Environmental DNA captures the genetic diversity of bowhead whales (*Balaena mysticetus*) in West Greenland

Dóra Székely¹ | Natasja Lykke Corfixen¹ | Louise Løvschall Mørch¹ | Steen Wilhelm Knudsen²,³ | Morgan L. McCarthy¹ | Jonas Teilmann⁴ | Mads Peter Heide-Jørgensen⁵ | Morten Tange Olsen¹

¹Evolutionary Genomics Section, Globe Institute, University of Copenhagen, Copenhagen, Denmark
²NIVA Denmark Water Research, Copenhagen, Denmark
³Natural History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark
⁴Marine Mammal Research, Department of Bioscience, Aarhus University, Aarhus, Denmark
⁵Department of Birds and Mammals, Greenland Institute of Natural Resources, Nuuk, Greenland

Correspondence
Dóra Székely and Morten Tange Olsen, Section for Evolutionary Genomics, GLOBE Institute, University of Copenhagen, Øster Farimagsgade 5, Copenhagen, Denmark. Emails: dora.szekely@sund.ku.dk (D.S.); morten.olsen@sund.ku.dk (M.O.)

Funding information
Den Grønlandske Fond, Office of Naval Research; Marie Skłodowska-Curie, Grant/Award Number: 801199; Carlsberg Foundation, Grant/Award Number: CF18-0754

Abstract
Environmental changes are prominent in Arctic ecosystems, where the distribution, abundance, life history, and health of marine organisms such as the bowhead whale (*Balaena mysticetus*) are tightly connected to sea ice and sea temperature. However, due to logistical and other challenges of data collection in the Arctic, appropriate assessments of past, present and future effects of climate change and human activities are lacking for many Arctic species. Environmental DNA (eDNA) is emerging as a non-invasive and cost-effective way of obtaining genetic material from the environment and has the potential to complement traditional methods for biodiversity and genetic monitoring. In this study, we investigate whether eDNA isolated from seawater samples has the capacity to capture the genetic diversity of bowhead whales in Disko Bay, West Greenland, for the implementation of long-term genetic monitoring programs of key Arctic marine species. A total of 41 eDNA “footprint” samples were obtained from the water surface after a whale had dived and an additional 54 eDNA samples were collected along transect lines. Samples were screened for bowhead DNA using a species-specific qPCR primer and probe assay, and a subset of 30 samples were successfully Sanger-sequenced to generate individual mitochondrial control region haplotypes. Moreover, by shotgun sequencing ten footprint samples on an Illumina NovaSeq platform we show that footprints generally contain less than 1% endogenous DNA, resulting in partial mitochondrial genomes in four samples out of ten samples. Our findings suggest that sampling in the footprint or wake of traveling animals is a promising method for capturing the genetic diversity of bowhead whales and other marine megafauna. With optimization of sampling and target DNA sequencing for higher endogenous DNA yield, seawater eDNA samples have a large potential for implementation in the long-term population genetic monitoring of marine megafauna in the Arctic and elsewhere.

KEYWORDS
Arctic, climate change, eDNA, marine mammal, mitochondrial DNA, noninvasive, population genetics, shotgun sequencing

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. Environmental DNA published by John Wiley & Sons Ltd
INTRODUCTION

The Arctic is experiencing dramatic changes as a consequence of global warming and increased human activities, including natural resource utilization, tourism and shipping (Stroeve et al., 2012; Yadav et al., 2020). The predicted effects on flora and fauna include changes in distribution, abundance, and population connectivity with likely adverse effects on Arctic species, and beneficial effects on more temperate species (Hauser et al., 2018; Laidre & Heide-Jørgensen, 2012; Wassmann et al., 2011).

Population genetic monitoring is a powerful tool for assessing temporal changes in species distribution, abundance, diversity, and connectivity, and for delineating units for conservation and management (Palsbøll et al., 2007; Palsbøll et al., 2013; Waples & Gaggiotti, 2006). However, due to the inherent vastness of the sea, the often cryptic nature of aquatic species, as well as logistical and methodological challenges associated with field work in the Arctic, genetic monitoring is often lacking for Arctic marine species (Laidre et al., 2015). Thus, a top priority for assessing the effects of environmental change in the Arctic is to develop and implement innovative, yet relatively simple, noninvasive, and cost-effective methods for collecting population genetic data from marine species.

Environmental DNA (eDNA) has recently emerged as a relatively simple, fast, noninvasive, and cost-effective tool for biodiversity monitoring (Bohmann et al., 2014; Thomsen & Willerslev, 2015). Methods utilizing eDNA are based on the premise that all organisms leave a genetic trace in the environment that can be detected through the collection, and processing of environmental samples. Numerous studies have successfully demonstrated the use of eDNA for monitoring the presence of marine species (Carroll et al., 2019; Foote et al., 2012; Miya et al., 2015; Sigsgaard, Torquato, et al., 2020; Thomsen et al., 2012; Valsecchi et al., 2020). However, the applicability of eDNA for population genetic inference has not been extensively investigated, although a few studies have indicated a promising potential for large marine vertebrates (Baker et al., 2018; Parsons et al., 2018; Sigsgaard, Jensen, et al., 2020; Sigsgaard et al., 2016). Cetaceans are known to slough skin regularly while passing through the water column, which may favor their detection by eDNA sampling (Foote et al., 2012). In particular, it has been shown that sampling the so-called cetacean “footprint”—the turbulent surface water that cetaceans leave behind following their breathing and diving sequence—is a successful way to collect genetic material via eDNA samples that contain a sufficient amount of species-specific DNA to sequence parts of the mitochondrial genome (Baker et al., 2018; Parsons et al., 2018). However, to date, the applicability of eDNA for collecting population genetic data in a logistically challenging and highly understudied environment such as the Arctic has not been evaluated.

Here, we explore the use of eDNA for population genetic inference of bowhead whales (Balaena mysticetus) in Disko Bay, West Greenland. The bowhead whale is endemic to Arctic and subarctic waters. Its global population is currently divided into five different management stocks: Bering–Chukchi–Beaufort Seas (BCB), Hudson Bay–Foxe Basin, Davis Strait–Baffin Bay, Svalbard–Barents Sea (Spitsbergen), and Okhotsk Sea (Rugh et al., 2003). These stocks were severely depleted during centuries of intense whaling (Allen & Keay, 2006), but have been recovering since the ban of commercial bowhead whaling in 1946. As a result, the bowhead whale is currently categorized as “Least Concern” by IUCN (Cooke & Reeves, 2018) and limited aboriginal subsistence whaling is allowed by the International Whaling Commission.

While bowhead whales are on the way to a successful recovery from whaling, environmental change is imposing a new challenge. Bowhead whales have a strict yearly migration pattern predominantly following the dynamic changes in seasonal sea ice coverage and spring blooms of Calanoid copepods (Calanus sp.)—its preferred prey (Laidre et al., 2007), (Moore & Reeves, 1993). Recent work has demonstrated that bowhead whales target a narrow range of sea surface temperatures from ~0.5 to 2°C and may be susceptible to thermal stress, changes in food availability, extensive habitat loss, and increased population connectivity (Chambault et al., 2018; Heide-Jørgensen et al., 2012). Previous genetic studies on bowhead whale population connectivity have been unclear, with some indicating population structuring (Postma et al., 2005), while others supporting increased connectivity among populations (Alter et al., 2012). Moreover, the discreteness between Hudson Bay–Foxe Basin and Davis Strait–Baffin Bay has been questioned due to movements of whales between these areas (Heide-Jørgensen et al., 2006). In order to understand the change in distribution and connectivity of bowhead whale populations as a consequence of environmental changes, a noninvasive and cost-effective method for generating additional population genetic data is required.

Here, we first used quantitative polymerase chain reaction (qPCR) to quantify and compare the amount of bowhead whale DNA in seawater samples collected in Disko Bay along survey transects versus water samples collected from the footprints of individual whales. Next, we tested whether eDNA samples—containing bowhead DNA of sufficient quantity and quality based on regular PCR and gel electrophoresis—were suited for Sanger sequencing of mitochondrial control region haplotypes. We compared the resulting population genetic data with published data that had been generated through more than a decade of biopsy sampling and sequencing. Finally, to explore the full potential of eDNA for population inference, we conducted “shotgun” high-throughput sequencing of ten selected footprint samples to evaluate the potential of generating complete mitochondrial genomes.

MATERIALS AND METHODS

Study site

Disko Bay in West Greenland constitutes an ideal study site for assessing the applicability of eDNA for genetic monitoring of bowhead whales. After centuries of absence, bowhead whales reappeared in 2000, showing an increasing population trend (Heide-Jørgensen...
et al., 2007). The region now constitutes an important bowhead whale habitat with about 1,500 animals aggregating each winter and spring to feed extensively on the abundant Calanus spp. before they migrate toward Baffin Bay in May and June (Heide-Jørgensen et al., 2007; Rekdal et al., 2015). The feeding animals are dominantly females (>85%) observed without calves and hypothesized to be part of a larger population ranging in Hudson Bay–Foxe Basin and Davis Strait–Baffin Bay (Heide-Jørgensen et al., 2006).

### 2.2 Seawater eDNA sampling

Seawater eDNA samples were collected in two consecutive field seasons in May 2017 and May 2018 in the northwestern part of Disko Bay (Table 1). In total, we collected 109 eDNA samples, comprising (a) 14 field controls consisting of tap water; (b) 54 samples collected along transect lines irrespective of the presence of bowhead whales; (c) 23 samples specifically collected from the footprint of diving whales; and (d) 18 samples collected from six individual footprints each sampled at three time points: 1, 5, and 10 min after the dive, respectively, to evaluate the temporal stability of the genetic signal from a footprint. During each temporal sampling session, the geographic position of the boat was monitored with the ship's GPS. Samples consisted of 1 liter of water collected at the surface during calm weather (Beaufort Sea State ≤ 2) at air temperatures < 5°C using a sterilized plastic bucket or capped bottle, and subsequently filtered with sterile 0.22 µm filter units (Sterivex) mounted to sterile syringes. Filtering was performed directly onboard the research boat, and if this was not possible, the water sample was kept cold and dark until filtering ashore on the same or following day. Filters were kept at −20°C until DNA extraction following Spens et al. (2017). To compare individual genotypes obtained from biopsy and environmental samples, we collected pairs of biopsy-footprint samples from nine individual whales. Biopsy samples were taken with biopsy darts mounted on a tagging pole or fired from a crossbow and were stored in DMSO at −20°C until DNA extraction following manufacturer’s instructions. These extractions were conducted in the same laboratory as the seawater eDNA samples, but at a later time point. Negative controls were included in both seawater eDNA (N = 10) and biopsy DNA (N = 4) extractions.

### 2.3 DNA extractions

All laboratory work was conducted in a sterile environment, strictly partitioning DNA extraction, amplification, and sequencing, respectively, into separate laboratories. Seawater eDNA extractions were conducted in a sterile laminar flow hood, following UV treatment and a thorough cleaning of all workspace and equipment with 5% bleach and 70% ethanol. Total genomic DNA was extracted from the filter units using a slightly modified QiaGen DNeasy Blood and Tissue Protocol as described by Sigsgaard et al. (2016). At the lysis step, the reagents were added inside the housing of the filters, samples were incubated at 56°C for 3 hr and were eluted twice in 60 µl elution buffer to yield a total of 120 µl purified DNA. Biopsy samples were extracted using the QiaGen DNeasy Blood and Tissue Kit following manufacturer’s instructions. These extractions were conducted in the same laboratory as the seawater eDNA samples, but at a later time point. Negative controls were included in both seawater eDNA (N = 10) and biopsy DNA (N = 4) extractions.

### 2.4 Design of bowhead whale qPCR primers and probe

The amount of bowhead whale DNA in each seawater eDNA sample was quantified using a novel set of species-specific qPCR primers and probe targeting a 191 bp fragment of the mitochondrial DNA (mtDNA) control region following previous similar approaches (Agersnap et al., 2017; Knudsen et al., 2019). Specifically, putative primers were first matched in silico to mtDNA data from bowhead whale (target) and all other cetacean species in the region (nontarget) (Table S3) from NCBI GenBank using Geneious v. 11.0.3 (Kearse et al., 2012), Primer3 v. 0.4.0 and v. 4.1.0 (Koressaar & Remm, 2007; Untergasser et al., 2012) and NCBI Primer-BLAST (Ye et al., 2012).

### TABLE 1 Characteristics of the eDNA samples collected from transects and bowhead whale footprints in Disko Bay, West Greenland

<table>
<thead>
<tr>
<th>eDNA samples</th>
<th>Bowhead qPCR detection</th>
<th>Bowhead mtDNA haplotype yield</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>N</td>
<td>Positives</td>
</tr>
<tr>
<td>Transect</td>
<td>54</td>
<td>44</td>
</tr>
<tr>
<td>Footprint</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>Temporal</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>Sum</td>
<td>95</td>
<td>85</td>
</tr>
</tbody>
</table>

Note: The footprint samples yielded more qPCR positives, had higher amount of bowhead whale DNA molecules, and resulted in better quality mtDNA haplotype data for population genetic inference.

N = sample size; DNA mol. = average number of bowhead whale DNA molecules per L of filtered seawater; SD = standard deviation; Single = a single haplotype; Multiple = multiple haplotypes (sequence ambiguities) in each sample; Other = other species were detected; None = no mtDNA haplotype was obtained.
Next, the primers were tested in vitro by regular PCR to confirm that the primers amplified DNA from the target species, but not any of the non-target species. Then, we designed a TaqMan hydrolysis probe labeled with a FAM (6-Carboxyfluorescein) fluorophore at the 5′-end, and a BHQ-1 dark quencher at the 3′-end and tested the entire assay on target and nontarget species by qPCR (Figure S1). Once initial screening yielded a successful primer–probe assay, the qPCR conditions were optimized for both primers and probe by adjusting annealing temperature and primer–probe concentrations.

2.5 | Bowhead whale environmental DNA quantification with qPCR

Among the tested primer–probe assays, the most successful combination (NFDL5–forward primer, NRDL5–reverse primer, NPDL14–probe) was used to screen eDNA samples (Table S1). The qPCRs were run on an Agilent MxPro 3005 qPCR System with each run containing eDNA samples in four technical replicates, non-template controls in triplicates, and a triplicate genomic DNA dilution series (10[^5]–10[^8] molecules/µl) for absolute quantification of bowhead DNA copy numbers. Field extraction and qPCR controls were treated as eDNA samples and tested alongside sea water eDNA samples in four technical replicates. The total reaction volume of each reaction was 12.5 µl consisting of 5 µl TaqMan Environmental Master Mix 2.0 (EM), 0.5 µl of each primer (NFDL5 (5 µM), NRDL5 (10 µM)), 0.5 µl probe (NPDL14 (7.5 µM)), 4 µl ddH2O, and 2 µl DNA template. The temperature conditions were as follows: preliminary denaturation at 50°C for 5 min and further denaturation at 95°C for 10 min. This was followed by 50 cycles at 95°C denaturation for 30 s and annealing at 60°C for 1 min. Fluorescence data were collected at the endpoint of each cycle, and estimated concentrations were calculated based on standard curve equations of the amplicon dilution series in each run. The statistical analyses of the results were assuming significant deviation from normal distribution (the Shapiro–Wilk test), and non-parametric test statistics were conducted to further investigate the eDNA approach, but not in the population genetic analyses. For these analyses, the resulting eDNA mtDNA control region data were aligned and trimmed to 397 bp to allow for comparison with a large existing mtDNA reference dataset obtained through more than a decade of biopsy collection from bowhead whales in Disko Bay (DB; n = 429; years 2000–2013) and Bering-Chukchi-Beaufort Seas (BCB; n = 371; years 2000–2013) (LeDuc et al., 2008; Rekdal et al., 2015; Wiig et al., 2011). To integrate our eDNA mtDNA data in a population genetic framework, haplotype frequency were estimated in DNAsp v. 6.10.04 (Rozas et al., 2017) and a median-joining haplotype network was constructed using PopART v. 1.7 (Leigh & Bryant, 2015). Moreover, to explore the effect of eDNA sample size on haplotype retrieval, rarefaction curves for the number of haplotypes as a function of sample size were constructed for the eDNA samples, as well as the Disko Bay and BCB population genetic reference data using the function rarecurve in the vegan v. 2.5-4 package (Oksanen et al., 2017) in RStudio v. 1.1.442 (RStudio Team, 2015) with R v. 3.4.4 (R Core Team, 2020).

2.6 | Bowhead whale mtDNA diversity in seawater eDNA

In order to explore the use of eDNA for basic mtDNA haplotype population genetic analyses, environmental samples were amplified using regular PCR and Sanger-sequenced targeting a 519 bp fragment of the mtDNA control region using forward (5′-TCAGCACCCAAAAAGCTGAATT-3′) and reverse (5′-CACCATCGACCCAAAAAGCTG-3′) primers modified from Rekdal et al. (2015). The PCRs included 2.5 µl buffer, 2.5 µl MgCl₂ (25 µM), 1 µl of each primer (10 µM), 1 µl BSA (20 mg/ml), 0.25 µl dNTP (10 µM of each), 0.2 µl AmpliTaq Gold Polymerase (5 U/µl), 3 µl DNA, and 13.55 µl ddH₂O per reaction to reach a final volume of 25 µl. The initial PCR conditions were set to 95°C for 5 min. This was followed by 40 cycles of 95, 55, and 72°C for 30 s each, composing denaturation, annealing, and elongation. The final elongation was set to 72°C for 7 min. PCR products were visualized using agarose gel electrophoresis to confirm that they were of the desired length and that negative controls did not amplify. Successful PCR amplifications were sent for purification, and forward and reverse Sanger sequencing at Macrogen Europe using the PCR primers listed above.

All chromatograms were mapped to a bowhead whale mtDNA reference sequence generated as a majority consensus sequence of all haplotypes detected in previous studies and visually inspected using Geneious mapper using standard high sensitivity settings with five iterations (Kearse et al., 2012). Environmental samples may contain DNA from multiple individual whales, resulting in chromatograms with base ambiguities, but overall high Q-scores (i.e., 95% of the bases having Phred scores Q > 30). These “mixed” chromatograms were included in the evaluation of the eDNA approach, but not in the population genetic analyses. For these analyses, the resulting eDNA mtDNA control region data were aligned and trimmed to 397 bp to allow for comparison with a large existing mtDNA reference dataset obtained through more than a decade of biopsy collection from bowhead whales in Disko Bay (DB; n = 429; years 2000–2013) and Bering-Chukchi-Beaufort Seas (BCB; n = 371; years 2000–2013) (LeDuc et al., 2008; Rekdal et al., 2015; Wiig et al., 2011). To integrate our eDNA mtDNA data in a population genetic framework, haplotype frequency were estimated in DNAsp v. 6.10.04 (Rozas et al., 2017) and a median-joining haplotype network was constructed using PopART v. 1.7 (Leigh & Bryant, 2015). Moreover, to explore the effect of eDNA sample size on haplotype retrieval, rarefaction curves for the number of haplotypes as a function of sample size were constructed for the eDNA samples, as well as the Disko Bay and BCB population genetic reference data using the function rarecurve in the vegan v. 2.5-4 package (Oksanen et al., 2017) in RStudio v. 1.1.442 (RStudio Team, 2015) with R v. 3.4.4 (R Core Team, 2020).

2.7 | Shotgun sequencing of bowhead whale footprints

To fully explore the potential of eDNA for population genetic inference, 10 footprint samples yielding unambiguous mtDNA control region haplotypes and containing the highest amounts of bowhead whale DNA copy numbers (18,000–158,000 molecules/L) as quantified by the qPCR assay were selected for shotgun sequencing in an attempt to retrieve full mitochondrial genomes (mitogenomes). Library blanks were used and treated as samples up until sequencing to track any possible contamination during the library preparation process. Shotgun sequencing was performed using a modified version of the Illumina library preparation protocol by Carèe et al. (2018). The extracted DNA was fragmented into 350 bp fragments in 15 µl volume using ME220 Ultrasonicator (Covaris™). Fragment lengths
and molarity were analyzed with a Tape Station (Agilent), adapter concentration was adjusted to match individual sample concentrations, and libraries were amplified with dual indices in 50 µl reaction volume using 10 µl of libraries. The thermocycler conditions were set to a starting phase of 10 min at 95°C, followed by multiple cycles of denaturing for 20 s at 95°C, annealing for 30 s, extension for 45 s at 72°C, finished with a final elongation of 5 min at 72°C, and final hold at 10°C. The optimal number of amplification cycles was determined individually for each sample using qPCR (MX3005P Agilent). Thermocycler reactions were set as described above with 40 cycles, dissociation curve, and endpoint data collection. The amplified and indexed libraries were purified with magnetic beads (MAGBIO), and the quality was controlled with Tape Station (Agilent). The ten amplified libraries (one for each sample) were shotgun sequenced as a single equimolar pooled library on a single lane of Illumina NovaSeq 6000 at the National High-throughput DNA Sequencing Centre, University of Copenhagen, Denmark.

The raw reads were demultiplexed, and sequence data were quality inspected with FastQC v. 0.11.8a (Andrews, 2010). Alignment files were generated with Geneious Prime v. 2020.1.2 and PALEOMIX (Schubert et al., 2014) in two separate runs using the bowhead whale mitogenome as reference (NCBI GenBank Accession NC_005268.1). Steps included trimming adapters, low-quality bases, and reads shorter than 25 nucleotides. Trimming settings were kept conservative due to the degraded and fragmented nature of eDNA, aiming to conserve any degraded genetic material of the target species in our downstream analysis. Overlapping mate pairs were collapsed with AdapterRemoval2 (Schubert et al., 2016). Reference genomes were indexed with SAMtools (Li et al., 2009), and reads were aligned using the BWA-MEM v. 0.7.17 (Li & Durbin, 2009) algorithm. Duplicates were removed with Picard MarkDuplicates (http://www.broad institute.github.io/picard) as part of the PALEOMIX pipeline, and BBTools (Bushnell, 2019) to remove duplicates from alignments generated by Geneious Prime. The mitogenome consensus sequences were called in Geneious Prime with a majority rule base call and minimum number of two reads.

To confirm that the resulting eDNA mitogenome data were of bowhead whale origin, all mapping reads were first analyzed by the Basic Local Alignment Search Tool (BLAST, Altschul et al., 1990) to confirm bowhead whale identity using 98% percent identity cutoff and only the first best hit was directed to the output. The results of the taxonomic assignment were visualized using KronaTools (Ondov et al., 2011). Second, the consensus sequences generated by the alignments by Geneious Prime and PALEOMIX were phylogenetically compared with the mitochondrial genomes of 14 baleen whale species. The phylogenetic relationships were inferred by PHYML (Guindon et al., 2010) using the nucleotide substitution model of Hasegawa–Kishino–Yano (Hasegawa et al., 1985) applying 1,000 bootstraps and a random seed, and using the hippopotamus (Hippopotamus amphibious) as an outgroup. The best supported phylogenetic tree was visualized in FigTree (Rambaut, 2007). Finally, to estimate the proportion of endogenous (bowhead whale) DNA reads relative to other taxa in the eDNA samples, a subsample of reads (10%, approximately 3,600,000 reads) were analyzed by the BLAST, using a percent identity cutoff of 99% and including maximum five target sequences. The results were visualized using KronaTools.

3 | RESULTS

3.1 | Bowhead whale eDNA sampling

Bowhead whale DNA was successfully detected in all 23 (100%) footprint samples and 44 out of 54 (81%) transect survey samples (Table 1). The amount of bowhead whale DNA was more than 15-fold higher in footprints (mean = 26,424 molecules/L, SD = 38,458) compared with transect samples (mean = 1,428 molecules/L, SD = 2,949) (Figure 1a), underlined by significant difference in means (Wilcoxon’s test; p < 0.05).
In the six additional footprints sampled at 1, 5, and 10 min after the dive, the amount of bowhead whale DNA confirmed a predicted decrease over time. Mean bowhead whale DNA copy number at the first minute was 18,117 molecules/L (SD = 27,225), at minute 5 it was 5,235 molecules/L (SD = 8,128), and 10 min after the dive it was 4,046 molecules/L (SD = 5,824). However, even after 10 min the amount of bowhead whale DNA was higher than in an average transect sample (Figure 1b). While the negative trend was visually confirmed, the difference in DNA copy numbers at different time points was not significant (Friedman test, $p = 0.115$). The footprints also provided more reliable population genetic data, with 14 (61%) of the samples resulting in unambiguous mtDNA control region haplotypes, compared with only six (11%) of the transect samples (Table 1). One of the transect samples yielded an mtDNA sequence identified as narwhal (Monodon monoceros; sample M52 > 99% identity to NCBI accession MK397810) in NCBI BLAST analysis. All 14 field controls, 10 extraction controls, 33 qPCR controls, and 28 PCR controls were negative.
3.2 | Bowhead whale mtDNA diversity in seawater eDNA

Despite the relatively small sample size, the bowhead whale genetic diversity in Disko Bay eDNA samples \( (N = 30; \text{Hd} = 0.885; \pi = 0.0076) \) resembles diversity estimates based on more than a decade of genetic biopsy sampling in Disko Bay \( (N = 429; \text{Hd} = 0.881; \pi = 0.0109) \) (Rekdal et al., 2015). The eDNA samples captured the two most common haplotypes reported in the large set of previously published bowhead whale population genetic data from the Disko Bay \( (N = 429) \) and BCB \( (N = 371) \) stocks (Figure 2a,c). Still, multiple haplotypes from previous studies were not detected by eDNA samples. Indeed, the rarefaction curves suggest that an increased eDNA sampling effort would more effectively capture the haplotypes reported for Disko Bay bowhead whales (Figure 2b). In the nine cases where pairs of footprint and biopsy samples were obtained from the same animal, the mtDNA control region haplotypes were identical in six of the footprint-biopsy pairs, whereas two of the pairs had mismatching base calls in three and five sites, and one pair could not be assessed as Sanger sequencing of the footprint failed.

3.3 | Shotgun sequencing of bowhead whale footprints

The attempt to retrieve mitogenomes from ten bowhead footprint samples was only partially successful. All samples had low coverage in terms of depth and breadth of coverage, and the endogenous DNA content—estimated as the fraction of reads mapping to the bowhead mitogenome—was low and only partly correlated with the amount of bowhead whale DNA quantified by qPCR (Figure 3; Table S2). BLAST analyses resulted in 2% of the reads (79,007) assigned to 34,186 different taxonomic groups based on a subset of the DNA reads. Exactly 65% of all identified reads were of bacterial origin, 32% were assigned to eukaryotic organisms with only 2% assigned to artiodactyls, including cetaceans (Figure 4). Still, four samples (M30, M32, M33, and ED47) yielded enough endogenous reads to obtain partial mitogenomes with >50% coverage. The authenticity of these eDNA mitogenomes was confirmed by phylogenetic analysis, clearly grouping them with the bowhead whale reference mitogenome (Figure 5), as well as by BLAST analyses identifying >95% of the mapped reads as bowhead whale (Figure S2).

4 | DISCUSSION

4.1 | Cetacean footprints are a promising source of genomic DNA

The amount of bowhead whale DNA detected from environmental samples was 15-fold higher in the footprint of a diving whale compared with water samples collected during transect surveys. Moreover, we found that bowhead whale DNA could be detected in footprints for at least 10 min after the dive, providing a short window of time in which to obtain population genetic data. However, some of the samples collected after 10 min resulted in mtDNA control region haplotypes with sequence ambiguities, which indicates either rapid DNA damage, or perhaps more likely mixing of DNA from multiple animals as ambiguities generally were found in mtDNA sites with known variation in bowhead whales (LeDuc et al., 2008; Rekdal et al., 2015; Wiig et al., 2011). Thus, for population genetic inference we recommend collecting footprint samples as fast as possible after the dive.

A recent study investigating the detection of killer whales \( (Orcinus orca) \) in Puget Sound in the Northeast Pacific using a droplet digital PCR (ddPCR) reported successful detection up to two hours after the presence of animals (Baker et al., 2018). Generally, ddPCR is considered to be a highly sensitive method for DNA
In the study of Baker et al. (2018), 15% of the eDNA killer whale detections resulted in mtDNA haplotype population genetic data and some of these from samples collected up to an hour after observing the animals. During sampling, their research vessel was drifting with the current, possibly tracking the same water over time, and it is possible that the obtained data stem from multiple animals, as killer whales tend to travel in family groups with similar mtDNA haplotypes.

Intriguingly, another recent eDNA study on killer whales reported false-negative detection despite animals being visually observed during sampling (Pinfield et al., 2019). The platform used for the study did not allow for footprint sampling; however, the animals were less than 20 m from the boat during sample collection. Our results demonstrated a significant difference in target DNA quantity when comparing footprint and transect sampling, which could in part explain the lack of positive detections in the study of Pinfield et al. (2019). Still, we did also detect target DNA in transect samples when no bowhead whales were visually observed in the close vicinity of the boat, indicating that the false-negative detections reported by Pinfield et al. (2019) could in part be owing to differences in study design (e.g., primer sensitivity and inhibitors), sampling environment (e.g., temperature and sea state), and/or target species (e.g., size of animal and rate of skin shedding and defecation).

Footprints are a relatively easy target when studying larger baleen whales, but eDNA from footprints has also proven successful for obtaining genetic information from the harbor porpoise (Phocoena phocoena) that is much smaller and faster species (Parsons et al., 2018). Thus, footprint sampling for eDNA may turn out to be a powerful noninvasive method for obtaining genomic DNA from a wide range of elusive, fast-swimming, or vulnerable cetacean species such as ziphids, phocoenids, and smaller delphinids that cannot be easily sampled by, for example, biopsy darts, and/or for regions or seasons where animals are difficult to approach. Though it may be difficult to discern individual footprints for social species traveling in pods, as these may contain a mixture of DNA from multiple individuals.

**FIGURE 4** Krona plot indicating the species composition of footprint sample M30 based on a random subsample of the 10% of the total collapsed reads out of the approximately 36,000,000 reads generated for the sample. The majority of DNA reads obtained by shotgun sequencing that could be assigned in NCBI BLAST search were Bacteria (~65%), less than 35% to Eukaryota, and 2% to Artiodactyla, including cetaceans [Colour figure can be viewed at wileyonlinelibrary.com]
4.2 Capturing bowhead whale genetic diversity in seawater eDNA

The mitochondrial control region sequences obtained from Sanger sequencing of bowhead footprint samples successfully captured the most common haplotypes reported in published studies on bowhead whales (LeDuc et al., 2008; Rekdal et al., 2015; Wiig et al., 2011). Conversely, several haplotypes detected by previous studies through biopsy samples were not recovered by environmental samples. Disko Bay is a known winter/spring feeding ground of bowhead whales before they continue their yearly migration across Baffin Bay (Laidre et al., 2007). The individuals that are observed in the area are mainly females without calves and are part of a bigger population ranging toward the Canadian Arctic, Foxe Basin–Hudson Bay, and Davis Strait–Baffin Bay (Heide-Jørgensen et al., 2006). Heide-Jørgensen et al. (2007) suggest that the animals that arrive to Disko Bay are part of a multiyear cycle where mature females without calves arrive to the area to feed before breeding, resulting in a yearly shift in the visiting migratory individuals. Because of this shift additional seasons, eDNA sampling would be needed to capture the genetic diversity of the entire group of bowhead whales visiting Disko Bay, which is also supported by the rarefaction curves on sample size and unique haplotype discovery. With just two relatively short field efforts of 14 days each, we were able to capture a broad representation of haplotypes found in more than a decade of sampling by conventional biopsy methods (LeDuc et al., 2008; Rekdal et al., 2015; Wiig et al., 2011). Indeed, when matching nine pairs of footprint and biopsy samples from individual whales we found six pairs had matching mtDNA haplotypes, one pair failed, and only two sample pairs provided different haplotypes. These two discrepancies could be due to either a mix of DNA from different haplotypes occurring in the footprint, or by losing track of individual whales in instances where multiple whales were present at the same time.

In contrast to our Sanger sequencing approach, our effort to obtain full bowhead whale mitogenomes from eDNA samples was only partly successful. Whole genome shotgun sequencing is often used to assemble mitogenomes from low target DNA concentration samples such as degraded, ancient, and historical specimens (Knapp & Hofreiter, 2010). To our knowledge, ours is the first to utilize it for assembling mitogenomes of marine megafauna extracted from seawater. Our results indicate a very low yield of endogenous bowhead whale DNA reads; however, we obtained a breadth of coverage above 50% for four samples out of the 10 samples that were tested. Of the reads that mapped to the bowhead reference mitogenome, >95% were confirmed to be bowhead whale by BLAST analysis. The “mismatching” reads were often shorter (38–68 bp), and these typically matched 100% to both bowhead and multiple other cetacean and mammalian species. We kept our data preprocessing measures conservative compared with standard recommendations for genomic shotgun data analysis. The quality of eDNA is often degraded, and the samples often innately include shorter free-ranging, extracellular DNA fragments that could potentially originate from the target species. A plausible reason why the shorter reads also matched other cetacean and mammalian species is that the sequences originated from conserved regions of the mitogenome and were simply too short to contain any variable region to properly discern between species. Another reason could be the existence of nuclear copies of mitochondrial DNA (NUMTs).
As with ancient DNA (aDNA) analyses, the main challenge associated with shotgun sequencing of seawater eDNA samples is DNA from nontarget species, such as fish, plankton, and bacteria.

In our attempt to characterize the species composition of the footprint sample that contained the highest number of endogenous reads (M30) we found that 97.8% of the subsampled reads were unassigned by BLAST which limits our understanding of the exact composition of the sample. The measures (85.9%) reported by Stat et al. (2017) for a BLAST analysis of a complete sample (~30 million reads) resemble the unassigned ratios in our study, indicating that the NCBI database can be restrictive for assessments of environmental samples. When nontarget sequences are more abundant in samples, either innately or as part of contamination, it reduces the chance of sequencing target species with sufficient depth. In aDNA studies, contamination is minimized by working in dedicated aDNA clean laboratories and genome coverage is often increased by sequencing deeper (Gilbert & Willerslev, 2007; Willerslev et al., 2003). However, although sequencing prices are declining, pooling fewer than 10 samples on a lane of an Illumina NovaSeq 6000—as done here—will most likely exceed the allocated budget of most marine mammal research funds.

4.3 | Perspectives and recommendations on population genetic monitoring using eDNA

In support of previous findings on whale sharks (Rhincodon typus) in tropical waters (Sigsgaard et al., 2016) and killer whales in temperate waters (Baker et al., 2018), we here obtain population genetic data from bowhead whales in Arctic waters to demonstrate that eDNA is a promising tool for population genetic monitoring of marine megafauna. In particular, we show that sampling in the footprints (or wake) of swimming animals substantially increase the amount of retrieved target DNA, resulting in a relatively high mtDNA haplotype sequencing success rate (>60%), and further methodological developments may make the retrieval of complete mitochondrial and nuclear genomes possible. Still, there are many challenges to overcome before seawater eDNA samples can be applied in routine genetic monitoring programs, as recently discussed by Sigsgaard, Jensen, et al. (2020). Marine megafauna are typically present in low abundances in marine ecosystems compared with other more dominant taxonomic groups such as prokaryotes and eukaryotic cellular organisms resulting in these typically dominating the eDNA sample (Stat et al., 2017). Moreover, once the genetic material is exposed to the environment it immediately starts to degrade due to several environmental factors (e.g., temperature, salinity, pH, UV, microbial activity) at a rate that is currently not completely understood, although most studies have found the degradation rate is exponential (Andruszkiewicz et al., 2017; Sigsgaard et al., 2016; Thomsen et al., 2012). Such degradation could compromise the quality and quantity of the genetic information extracted from water samples.

There are several potential approaches that could be tested to increase the quality and quantity of population genetic data from the target species: (a) filtering the environmental sample in multiple steps with decreasing pore sizes could be tested as a method to capture target cells and DNA, but not, for example, larger sized bacteria, phytoplankton, and zooplankton; (b) applying target capture methods and designing a species-specific bait, which would facilitate amplifying target DNA—both nuclear and mitochondrial sequences—to an amount that could be assembled to genetically informative data (Aylward et al., 2018; Carpenter et al., 2013; Maricic et al., 2010); (c) if interest is mainly in a shorter mtDNA fragment such as the control region, NGS-targeted amplicon sequencing could be further explored to escape some of the limitations of Sanger sequencing, as also demonstrated by Sigsgaard et al. (2016); and ultimately (d) examine the ability of single-molecule real-time sequencing techniques to capture individual genotypes, and aim to develop statistical tools for handling the mixed population genetic data from multiple animals.

Given the ease of collecting environmental samples for eDNA analysis, we see a great potential in involving and training local communities in routine sampling for a community-driven genetic monitoring network. Moreover, with the rapid speed of developments within DNA sequencing technologies, we envision that future monitoring can be performed in real-time via fully automated eDNA-monitoring platforms (e.g., autonomous floating drones or buoys), which integrate eDNA sampling, sequencing (e.g., Oxford Nanopore’s MiniION), analysis, and GSM/satellite-linked data transmission. Such eDNA-sampling buoys termed “Environmental Sample Processors (ESPs)” are currently being developed and tested at several research institutes (Hansen et al., 2020) and may find their way into standard population genetic monitoring programs in the near future. Perhaps with time, eDNA methodology will be developed further to provide information on organismal age, sex, gene expression profiles, pathogens, microbiome, immunological status, and much more, to monitor not just genetic diversity, but also species demography, life history, health, and function.

ACKNOWLEDGMENTS

The authors would like to sincerely thank Abel Brandt, Johannes Mølgård, Ado Isaksen, Tarfi Mølgård, Karl-Jørgen Broberg, Grete Teilmann, Svenja Christiansen, Mette Dalgaaard Agersted, Kim Gustavson, and Eva Friis Møller for assistance and good company in the field. Phil Morin, Andy Foote, Philip Francis Thomsen, Eva Sigsgaard, Peter Rask Møller, Marta Maria Ciucani, Fátima Sánchez Barreiro, Xénia Keighley, Mikkel Skovrind, Mikkel Sinding, and the staff at the National High-throughput Sequencing Centre, University of Copenhagen, Denmark, are thanked for advice and fruitful discussions about eDNA methodology. The project was supported by Den Grønlandske Fond to NLC and LLM, the Office of Naval Research to MPHJ and JT, EU Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 801199 to DSZ, and Carlsberg Foundation grant no. CF18-0754 to MTO.

CONFLICT OF INTEREST

The authors declare no conflicts of interests.
AUTHOR CONTRIBUTION

MTO and MPJ conceived the study; DS, NLC, LJM, JT, MPJ, and MTO provided funding; NLC, LJM, JT, MPJ, and MTO designed and performed the fieldwork; DS, NLC, LJM, and SWK designed and performed the molecular laboratory work; DS, NLC, LJM, SWK, MLM, and MTO analyzed the data; DS, NLC, LJM, and MTO drafted the manuscript; and all authors approved the final version.

DATA AVAILABILITY STATEMENT

All data will be deposited in open-access repository upon acceptance of the manuscript.

ORCID

Morten Tange Olsen https://orcid.org/0000-0001-6716-6345

REFERENCES


**SUPPORTING INFORMATION**

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