Orcinus orca*) and Dall's porpoise (*Phocoenoides dalli*), using DNA extracted from vouchered tissue samples (DNA provided by P. Morin, Southwest Fisheries Science Center, NOAA, National Marine Fisheries Service).

Table 1. Harbour porpoise eDNA mitochondrial primers. Sequence in bold is the Illumina sequencing primer and adapter sequence.

primer name	sequence (5′-3′)
Ppho_Con1F	TAC TCC TTG AAA AAG CCC ATT GTA
Ppho_Con7R	ATG GTC CTG AAG TAA GAA CCA GAT G
Ppho_Con1F-Illumina	**TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG** TAC TCC TTG AAA AAG CCC ATT GTA
Ppho_Con7R-Illumina	**GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G**AT GGT CCT GAA GTA AGA ACC AGA TG

Real-time PCRs were performed on an Applied Biosystems 7900 in 10 µl reaction volumes containing 0.3 µM probe, 0.8 µM each primer, 1× TaqMan Environmental Master Mix 2.0 (Applied Biosystems) and 3 µl template eDNA. Following Applied Biosystems TaqMan protocol, thermal cycling conditions were set at 95°C for 10 min, followed by 50 cycles of 95°C for 15 s and 60°C for 1 min. Each set of reactions contained a minimum of two no-template controls to detect contamination with exogenous DNA. Relative standard curves (RSCs) were developed from 1 : 10 serial dilutions ($1.3 \, \text{ng} \, \mu\text{l}^{-1}$ to $1 \times 10^{-5} \, \text{ng} \, \mu\text{l}^{-1}$) of quantified target-species DNA extracted from a vouchered harbour porpoise tissue sample quantified on a Qubit™ (Invitrogen). The RSC slope and y-intercept were used to calculate the DNA concentration for the measured threshold cycle (C_t) values.

2.4. Harbour porpoise eDNA primer development

Primer design for next-generation sequencing focused on the 5′ region of the mitochondrial control region. This section of the mitogenome is known to be hypervariable from ongoing genetic studies of North Pacific harbour porpoise [18,19] and captures phylogenetically informative sites. Primer design was based on a ClustalW alignment of 58 known North Pacific harbour porpoise haplotypes from sequence data generated from Sanger sequencing, in collaboration with the Southwest Fisheries Science Center, and from published sequences [20]. Primers were designed using NCBI's Primer-BLAST application [21] to capture the species-specific target region (approx. 450 bp). Each primer was tested *in silico* using the GenBank database (adjusting parameters for short input sequences), and *in vitro* with DNA extracted from vouchered tissues to confirm species specificity and ensure that the primers did not amplify the two commonly co-occurring, non-target odontocetes: killer whales and Dall's porpoise. PCR conditions were optimized to amplify the target fragment reliably in low copy number eDNA samples. PCR reactions were carried out in 10 µl volumes containing 2 µl of template eDNA, 1.2 mM MgCl$_2$, 0.2 mM dNTP, 1.0 µM each forward and reverse primer (table 1), $2 \, \mu\text{g} \, \mu\text{l}^{-1}$ bovine serum albumin and 0.6 U Promega GoTaq (Promega, Madison, WI). PCR cycling conditions were as follows: 94°C for 2 min, followed by 40 cycles of 94°C for 1 min, 49°C for 1 min, 72°C for 1 min and a final extension of 5 min at 72°C.

2.5. Next-generation eDNA amplicon sequencing

Amplifications for Illumina sequencing were carried out in triplicate on all eDNA and positive control samples. Positive controls were generated using genomic DNA from four Sanger sequenced, vouchered harbour porpoise samples to test for PCR or sequencing error. DNA from each control sample was quantified using a Qubit™ 3.0 (ThermoFisher, Waltham, MA) and the Qubit DS Broad Range Assay (ThermoFisher, Waltham, MA) according to the manufacturer's instructions, and samples were normalized to a concentration of $5 \, \text{ng} \, \mu\text{l}^{-1}$. Replicate 10 µl reactions were performed for all eDNA and control samples, using optimized conditions (as above) and modified primers containing Illumina sequencing primer and adapter sequences. Each sample was amplified in three independent 10 µl reactions to ensure sufficient PCR product for next-generation sequencing. Products from all three PCRs were pooled and the presence of the desired amplicon size was confirmed by electrophoresis of the pooled 30 µl in a 2% agarose gel stained with SyberSafe (ThermoFisher, Waltham, MA). Amplicon size was determined with reference to a 100 bp DNA ladder (New England Biolabs, Ipswich, MA). Two negative control reactions were run in each set of PCR reactions, including template from one extraction blank and a no-template PCR sample. Bands from successful PCR of the eDNA samples and positive controls were gel excised using clean techniques to prevent cross-contamination between amplicons and purified using the Qiagen MinElute Gel Extraction spin-column protocol with a 30 µl elution volume

(Qiagen, Germantown, MD). Gel electrophoresis of both negative controls confirmed the absence of detectable PCR product and these samples were not used for library generation.

Sequencing capture adapters with unique pairs of Illumina Nextera XT index tags were added to each sample using a second reduced-cycle PCR protocol according to the Illumina 16S metagenomic sequencing library preparation protocol as follows. This second PCR step was performed in 50 µl reaction volumes containing 8 µl of gel-purified PCR product from the first PCR, 5 µl each of one Illumina Nextera forward and reverse index tag, and 1× NEB Phusion High-Fidelity PCR Master Mix (New England Biolabs, Ipswich, MA). PCR cycling conditions were as follows: 95°C for 3 min, followed by 8 cycles of 95°C for 30 s, 55°C for 30 s, 72°C for 30 s, and a final extension of 5 min at 72°C. Indexed amplicons were normalized across samples using the SequalPrep Normalization Plate Kit (Applied Biosystems) and eluted in a 20 µl volume. Following normalization, indexed amplicons were pooled, cleaned using Ampure XP beads (Beckman Coulter, Brea, CA) and eluted in 25 µl of 10 mM Tris (pH 8.5). The indexed and pooled amplicon library was quantified via qPCR using the KAPA Library Quantification Kit for Illumina libraries (KAPA Biosystems, Wilmington, MA) according to the manufacturer's protocols and diluted to 4 nM. The final library was diluted to 12 pM and sequenced on an Illumina MiSeq Platform using the V3-600 paired-end sequencing chemistry (Illumina, San Diego, CA) with a PhiX spike-in of 10%.

2.6. Next-generation sequencing data analysis

The data analysis pipeline described below was first performed on sequence data for the four positive controls to calculate the expected rate of error generated during PCR and sequencing. Using data from each of the controls, we identified the most abundant sequence that had less than 100% identity to the known control region haplotype for that sample, and compared the read count to the total read count for the sample to derive an overall error rate due to non-specific priming and sequencing error [22]. The observed error rate was used to set a filter threshold for eDNA libraries; haplotypes appearing at a frequency below this threshold were removed from the data as described below.

Cutadapt (v. 1.13) [23] was used to trim Illumina adapters and primers from each sequence and quality filter each paired-end datafile separately using a threshold of -q 20. Read pairs were merged with VSEARCH (v. 2.4.3) [24] *fastq_mergepairs*, with a minimum overlap of 50 bp. VSEARCH was used to cluster identical sequences (or de-replicate), and restrict sequences to a minimum aligned length of 379 bp. Chimeric sequences were identified and discarded using the VSEARCH command *uchime denovo*, and sequences with a frequency of occurrence less than the empirically determined error rate (above) were removed from the analysis. Remaining sequences were compared to a local database of 58 known harbour porpoise control region haplotypes from the eastern North Pacific using the pairwise BLAST (Basic Local Alignment Search Tool) algorithm executed through the suite of command-line tools in the BLAST+ executables [25]. Sequences that failed to align to the local database were BLASTed against the NCBI database. Novel sequences were aligned in MAFFT [26] to confirm unique haplotypes. Phylogenetic relationships among all resolved haplotypes were explored using the program IQ-TREE [27]. The best model of nucleotide substitution and the maximum-likelihood tree were simultaneously inferred using the IQ-TREE algorithm using an alignment of all reference sequences trimmed to the 379 bp length of the eDNA consensus sequences. Branch support for maximum-likelihood trees was estimated using the ultrafast bootstrap approach implemented in IQ-TREE [28] using 5000 bootstrap replicates. Both Phi$_{ST}$ and χ^2 were used to generate preliminary estimates of genetic differentiation within southeast Alaska from mtDNA sequence data using the R package STRATAG [29].

3. Results

Forty surface seawater eDNA samples were collected during shipboard surveys in southeast Alaska inland waters on 10 different days between July and September 2016 (figure 1). The number of porpoise within visual range of the focal fluke print at the time of sample collection ranged from 1 to 15 (mean 3.95 ± 3.74). Harbour porpoise eDNA amplified in all 36 seawater samples collected from porpoise fluke prints. All field negative controls ($n = 4$) failed to amplify above the amplification threshold within 45 cycles with species-specific primers and the TaqMan probe. The median concentration of porpoise DNA for the 36 eDNA samples was 0.0033 ng µl^{-1}, with values ranging from 8.96×10^{-5} to 0.112 ng µl^{-1} (average $= 0.012$ ng µl^{-1}, s.d. $= 0.022$; electronic supplementary material). No cross-species amplification with harbour porpoise qPCR probes was detected in tests with genomic DNA from killer whale and Dall's porpoise.
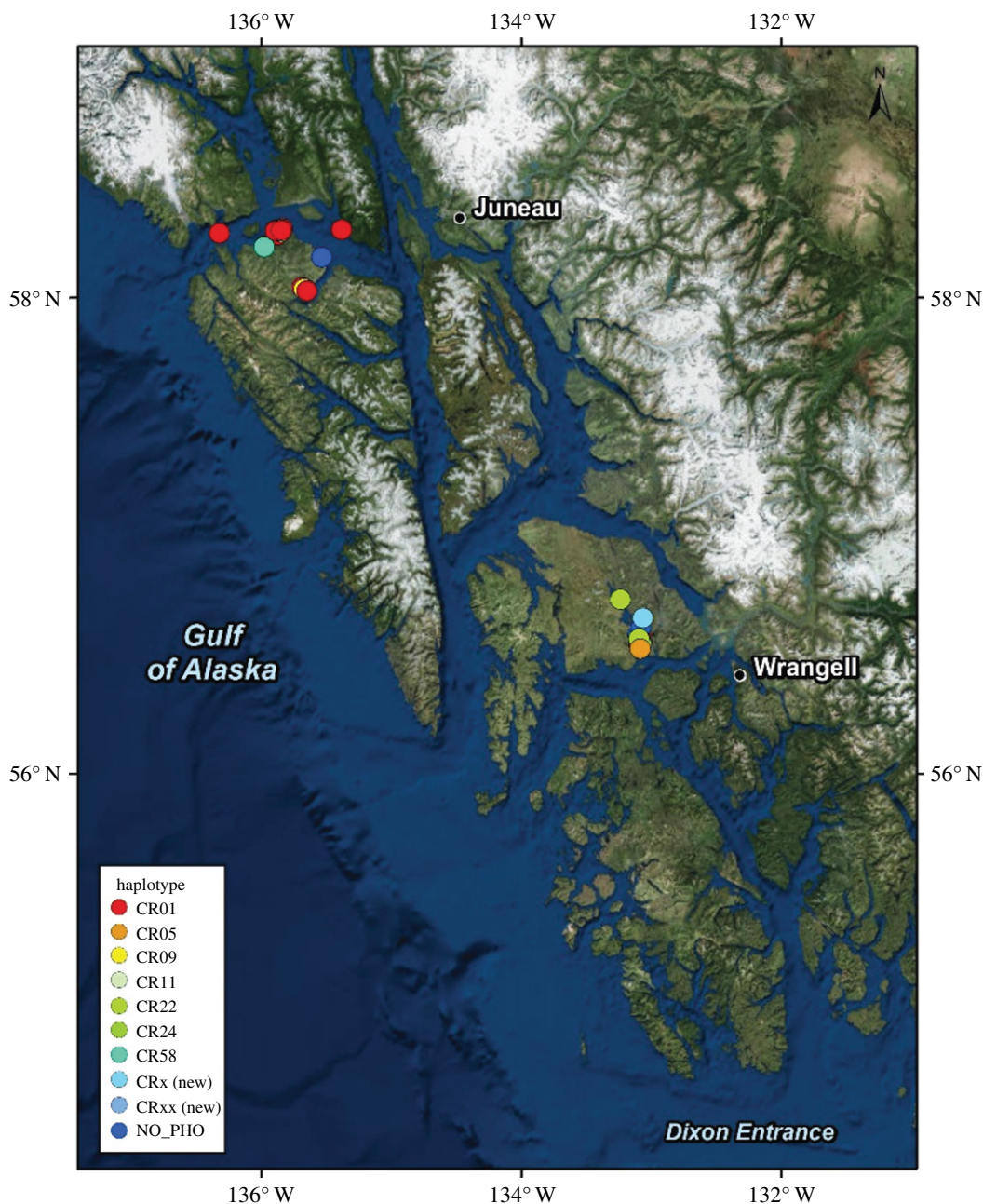
Figure 1. Environmental DNA (eDNA) sampling locations in southeast Alaska. Colours indicate mitochondrial control region haplotypes resolved from 36 eDNA samples (basemap source: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community).

A total of 34 874 482 raw paired-end reads were produced from a pooled library consisting of eDNA samples and positive controls, sequenced on a single Illumina MiSeq flow cell. The average number of reads passing filter per sample was 237 434.50 (s.d. = 74 914.37). Positive controls had an average of 101 643.75 reads after quality filtering, and between 93.78% and 95.56% of these reads represented the true, known control region haplotype for each sample. The maximum error rate for controls was 1.70% (average = 1.36%; s.d. = 0.29%) calculated based on the proportion of processed reads representing the most abundant sequence that was not identical to the known haplotype. This maximum error rate calculated from the control samples was carried forward in the analysis of eDNA samples; all de-replicated sequences representing less than 2% of processed reads were assumed to be either erroneous or non-target and eliminated from the dataset. All negative controls, including no-template PCR negatives, extraction blanks and four field negatives, failed to produce any amplification product detectable on the agarose gel.
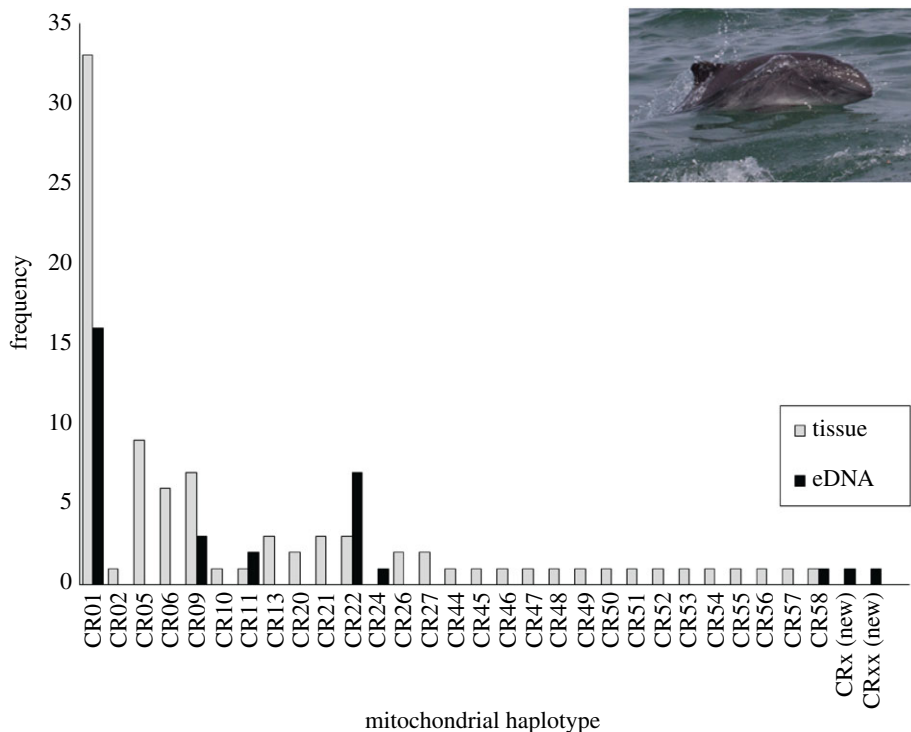
Figure 2. Frequency distribution of Alaska harbour porpoise control region (mitochondrial) haplotypes from tissue ($n = 88$; unpublished) and seawater eDNA ($n = 36$) samples (GenBank accession number available in electronic supplementary material, table S2). Photo credit D. Webster, NOAA National Marine Fisheries Service.

Harbour porpoise control region sequences with a minimum length of 379 bp were successfully generated from 29 of 36 (80.56%) targeted eDNA samples. Gel-purified product from the libraries of seven samples failed to generate sequences of sufficient read depth to meet our stringent error rate criterion of greater than 2% of processed reads per sample, based on the analysis of sequences from tissue controls (above). The most abundant sequence for these eDNA samples either failed to align to any GenBank sequence, or aligned to the genome of a marine plankton (DQ257435) when BLASTed against the NCBI database. The total number of de-replicated reads for one sample (2016HP46) was atypically low (510) and although the sequence showed a high degree of similarity (99.74% identity) to a known harbour porpoise haplotype, it was not included as a newly discovered haplotype because the processed read depth relative to the total reads (221 922 merged pair contigs) was extremely low (less than 1% of all merged pair contigs) and may represent sequencing error.

Of the 29 eDNA samples that produced unequivocal harbour porpoise control region sequences, 26 samples yielded a single control region haplotype and three samples contained eDNA from multiple porpoise having two different haplotypes (electronic supplementary material). In all samples with multiple haplotypes, the less abundant haplotype was represented by greater than 20% post-processed reads. Eight unique haplotypes (based on a 379 bp fragment of the control region) were resolved from the eDNA samples, including two previously unreported haplotypes. Both were found in samples (2016HP08 and 2016HP78) that yielded multiple haplotypes (electronic supplementary material; figure 2). The high read depth (21 427 and 18 076, respectively) and the proportion of de-replicated sequences that comprised greater than 20% of post-processed reads (22.52% and 21.53%, respectively) suggest a very high probability that these secondary haplotypes are true control region haplotypes rather than sequencing error. The most common haplotype (CR01) occurred in 16 eDNA samples collected on 6 different days across the geographical survey range (figures 1 and 2). Of the eight mitochondrial haplotypes found in eDNA samples, five were previously known from tissue samples from Alaska harbour porpoise carcasses, and another had been previously identified from porpoise carcasses recovered in central and northern California (SWFSC 2017, unpublished). The most frequently observed haplotype in the eDNA samples was also the most common haplotype among a set of 88 tissue samples collected over multiple decades (1988–2016) from porpoise strandings and fisheries by-catch in Alaska waters (K. Parsons and P. Morin unpublished data, NOAA, National Marine Fisheries Service;
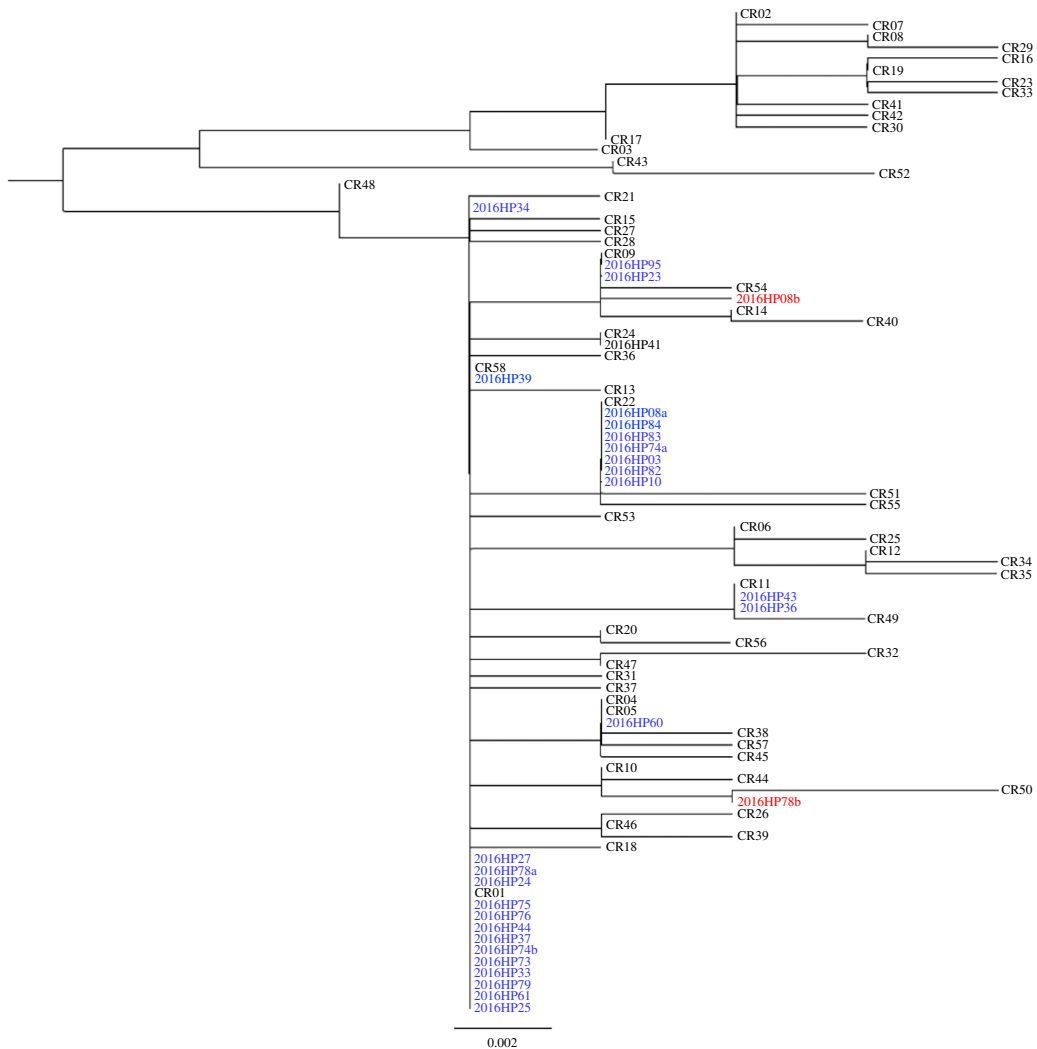
Figure 3. Unrooted maximum-likelihood bootstrap consensus tree based on the best fit substitution model (HKY_G4) determined by Bayesian information criteria. Known harbour porpoise haplotypes are indicated by CR01–CR58 (GenBank accession numbers available in electronic supplementary material, table S2). eDNA samples are denoted by blue text and samples with new haplotypes are labelled with red text. Sample names followed by 'a' or 'b' denote samples with multiple control region haplotypes.

figure 2). A phylogenetic tree based on the best fit substitution model (HKY_G4) and 28 informative sites across the 379 nucleotide sites highlights the similarity of the two new porpoise haplotypes from eDNA samples to previously known control region haplotypes (figure 3). Estimates of genetic differentiation (Phi$_{ST}$ = 0.132, $p = 0.022$; $\chi^2 = 16.24$, $p = 0.002$) based on the geographical distribution of eDNA haplotypes suggest that genetic structure within southeast Alaska is likely.

4. Discussion

Identifying spatial genetic structure in natural populations is critical for defining management stocks and identifying region-specific susceptibility to anthropogenic and environmental stressors. Genetic data are particularly informative for marine vertebrates, such as cetaceans, that tend to be broadly distributed and capable of long-range dispersal. Characterizing this structure in harbour porpoise has, until now, been limited to the analysis of samples collected from beach-cast carcasses or fisheries by-catch; such sample types are typically restrictive because the rate of sample recovery from these opportunistic sources is low, and sample origin is often uncertain. The technological advances associated with next-generation sequencing have increased both the sensitivity and sequencing read depth, making it possible to generate mitochondrial sequence data from dilute target DNA. In the present study, we developed species-specific

quantitative assays and generated mitochondrial haplotypes from eDNA seawater samples collected in the fluke prints of harbour porpoise. The high success rate of this approach and the amplification of sequence data that can contribute to reliable estimates of genetic diversity offer a valuable new avenue for generating key genetic data to guide the conservation and management of this small cetacean.

All surface seawater samples were collected in the fluke prints of diving porpoise in an effort to maximize the amount of target DNA in eDNA samples. Although the concentration of porpoise eDNA estimated through qPCR revealed considerable heterogeneity among samples (electronic supplementary material), harbour porpoise successfully amplified above the minimum amplification threshold from all eDNA samples. eDNA concentration has been correlated with the density and biomass of targeted species [22,30–32], and as such, eDNA sampling can be an informative addition to surveys aimed at estimating seasonality and distribution of aquatic species [9,22,33–36]. In addition to the biomass of target species, other biotic and abiotic factors are likely to affect the detectability of eDNA in seawater samples. A number of studies have attempted to examine the fate of eDNA in different aquatic environments and identify contributing factors that may act to preserve or degrade eDNA. Prolonged exposure to ultraviolet light [31], surface water temperature and pH [37,38], microbial community composition and the presence of tannins and other inhibitors of downstream PCR assays and the degradation of DNA in cellular material due to the time elapsed since deposition or shedding will all affect the quality and quantity of detectable eDNA in varied (and often unpredictable) ways [39,40]. Shedding rates of cellular material into the environment also appear to vary among species and among individuals in relation to behavioural states [37,38,41].

Designing assays to target organellar genomes, such as mitochondria, will help maximize the detection of target DNA in environmental samples due to the high copy number per cell compared to the nuclear genome [12]. Targeting mitochondrial gene regions maximizes the probability of detection not only due to the relative availability of target DNA per cell, but also extended persistence of detectable DNA in seawater due to the protection afforded by organellar membranes [42,43]. In the present study, we targeted regions of the mitochondrial genome for both qPCR and amplicon sequencing not only to maximize the probability of detection due to the inherent characteristics mentioned above, but also because the targeted 5′ section of the mitochondrial control region encompasses the specific sequence of interest used for past (and future) population genetic analyses. The ability to generate population-specific mitochondrial haplotypes from seawater samples dramatically expands future opportunities to examine the population genetic structure of small, threatened cetacean populations. Furthermore, the focused collection of seawater samples from porpoise fluke prints provides additional metadata useful to the specific needs of stock structure analyses because, unlike beach-cast strandings, the spatio-temporal attributes of each sample are known with greater certainty.

Harbour porpoise mtDNA haplotypes were generated from 80% of seawater samples and the difference in relative success rates between the real-time PCR and next-generation sequencing may reflect differences in the relative sensitivities of the two PCR-based approaches with greater sensitivity for the TaqMan™ probe in detecting low concentration target DNA [44]. Three of the 30 samples that generated harbour porpoise sequence data contained DNA from multiple porpoise with different haplotypes, with two different haplotypes resolved in each of the three samples. In all three multi-haplotype samples, one haplotype was dominant, comprising more than 59% of the read depth after quality filtering sequences (electronic supplementary material). In southeast Alaska, harbour porpoise group size varies seasonally, but typical groups range in size from singletons to groups of three individuals [45]. Seawater samples for the present study were collected in the fluke prints of a single porpoise, but the number of porpoises within the sampled group ranged from one to a maximum of 15. The persistence of eDNA in aquatic samples, and the spread of genetic material from its source organism is subject to a large number of variables; as such, it is difficult to predict with any certainty the number of species or organisms contributing genetic material to a given seawater sample [37,39,46]. Furthermore, while reliance upon mtDNA markers has many advantages, it also limits our ability to accurately estimate the number of genetic contributors within a given sample because of its unilateral mode of inheritance and highly conserved nature. Future development of methodologies that allow us to use species-specific nuclear DNA markers, such as single nucleotide polymorphisms, in combination with next-generation sequencing of eDNA will further enhance the value of these samples by allowing us to generate nuclear genotypes and apply models to estimate the minimum number of individuals within each seawater sample for target species.

We demonstrate here how high-throughput sequencing of seawater eDNA can generate mtDNA haplotypes for characterizing genetic diversity for marine vertebrate populations. Prior knowledge of haplotypic sequence diversity in Alaska harbour porpoise provided baseline data for comparison with

eDNA haplotypes. Of the eight unique mtDNA sequences generated from eDNA samples, five were represented in the Alaska porpoise baseline dataset, with the most common eDNA haplotype also being the predominant haplotype from porpoise tissue samples (unpublished data). The two new haplotypes (CRx and CRxx) differed from two previously documented haplotypes (CR54 and CR50, respectively) by only a single nucleotide, and with representation by greater than 20% of final processed reads, these haplotypes were determined to be true genetic variants that have not yet been found in tissue samples from fisheries by-catch and stranded animals. While it is difficult to determine empirically whether or not the new haplotypes are true haplotypes or artefacts generated by early PCR error, resequencing of these samples and the collection of additional samples throughout the surveyed waters will help to determine the veracity of new sequence variants. The discovery of new haplotypes was not unexpected, considering the relatively high diversity of control region haplotypes and frequency of unique haplotypes found in single samples from the opportunistic tissue samples collected from Alaska harbour porpoise (figure 2). Further genetic sampling of harbour porpoise in inland Alaska waters will provide additional confirmation of these new haplotypes through discovery in additional tissue and eDNA samples.

Our goal was to examine whether population genetic data could be generated for small cetaceans through next-generation sequencing of eDNA samples. The high success rate and quality of mitochondrial sequences obtained by collecting surface seawater in the wake of focal animals offers a valuable alternative to traditional methods of collecting genetic samples that have proved to be largely ineffective for real-time biosampling of small, elusive cetaceans such as harbour porpoise. Amplicon sequencing using eDNA offers an approach that should allow the collection of a large number of samples and more thorough coverage of the surveyed geographical range, avoiding common complications in population genetic studies confronted by under-representation or bias in opportunistically collected samples. For example, Chivers *et al.* [18] noted that samples collected from stranded porpoise represented discrete locales within the population's range that did not necessarily correspond to the species' regional distribution, leaving some large numbers of porpoise essentially unsampled. In this case, disjunct sampled areas dictated *a priori* strata rather than true geographical breaks in the population's distribution. Although this limits inference of intraspecific population structure, significant genetic differentiation was detected between the two geographically sampled regions, warranting dedicated sampling throughout suitable habitat to identify significant population boundaries. In southeast Alaska, concentrations of harbour porpoise were consistently found in the same patchy regions throughout inland waters with concentrations in the northern and southern reaches [6]. If significant intraspecific structuring exists, as has been found in other regions [18,47,48], discontinuous distributions may indicate a highly stratified population. Smaller populations may not be able to withstand existing pressures of fisheries by-catch and other habitat altering anthropogenic activities, highlighting the need to determine biologically relevant management units.

Population genetics remains key for the conservation and management of natural populations, and contributes essential information needed to understand and mitigate human-mediated impacts on coastal marine species. This is particularly relevant for higher trophic, long-lived marine mammals which can serve as important barometers of the health of marine ecosystems [1,49–51]. The impacts of anthropogenic activities on cetacean populations can be complex and synergistic. While directed takes such as subsistence hunts can be measured empirically, other activities such as incidental takes and indirect effects due to environmental and prey impacts are best assessed by looking at temporal trends in population dynamics and changes in population size and structure. Genetic data provide a powerful tool for monitoring marine mammal populations, but collecting the requisite tissue samples from some of the smaller cetacean species can be problematic. Recent studies exploring the use of eDNA sampling for genetic monitoring of cetacean populations have demonstrated promise for future applications including detection and species identification [36,52–53]. This study represents a significant step in designing an approach to generate population genetic data for small and threatened marine vertebrates through eDNA methodologies.

References

1. Wells RS *et al.* 2004 Bottlenose dolphins as marine ecosystem sentinels: developing a health monitoring system. *EcoHealth* **1**, 246–254. (doi:10.1007/s10393-004-0094-6)

2. Bosart GD. 2006 Marine mammals as sentinel species for oceans and human health. *Oceanography* **19**, 134–137.

3. Jefferson TA, Curry BE. 1994 A global review of porpoise (*Cetacea: Phocoenidae*) mortality in gillnets. *Biol. Conserv.* **67**, 167–183. (doi:10.1016/0006-3207(94)90363-8)

4. Read AJ. 1994 Interactions between cetaceans and gillnet and trap fisheries in the Northwest Atlantic. *Spec. Issue Rep. Int. Whaling Comm.* **15**, 133–147.

5. Barlow J, Swartz SL, Eagle TC, Wade PR. 1995 *U.S. marine mammal stock assessments: guidelines for preparation, background and a summary of the 1995 assessments*.

6. Dahlheim ME, Zerbini AN, Waite JM, Kennedy AS. 2015 Temporal changes in abundance of harbor porpoise (*Phocoena phocoena*) inhabiting the inland waters of Southeast Alaska. *Fish. Bull.* **113**, 242–256. (doi:10.7755/FB.113.3.2)

7. Creer S, Seymour M. 2017 Marine ecology: genetics from a drop in the ocean. *Nat. Ecol. Evol.* **1**, 37. (doi:10.1038/s41559-016-0037)

8. Goldberg CS, Pilliod DS, Arkle RS, Waits LP. 2011 Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. *PLoS ONE* **6**, e22746. (doi:10.1371/journal.pone.0022746)

9. Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008 Species detection using environmental DNA from water samples. *Biol. Lett.* **4**, 423–425. (doi:10.1098/rsbl.2008.0118)

10. Valentini A *et al.* 2016 Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Mol. Ecol.* **25**, 929–942. (doi:10.1111/mec.13428)

11. Leray M, Knowlton N. 2015 DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proc. Natl Acad. Sci. USA* **112**, 2076–2081. (doi:10.1073/pnas.1424997112)

12. Thomsen PF, Willerslev E. 2015 Environmental DNA: an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* **183**, 4–18. (doi:10.1016/j.biocon.2014.11.019)

13. Sigsgaard EE *et al.* 2016 Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nat. Ecol. Evol.* **1**, 0004. (doi:10.1038/s41559-016-0004)

14. Longmire JL, Maltbie M, Baker RJ. 1997 *Use of 'lysis buffer' in DNA isolation and its implication for museum collections*. Lubbock, TX: Museum of Texas Tech University.

15. Sambrook J, Fritsch EF, Maniatis T. 1989 *Molecular cloning: a laboratory manual*, 2nd edn. New York: NY: Cold Spring Harbor Laboratory Press.

16. Sambrook J, Russell DW. 2001 *Molecular cloning: a laboratory manual*, 3rd edn. New York: NY: Cold Spring Harbor Laboratory Press.

17. Renshaw MA, Olds BP, Jerde CL, McVeigh MM, Lodge DM. 2015 The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol–chloroform–isoamyl alcohol DNA extraction. *Mol. Ecol. Resour.* **15**, 168–176. (doi:10.1111/1755-0998.12281)

18. Chivers S, Dizon A, Gearin P, Robertson K. 2002 Small-scale population structure of eastern North Pacific harbour porpoises (*Phocoena phocoena*) indicated by molecular genetic analyses. *J. Cetacean Res. Manage.* **4**, 111–122.

19. Rosel PE, Dizon AE, Haygood MG. 1995 Variability of the mitochondrial control region in populations of the harbour porpoise, *Phocoena phocoena*, on interoceanic and regional scales. *Can. J. Fish. Aquat. Sci.* **52**, 1210–1219. (doi:10.1139/f95-118)

20. Crossman CA, Barrett-Lennard LG, Taylor EB. 2014 Population structure and intergeneric hybridization in harbour porpoises, *Phocoena phocoena*, in British Columbia, Canada. *Endangered Species Res.* **26**, 1–12. (doi:10.3354/esr00624)

21. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012 Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinf.* **13**, 134. (doi:10.1186/1471-2105-13-134)

22. Thomsen PF, Møller PR, Sigsgaard EE, Knudsen SW, Jørgensen OA, Willerslev E. 2016 Environmental DNA from seawater samples correlate with trawl catches of subarctic, deepwater fishes. *PLoS ONE* **11**, e0165252. (doi:10.1371/journal.pone.0165252)

23. Martin M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12. (doi:10.14806/ej.17.1.200)

24. Rognes T, Flouri T, Nichols B, Quince C, Mahe F. 2016 VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584. (doi:10.7717/peerj.2584)

25. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST+: architecture and applications. *BMC Bioinf.* **10**, 421. (doi:10.1186/1471-2105-10-421)

26. Katoh K, Misawa K, Kuma KI, Miyata T. 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.* **30**, 3059–3066. (doi:10.1093/nar/gkf436)

27. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. 2016 W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids Res.* **44**, W232–W235. (doi:10.1093/nar/gkw256)

28. Minh BQ, Nguyen MAT, von Haeseler A. 2013 Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195. (doi:10.1093/molbev/mst024)

29. Archer FI, Adams PE, Schneiders BB. 2017 STRATAG: an R package for manipulating, summarizing and analysing population genetic data. *Mol. Ecol. Resour.* **17**, 5–11. (doi:10.1111/1755-0998.12559)

30. Takahara T, Minamoto T, Yamanaka H, Doi H, Kawabata ZI. 2012 Estimation of fish biomass using environmental DNA. *PLoS ONE* **7**, e35868. (doi:10.1371/journal.pone.0035868)

31. Pilliod DS, Goldberg CS, Arkle RS, Waits LP. 2014 Factors influencing detection of eDNA from a stream-dwelling amphibian. *Mol. Ecol. Resour.* **14**, 109–116. (doi:10.1111/1755-0998.12159)

32. Pilliod DS, Goldberg CS, Arkle RS, Waits LP. 2013 Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Can. J. Fish. Aquat. Sci.* **70**, 1123–1130. (doi:10.1139/cjfas-2013-0047)

33. Stewart K, Ma H, Zheng J, Zhao J. 2017 Using environmental DNA to assess population-wide spatiotemporal reserve use. *Conserv. Biol.* **31**, 1173–1182. (doi:10.1111/cobi.12910)

34. Sigsgaard EE, Nielsen IB, Carl H, Krag MA, Knudsen SW, Xing Y, Holm-Hansen TH, Møller PR, Thomsen PF. 2017 Seawater environmental DNA reflects seasonality of a coastal fish community. *Mar. Biol.* **164**, 128. (doi:10.1007/s00227-017-3147-4)

35. Takahara T, Minamoto T, Doi H. 2013 Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE* **8**, e56584. (doi:10.1371/journal.pone.0056584)

36. Foote AD *et al.* 2012 Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE* **7**, e41781. (doi:10.1371/journal.pone.0041781)

37. Strickler KM, Fremier AK, Goldberg CS. 2015 Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biol. Conserv.* **183**, 85–92. (doi:10.1016/j.biocon.2014.11.038)

38. Klymus KE, Richter CA, Chapman DC, Paukert C. 2015 Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biol. Conserv.* **183**, 77–84. (doi:10.1016/j.biocon.2014.11.020)

39. Barnes MA, Turner CR, Jerde CL, Renshaw MA, Chadderton WL, Lodge DM. 2014 Environmental conditions influence eDNA persistence in aquatic systems. *Environ. Sci. Technol.* **48**, 1819–1827. (doi:10.1021/es404734p)

40. Wilcox TM, McKelvey KS, Young MK, Sepulveda AJ, Shepard BB, Jane SF, Whiteley AR, Lowe WH, Schwartz MK. 2016 Understanding environmental DNA detection probabilities: a case study using a

stream-dwelling char *Salvelinus fontinalis*. *Biol. Conserv.* **194**, 209–216. (doi:10.1016/j.biocon.2015.12.023)

41. Sassoubre LM, Yamahara KM, Gardner LD, Block BA, Boehm AB. 2016 Quantification of environmental DNA (eDNA) shedding and decay rates for three marine fish. *Environ. Sci. Technol.* **50**, 10 456–10 464. (doi:10.1021/acs.est.6b03114)

42. Walters SP, Yamahara KM, Boehm AB. 2009 Persistence of nucleic acid markers of health-relevant organisms in seawater microcosms: implications for their use in assessing risk in recreational waters. *Water Res.* **43**, 4929–4939. (doi:10.1016/j.watres.2009.05.047)

43. Mann EE *et al.* 2009 Modulation of eDNA release and degradation affects *Staphylococcus aureus* biofilm maturation. *PLoS ONE* **4**, e5822. (doi:10.1371/journal.pone.0005822)

44. Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, Whiteley AR, Schwartz MK. 2013 Robust detection of rare species using environmental DNA: the importance of primer specificity. *PLoS ONE* **8**, e59520. (doi:10.1371/journal.pone.0059520)

45. Dahlheim ME, White PA, Waite JM. 2009 Cetaceans of Southeast Alaska: distribution and seasonal occurrence. *J. Biogeogr.* **36**, 410–426. (doi:10.1111/j.1365-2699.2008.02007.x)

46. O'Donnell JL, Kelly RP, Shelton AO, Samhouri JF, Lowell NC, Williams GD. 2017 Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ* **5**, e3044. (doi:10.7717/peerj.3044)

47. Tolley KA, Vikingsson GA, Rosel P. 2001 Mitochondrial DNA sequence variation and phylogeographic patterns in harbour porpoises (*Phocoena phocoena*) from the North Atlantic. *Conserv. Genet.* **2**, 349–361. (doi:10.1023/A:1012561811074)

48. Rosel PE, France SC, Wang JY, Kocher TD. 1999 Genetic structure of harbour porpoise *Phocoena phocoena* populations in the northwest Atlantic based on mitochondrial and nuclear markers. *Mol. Ecol.* **8**, S41–S54. (doi:10.1046/j.1365-294X.1999.00758.x)

49. Fleishman E *et al.* 2016 Monitoring population-level responses of marine mammals to human activities. *Mar. Mamm. Sci.* **32**, 1004–1021. (doi:10.1111/mms.12310)

50. Veldhoen N, Ikonomou MG, Helbing CC. 2012 Molecular profiling of marine fauna: integration of omics with environmental assessment of the world's oceans. *Ecotoxicol Environ. Saf.* **76**, 23–38. (doi:10.1016/j.ecoenv.2011.10.005)

51. Bik EM *et al.* 2016 Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nat. Commun.* **7**, 10516. (doi:10.1038/ncomms10516)

52. Baker CS, Steel D, Nieukirk S, Klinck H. 2018 Environmental DNA (eDNA) from the wake of the whales: droplet digital PCR for detection and species identification. *Front. Marine Sci.* **5**, 133. (doi:10.3389/fmars.2018.00133)

53. Ma H, Stewart K, Lougheed S, Zheng J, Wang Y, Zhao J. 2016 Characterization, optimization, and validation of environmental DNA (eDNA) markers to detect an endangered aquatic mammal. *Conserv. Genet. Resour.* **8**, 561–568. (doi:10.1007/s12686-016-0597-9)

54. Parsons KM, Everett M, Dahlheim M, Park L. 2018 Data from: Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. (http://dx.doi.org/10.5061/dryad.77h8fd8)