NOAA-TM-NMFS-SWFSC-635
September 2020

PROGRESS REPORT ON GENETIC ASSIGNMENT OF HUMPBACK WHALES FROM THE CALIFORNIA-OREGON FEEDING AGGREGATION TO THE MAINLAND MEXICO AND CENTRAL AMERICA WINTERING GROUNDS

Karen K. Martien, Brittany L. Hancock-Hanser, Morgane Lauf, Barbara L. Taylor, Frederick I. Archer, Jorge Urbán Ramírez, Debbie Steel, C. Scott Baker, & John Calambokidis

NOAA Technical Memorandum NMFS

SEPTEMBER 2020

PROGRESS REPORT ON GENETIC ASSIGNMENT OF HUMPBACK WHALES FROM THE CALIFORNIA-OREGON FEEDING AGGREGATION TO THE MAINLAND MEXICO AND CENTRAL AMERICA WINTERING GROUNDS

Karen K. Martien[1], Brittany L. Hancock-Hanser[1], Morgane Lauf[2], Barbara L. Taylor[1], Frederick I. Archer[1], Jorge Urbán[3], Debbie Steel[4], C. Scott Baker[4], John Calambokidis[5]

[1]NOAA Southwest Fisheries Science Center
[2]Ocean Associates, Inc.
[3]Universidad Autonoma de Baja California Sur
[4]Oregon State University
[5]Cascadia Research Collective

NOAA-TM-NMFS-SWFSC-635

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

About the NOAA Technical Memorandum series

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency which establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

SWFSC Technical Memorandums are available online at the following websites:

SWFSC: https://swfsc.noaa.gov

NOAA Repository: https://repository.library.noaa.gov/

NTIS National Technical Reports Library: https://ntrl.ntis.gov/NTRL/

Accessibility information

NOAA Fisheries Southwest Fisheries Science Center (SWFSC) is committed to making our publications and supporting electronic documents accessible to individuals of all abilities. The complexity of some of SWFSC's publications, information, data, and products may make access difficult for some. If you encounter material in this document that you cannot access or use, please contact us so that we may assist you.
Phone: 858-546-7000

Recommended citation

Karen K. Martien, Brittany L. Hancock-Hanser, Morgane Lauf, Barbara L. Taylor, Frederick I. Archer, Jorge Urbán, Debbie Steel, C. Scott Baker, John Calambokidis. 2020. Progress report on genetic assignment of humpback whales from the California-Oregon feeding aggregation to the mainland Mexico and Central America wintering grounds, U.S. Department of Commerce, NOAA Technical Memorandum NMFS-SWFSC-635.

Progress report on genetic assignment of humpback whales from the California-Oregon feeding aggregation to the mainland Mexico and Central America wintering grounds

Karen K. Martien, Brittany L. Hancock-Hanser, Morgane Lauf, Barbara L. Taylor, Frederick I. Archer, Jorge Urbán, Debbie Steel, C. Scott Baker, and John Calambokidis

Executive Summary

Humpback whales undertake annual migrations from low-latitude wintering grounds to higher-latitude feeding grounds. Individuals learn their migratory routes from their mothers and exhibit strong fidelity to both feeding and wintering grounds. This migratory behavior results in a population structure best characterized as migratory herds, with members of the same herd sharing the same wintering and feeding grounds and facing the same environmental conditions and threats throughout the entire year. Published genetic studies have focused on feeding and wintering grounds as the strata of interest. However, understanding the relationships between herds is fundamental to understanding the population structure, dynamics, and life history of humpback whales, and to appropriate management. We use mitochondrial sequence data to genetically characterize the two migratory herds that use the California/Oregon (CA/OR) feeding ground and compare them to each other and to the wintering aggregations of which they are a part. We generated two new data sets for this study. The first consists of full mitochondrial genome sequences (16,384 basepairs) from the herd that feeds off of CA/OR and winters in Central America (henceforth the CentAm-CA/OR herd; n=65), and the herd that also feeds off of CA/OR but winters off of mainland Mexico (henceforth the MMex-CA/OR herd; n=50). The second new data set consists of mitochondrial control region sequences (389 basepairs) from humpback whales sampled off of CA/OR during the 2018 California Current Ecosystems Studies (CCES) survey (n=227). We compare these new data sets to published (Baker *et al.* 2013) mitochondrial control region data sets collected from the mainland Mexico wintering aggregation (n=62), the CA/OR feeding aggregation in 2004 (n=123), and the CA/OR feeding aggregation in 1988-89 (n=49). Our results show that the CentAm-CA/OR herd and MMex-CA/OR herds are significantly genetically differentiated from each other (Φ_{ST} = 0.054 and 0.044 for full mitogenome and control region sequences, respectively). However, the herds share a high proportion of haplotypes, even when using full mitogenome sequences. Consequently, many individuals cannot be reliably assigned to a herd using only mitochondrial data. The MMex-CA/OR herd is significantly differentiated from the mainland Mexico wintering aggregation (Φ_{ST} = 0.044), confirming that wintering ground haplotype frequencies cannot be used as a proxy for haplotype frequencies in the herds that utilize a wintering ground. The 2018 CCES samples did not differ significantly from samples collected from the same feeding aggregation in 1988-89 (Φ_{ST} = 0.015) and 2004 (Φ_{ST} = 0.009), indicating that the genetic composition of the herds occupying the feeding ground has not changed significantly despite a near two-fold increase in abundance during the timespan of the two data sets. Finally, we found evidence that animals that are photographically identified as being part of the MMex-CA/OR herd are not fully representative of that herd. This bias could be due to Central America animals being photographed as they migrate past mainland Mexico and mistakenly identified as mainland Mexico animals, or it could result from spatial or temporal segregation of mainland Mexico animals on the wintering grounds such that some portion of the wintering aggregation is being over- or under-represented in the photographic catalog. Future work should investigate the level of nuclear genetic differentiation between herds, which may allow improved assignment of

animals to migratory herd and yield insight to the potential sources of bias apparent in the MMex-CA/OR herd.

Introduction

Much of what is known about humpback whale migratory routes between low-latitude wintering grounds and high-latitude feeding grounds in the north Pacific comes from the Structure of Populations, Levels of Abundance, and Status of Humpbacks (SPLASH) study (Calambokidis *et al.* 2008, Barlow *et al.* 2011, Baker *et al.* 2013). The SPLASH study was an international collaboration designed to collect photographs and tissue samples from all known wintering and feeding grounds in the north Pacific during the period from 2004 to 2006. The SPLASH data revealed complex population structure, with five or six wintering grounds and six feeding aggregations in the North Pacific. However, there is not a one-to-one correspondence between feeding and wintering populations. Rather, in most cases wintering grounds are shared by animals from multiple feeding aggregations and feeding aggregations are comprised of animals from multiple wintering grounds.

Animals that share the same feeding ground and wintering ground can be thought of as migratory herds, similar to herds in terrestrial migratory species. Members of a migratory herd face the same environmental conditions and threats throughout the entire year. Members of different herds, in contrast, face different conditions and threats through much of the year, which could result in differences in trends and vital rates. Even for herds that share the same feeding ground, for example the CentAm-CA/OR herd and the MMex-CA/OR herd, the proportional representation of those herds differs across the range of the feeding ground (Calambokidis, pers, comm.) and hence, threats in one area may affect one herd more than the other herd.

Though there are multiple published and ongoing genetic studies of north Pacific humpbacks that focus on feeding aggregations and/or wintering grounds, there have been no attempts to genetically characterize migratory herds. However, understanding the genetic relationships among herds is critical to appropriate management of north Pacific humpback whales. Evaluation of the SPLASH data resulted in the recognition of five Distinct Population Segment (DPS) in the wintering grounds, three of which are listed as Threatened or Endangered under the U.S. Endangered Species Act (Bettridge *et al.* 2015). Because most of the wintering grounds are outside of U.S. waters, proper management of these DPSs under the ESA requires apportioning abundance estimates generated on the feeding grounds to DPSs and determining whether threats encountered on feeding grounds are disproportionately affecting specific DPSs. This can most easily be accomplished by quantifying abundances of and impacts to individual herds and then combining data across herds from a given DPS.

Understanding the degree of genetic and demographic independence among herds that make up a particular DPS is also important for management and assessment of humpback whales under the U.S. Marine Mammal Protection Act (MMPA). Because demographic independence is the primary criterion for determining whether two groups should be managed as separate stocks under the MMPA (Angliss and Wade 1997, Moore and Merrick 2011), understanding the relationships among herds will inform upcoming discussions of how to designate stocks.

In this study, we focus on two questions that are central to understanding and managing humpback whales in the north Pacific. First, are the different migratory herds genetically differentiated from the other migratory herds that share a wintering ground? Second, can genetic mitochondrial genetic data be used to assign animals sampled or killed on feeding grounds to a migratory herd, and therefore to a specific wintering ground/DPS? We address these questions

in the context of the feeding aggregation that occurs off the coasts of California and Oregon (henceforth the CA/OR feeding aggregation).

The CA/OR feeding aggregation is comprised of two herds (Calambokidis *et al.* 2008). The MMex-CA/OR herd migrates between a wintering ground off of mainland Mexico (MMex) and the coast of CA/OR and is part of the Threatened Mexico DPS (Figure 1). The CentAm-CA/OR herd is the only herd that winters off the coast of Central America (CentAm), and therefore encompasses the entire endangered Central America DPS. Both herds migrate along the Baja Peninsula, and therefore have overlapping ranges for most of the year. Both herds are subject to high rates of fishing gear entanglement and ship strikes while feeding off of CA/OR (Rockwood *et al.* 2017, Carretta *et al.* 2019). Monitoring and management of these impacts requires both estimating the abundance of each herd and attributing entangled and struck whales to their herd of origin.

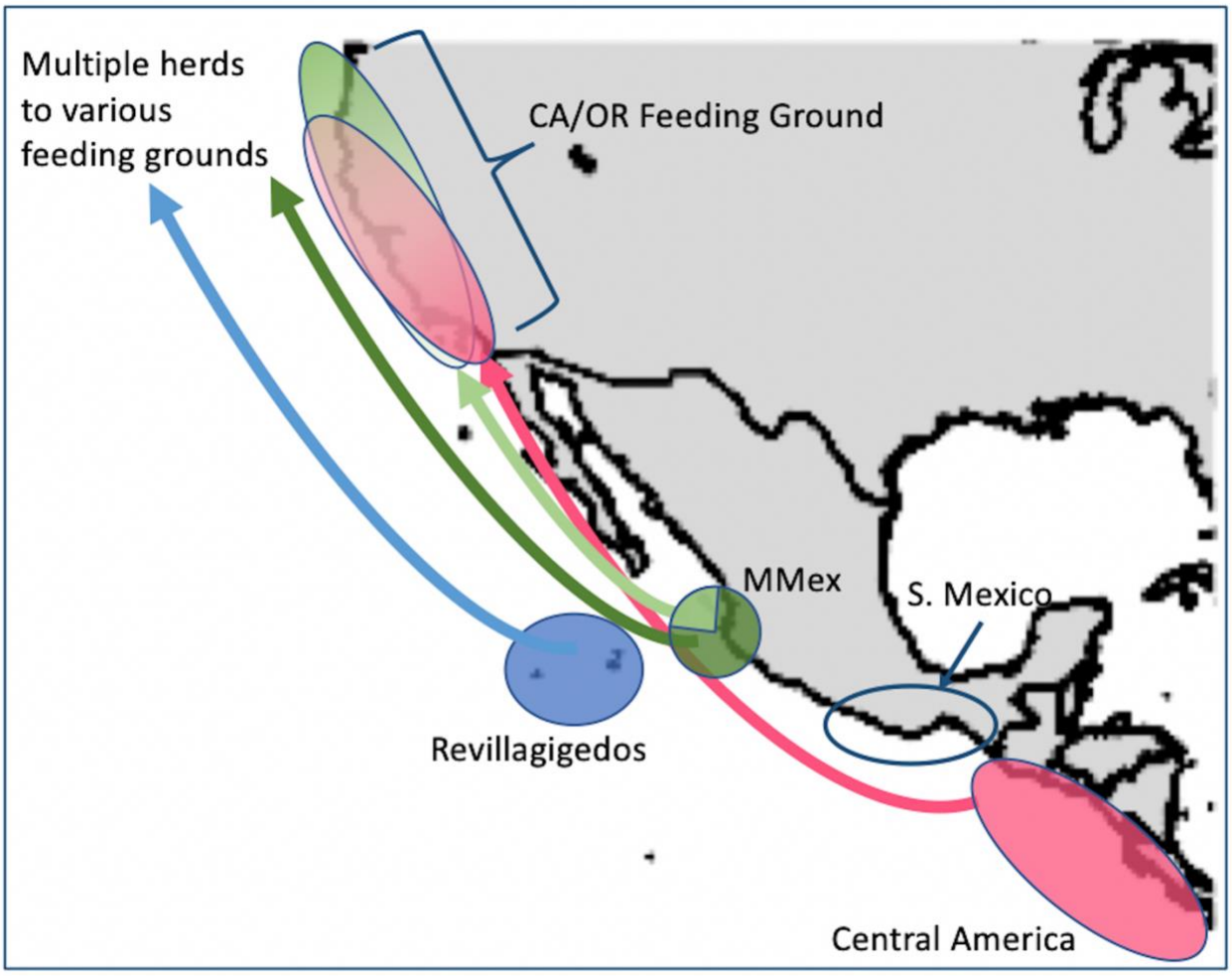

Figure 1. Three wintering grounds south of the U.S./Mexico border – Central America (red), mainland Mexico (MMex; green), and Revillagigedos (blue). The latter two comprise the Mexico DPS, while Central America is listed as a separate DPS. The DPS affiliation of animals off southern Mexico (S. Mexico) is unknown. All Central America animals migrate to the CA/OR feeding ground, as do a portion of the MMex animals. The remaining MMex animals and all Revillagigedos animals migrate to various feeding grounds further north (not shown).

Baker *et al.* {, 2013 #700} found strong frequency-based differences in the mitochondrial control region haplotype distributions of the Central America and mainland Mexico wintering grounds, as is expected between groups with maternally-driven site fidelity. Most common haplotypes were shared between wintering grounds, though some haplotypes that are common in mainland Mexico were not found in Central America. A traditional frequency-based assignment test indicated that individuals could be correctly assigned to the two wintering grounds with 68% to 79% accuracy based on their mitochondrial control region haplotypes. However, the maternally-inherited nature of mitochondrial DNA combined with the fact that humpback whales exhibit strong fidelity to the feeding ground of their mother mean that haplotype frequencies of a migratory herd may differ from the overall frequencies of the wintering ground from which the herd originates. Consequently, an assignment test based on mitochondrial DNA that uses wintering aggregations as source population references could either over- or under-estimate the assignability of individuals to wintering grounds/DPSs.

We use mitochondrial sequence data to compare the two CA/OR herds to each other and to compare the MMex-CA/OR herd to the mainland Mexico wintering aggregation as a whole. We sequenced the full mitogenomes of 156 individuals, 115 of which are known to be members of the MMex-CA/OR or CentAm-CA/OR herds based on photographic matches, and use the data to investigate the assignability of individuals sampled off of CA/OR to migratory herds. The increased phylogenetic resolution offered by full mitogenome sequencing could result in the separation of shared control region haplotypes into stratum-specific mitogenome haplotypes, as has been observed in previous studies of cetacean species (Morin *et al.* 2010, Archer *et al.* 2013). The resulting phylogeographic signal would increase assignment power. Furthermore, by restricting our source populations to only known members of the two herds that occur off of CA/OR rather than using the entire mainland Mexico wintering ground sample, we can obtain a more accurate estimate of assignability of unknown individuals sampled off of CA/OR to migratory herds.

We sequenced the mitochondrial control regions of 324 animals sampled off the coasts of California, Oregon, and Washington during the 2018 California Current Ecosystem Studies (CCES) survey. We compared the CCES data set to data from SPLASH samples (Baker *et al.* 2013) to look for temporal changes in haplotype frequencies in the CA/OR feeding aggregation. Finally, we compare the control region sequence haplotype frequencies from the MMex-CA/OR herd samples to published haplotype frequencies for the mainland Mexico wintering aggregation (Baker *et al.* 2013) in order to determine whether the herd is genetically differentiated from other mainland Mexico herds.

Materials and methods

Samples and stratification

All samples used in this study are housed in the NMFS Marine Mammal and Sea Turtle Research Collection (MMASTR) at the Southwest Fisheries Science Center (SWFSC). Samples were either fixed in salt-saturated 20% DMSO solution (Amos 1997) or ethanol and archived in a -20° C freezer or stored in a -80° C freezer with no preservative.

For the migratory herd data set, photographic identification data were used to select samples from animals that had been sighted off the U.S. west coast and had also been sighted off the coasts of Mexico or Central America. This approach allowed us to assign samples to two

migratory herds. The CentAm-CA/OR herd winters in Central America (Figure 1) and then migrates to the feeding aggregation off the coasts of California and Oregon (CA/OR). All animals that are part of the Central America DPS migrate to CA/OR and are therefore part of the CentAm-CA/OR herd. The MMex-CA/OR herd is comprised of animals that winter off of mainland Mexico (Figure 1) and migrate to the feeding aggregation off the coasts of CA/OR. Due to data indicating strong differences between mainland Mexico and the Revillagigedos Archipelago (Calambokidis *et al.* 2008, Baker *et al.* 2013), these two areas are considered separate wintering grounds, though they are currently part of the same DPS. Furthermore, animals that winter off Revillagigedos do not feed off CA/OR. The Revillagigedos animals are therefore not part of the MMex-CA/OR herd as defined in this study.

To assign animals from the CA/OR feeding aggregation to migratory herds, all sightings of those animals that occurred south of the U.S./Mexico border were stratified into four regions – Baja, mainland Mexico, southern Mexico, and Central America (from Guatemala south to Panama/Columbia border (Figure 1). Animals that winter in Central America must pass through Mexican waters during their migration. Consequently, it is possible that some animals sighted off of Mexico are actually part of the Central America DPS. We therefore considered any animal that had been sighted in Central America to be part of the CentAm-CA/OR herd, even if it had also been sighted in mainland Mexico.

All animals sighted in the mainland Mexico region (Figure 1) but never sighted south of 20.3° N latitude were considered to be part of the MMex-CA/OR herd. The DPS affiliation of animals sighted in southern Mexico between 20.3° N latitude and the Mexico/Guatemala border is uncertain, so animals sighted in that region but not within the range of the Central America DPS were not assigned to a herd. Similarly, the waters off of the Baja Peninsula are traversed by animals from the Central America DPS and the mainland Mexico and Revillagigedos populations of the Mexico DPS, and therefore were not assigned to a herd.

Previous studies have found that the F2 haplotype that is common in Central America is identical to haplotype SP88, which was identified in southern hemisphere humpback whales (Jackson *et al.* 2014). We therefore included nine samples from southern hemisphere animals that possess either haplotype SP88 (henceforth referred to as F2) or the closely-related haplotype SP90 in our data set in order to look for phylogeographic structure between the northern and southern hemisphere F-type animals.

In addition to the migratory herd data set comprised of CentAm-CA/OR and MMex-CA/OR herds samples, we generated control region sequence data for all humpback whale samples collected during the SWFSC's California Current Ecosystem Studies (CCES) shipboard survey, which took place between June and December, 2018 (Henry *et al.* 2020). We stratified the CCES samples geographically, with samples collected between the U.S./Mexico border and the Oregon/Washington border representing the CA/OR feeding aggregation.

Laboratory processing

Genomic DNA was extracted from the tissues using magnetic bead-based DNA extraction (NucleoMag®Tissue, Macherey-Nagel, Bethlehem, PA, USA), phenol/chloroform extraction (Sambrook *et al.* 1989), or sodium chloride protein precipitation (Miller *et al.* 1988). The sodium chloride precipitation protocol was sometimes modified to include additional proteinase K and overnight digestion at 37° C to assist with breakdown of the skin sample during the initial digestion.

All samples collected during the CCES survey were Sanger sequenced for a 390 bp portion at the 5' end of the hypervariable portion of the control region. In addition, we Sanger sequenced the control regions of all migratory herd samples for which control region sequences were not already available. We used standard Sanger-sequencing with primers TRO (5′-CCTCCCTAAGACTCAAGG-3′; developed at SWFSC) and D (5′-CCTGAAGTAAGAACCAGATG-3′; Rosel *et al.* 1994). We performed PCR reactions in 25ul volumes using 1 μL (approximately 5–25 ng) genomic DNA, 1× PCR buffer [10 mM Tris–HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$], 0.3 μM of each primer, 200 μM of each dNTP and 0.5 units of *Taq* DNA polymerase. The PCR thermal profile consisted of an initial denature at 94°C for 2.5 min, followed by 35 cycles of 94°C for 45 s, 1 min at 60°C annealing temperature, and 72°C for 1.5 min, then a final extension at 72°C for 5 min. Sequencing of the PCR product in both directions was performed using the ABI 3100 and 3130XL Automated Sequencers (Applied Biosystems Inc., Foster City, CA). All sequences were aligned using Geneious R8-10 software (Biomatters Ltd. Auckland, New Zealand). To test for errors in sequencing, a random 10% replication of all samples was completed. If a discrepancy was found, the sample was re-sequenced from extracted DNA. Unresolved discrepancies and all rare haplotypes were re-extracted from tissue and re-sequenced for haplotype confirmation.

Samples were genetically sexed using Real-Time PCR (Stratagene, La Jolla, CA) of the zinc finger (ZFX and ZFY) genes, as described in Morin *et al.* (2005).

Mitogenome library preparation closely followed a modified capture array method outlined in Hodges *et al.* (2009) with a few modifications as detailed in Hancock-Hanser *et al.* (2013) and in this document. Modifications to Hodges et al (2009) and Hancock-Hanser *et al.* (2013) include using dual indexing in the index PCR. Instead of using one i5 index for all the samples, several were used to increase the uniqueness of each sample. Prior to pooling, samples were quantified using a sybergreen qPCR method with Illumina sequencing primers (Sybergreen Universal Master Mix®, Biorad, Hercules, CA, USA). Although the above methods describe using a capture array to capture target fragments of the genome prior to sequencing, the library preparation used with these samples did not include hybridization to a capture array. Instead, "shotgun sequencing" (sequencing all fragments prepared in the library without regard to size or location in the genome) was used. The pool of product was gel size selected to isolate fragments in the range of ~250-500 base pairs prior to single-end sequencing on a NextSeq Illumina sequencer with a read length of 75 base pairs (Illumina, Inc, San Diego, CA, USA).

Mitogenome assembly

We generated consensus sequences from mitogenome FASTQ files using custom R scripts developed at SWFSC. The scripts utilize the Burrows-Wheeler alignment method (BWA-aln; Li and Durbin 2009) with a maximum edit distance of 4, seed length of 20, and indels disallowed within 5 base pairs of the end of a sequence read. All other parameters were left at their default settings. All mitogenomes were initially assembled to a published reference sequence (Genbank accession number AP006467; Sasaki *et al.* 2005) that we modified by adding 40 base pairs from each end of the sequence to the opposite end in order to avoid an artificial drop in coverage near the ends of the mitogenomes due to linearization of the sequence. However, approximately half of the samples possessed a 5 base pair deletion in the 16S region that resulted in poor alignment to the reference and large numbers of ambiguous bases. We therefore chose one sample (lab ID z0056398) that possessed the deletion but nonetheless assembled to the Genbank reference with no ambiguities due to its high sequencing coverage

(average coverage = 69.6 reads per base) and used it as a reference when assembling the remaining samples with the deletion. We again padded the ends of the z0056398 with 40 base pairs from the opposite ends when using it as a reference.

Once all of the reads for a sample were aligned, we generated a consensus sequence for the sample. Nucleotides were called if there were at least 5 reads and the nucleotide was called in greater than 75% of the reads, or if there were at least 3 reads with 100% agreement in their calls. If neither of these conditions was met, the nucleotide was called as an N. We did not incorporate any correction for 'index hopping' (Jun *et al.* 2012, Kircher *et al.* 2012, Van Cise *et al.* 2019) as our laboratory processing did not involve amplification of pooled product and sequencing was conducted on an Illumina NextSeq, which uses multiplexing chemistry that minimizes the likelihood of index hopping.

Data review

Initial inspection of the consensus sequences revealed a variable repeat region starting at position 1126 (within the 16S region) that consisted of 8 to 16 Cs, followed by 7 to 12 Ts, followed by the aforementioned 5 base pair insertion-deletion. We visually examined the BAM files for every sample in order to confirm the number of Cs and Ts in the repeats and confirm the presence or absence of the insertion. We were unable to reliably call the number of Cs in the repeat region, so that repeat was deleted from all consensus mitogenome sequences and replaced with a single N.

All Ns in all consensus sequences were checked by visual inspection of the BAM files and were manually called if they were due to obvious alignment errors. The control region of each consensus mitogenome sequence was extracted and compared to that sample's Sanger control region sequence and mismatches were resolved by visual inspection of the BAM file and the Sanger trace file as necessary. In cases where the control region from the mitogenome consensus sequence contained an N due to low coverage, the N was replaced by the nucleotide from the Sanger sequence.

Data analysis

We estimated the degree of genetic differentiation between strata using Φ_{ST} and evaluated its statistical significance using a permutation test with 10,000 permutations. We chose Φ_{ST} because it accounts for within-population diversity and produces an unbiased estimate of differentiation. F_{ST}, in contrast, exhibits a known downward bias and reduced statistical power when within-population diversity is high (Meirmans and Hedrick 2011), as is the case in our data set. When comparing the two migratory herds, we calculated all statistics using both the full mitogenome sequences and control region sequences. For all other comparisons we only used control region sequences since full mitogenomes were not generated for the CCES or SPLASH samples. We used the *strataG* package (Archer 2016) in R v3.6.1 (Team 2019) for all genetic differentiation analyses.

We used the ensemble-based classification method Random Forests to develop a classification model that could assign samples to migratory herds. Random Forests is an extension of the CART approach in which multiple classification trees (the forest) are produced through stochastic sub-selection of the data set. We developed the classification model using the herd data set, for which the herd assignment of individual samples was known. The model was internally validated because each tree in the forest was grown using only a subset of the samples, with the remaining 'out-of-bag' (OOB) samples used to validate that tree. We largely followed

the methods of Archer *et al*. (2017), who were the first to apply Random Forests to sequence data. We ran Random Forest in R v3.6.1 (Team 2019) using the *randomforest* package (Liaw and Wiener 2002) and custom scripts. In all analyses we built 10,000 trees, which resulted in stable estimates of assignment error. Each tree was built with the sample sizes from both herds equal to 25 (half the sample size of the smaller stratum, the MMex-CA/OR herd). All other variables were left at default values.

The one way in which our implementation of Random Forests differed from that of Archer *et al*. (2017) was in the choice of sites to use as variables. Archer *et al*. included all variable sites in the analysis, except that they excluded sites that were variable due to a substitution in a single individual. We further restricted the sites used in the analysis by calculating the correlation coefficient between variable sites. Sites that are perfectly correlated with each other (correlation coefficient equals 1.0) do not contribute unique information to the analysis. We therefore used custom R scripts to identify groups of sites that were all perfectly correlated and excluded all but one site from each group from the Random Forest analysis. For each such group we retained the highest position number, ensuring that the maximum number of control region sites were retained.

The assignment algorithm produced by Random Forests is predicated on the assumption that the samples used to develop the classification model are correctly stratified by herd. However, because Central America animals migrate past mainland Mexico, there is potential for CentAm-CA/OR animals to be photographed off of mainland Mexico during migration and incorrectly included in the MMex-CA/OR herd. In order to determine the assignment accuracy that could be achievable if there were no assignment errors in our MMex-CA/OR sample, we conducted a simulation that utilized the control region haplotype frequencies from the CentAm-CA/OR herd and the CCES samples to estimate the control region haplotype frequencies expected in the MMex-CA/OR herd. For each iteration of the simulation, we assumed that individuals from the CentAm-CA/OR herd comprised *p* percent of the CA/OR feeding aggregation. We then calculated the expected haplotype frequencies of the MMex-CA/OR herd by subtracting the weighted control region haplotype frequencies of the CentAm-CA/OR herd from the observed haplotype frequencies of the CCES samples. We used the resulting expected frequencies to generate a simulated MMex-CA/OR sample, which we refer to as the SimMex sample. We expanded the SimMex data set to full mitogenome haplotypes using the frequencies of mitogenome haplotypes relative to their corresponding control region haplotypes within the empirical MMex-CA/OR herd sample. We re-ran the RandomForest analysis using the SimMex sample and the empirical CentAm-CA/OR herd sample and recorded the assignment accuracy. We conducted 1,000 simulations for each value of *p* and examined five values of *p* ranging from 0.1 and 0.5.

Results

Genetic diversity

Migratory herd data set

We generated mitogenome assemblies for 156 samples known to feed off of the U.S. west coast and nine samples from the southern hemisphere. Nine samples were excluded from the data set due to low coverage resulting in more than 1,600 Ns in their final assemblies. Three samples were excluded because they showed evidence of contamination, and two samples were excluded due to uncertain provenance. After exclusions, our final data set included 151 samples,

nine of which were collected off of Washington state and nine of which were from the southern hemisphere. The remaining 133 samples are from animals known to be part of the CA/OR feeding aggregation, 115 of which could be assigned to either the MMex-CA/OR or CentAm-CA/OR herds (Table 1).

Table 1. Control region haplotype frequencies from the migratory herd data set. Samples in the Baja, MMex-CA/OR, Southern Mexico, and CentAm-CA/OR strata have been photographically identified as being part of the CA/OR feeding aggregation, but the wintering ground affiliation of the Baja and Southern Mexico samples is unknown. Haplotype designations match those from Baker et al. (2013) and Jackson et al. (2014), with the exceptions of E4.5, H35, H36, and H37, all of which are newly-identified haplotypes. Haplotypes that are shared between the MMex-CA/OR and CentAm-CA/OR herds are bolded.

Control region haplotype	**Washington State**	**Baja**	**MMex-CA/OR herd**	**Southern Mexico**	**CentAm-CA/OR herd**	**Southern Hemisphere**
A3	0	0	2	1	0	0
A-	1	0	0	0	0	0
A+	3	1	3	1	0	0
E1	2	0	**10**	3	**15**	0
E10	0	1	**3**	0	**1**	0
E13	0	0	**3**	1	**4**	0
E2	0	0	1	0	0	0
E3	0	0	1	0	0	0
E4	1	2	**2**	0	**3**	0
E4.5	1	0	2	0	0	0
E5	0	0	**1**	1	**2**	0
E6	0	0	**2**	1	**1**	0
E7	1	0	1	0	0	0
F1	0	0	**2**	0	**1**	0
F2	0	0	**14**	5	**25**	4
F3	0	0	0	0	7	0
F4	0	0	1	0	0	0
F6	0	1	**1**	0	**2**	0
F7	0	0	1	0	0	0
H35	0	0	0	0	2	0
H36	0	0	0	0	1	0
H37	0	0	0	0	0	1
SP90	0	0	0	0	1	4
Total	**9**	**5**	**50**	**13**	**65**	**9**

We identified 23 control region haplotypes among the samples (Table 1), 19 of which are already known from published studies (Baker *et al.* 2013, Jackson *et al.* 2014). Three of the newly-identified control region haplotypes (H35, H36, H37) are most similar to haplotype F1, differing from it by 2, 5, and 7 substitutions, respectively. The fourth new haplotype (E4.5) was

assigned to individuals who are heteroplasmic for the single substitution that distinguishes haplotypes E4 and E5.

Eight out of 23 control region haplotypes (35%) were shared between the MMex-CA/OR and CentAm-CA/OR herds. These included the two most common haplotypes in each herd (E1 and F2), and therefore represented 72% and 82% of the individuals in the MMex-CA/OR and CentAm-CA/OR herds, respectively

We identified 45 unique mitogenome haplotypes among the samples (Table 2). While some of the most common control region haplotypes (e.g., E1 and F2) were split into multiple haplotypes when the full mitogenome sequence is considered, most control region haplotypes were only represented by one or two full mitogenome haplotypes. Even in cases where shared control region haplotypes were split into multiple mitogenome haplotypes, many of the resulting mitogenome haplotypes were still shared between herds. In the mitogenome data set, 66% of the individuals in the MMex-CA/OR herd and 62% of the individuals in the CentAm-CA/OR herd possess a mitogenome haplotype that is shared with the other herd.

Table 2. Full mitogenome haplotype designations and frequencies from the migratory herd data set. Samples in the Baja, MMex-CA/OR, Southern Mexico, and CentAm-CA/OR strata have been photographically identified as being part of the CA/OR feeding aggregation, but the wintering ground affiliation of the Baja and Southern Mexico samples is unknown. The association between control region and mitogenome haplotypes is given in the first two columns. Haplotypes that are shared between the MMex-CA/OR and CentAm-CA/OR herds are bolded.

Control region haplotype	Mitogenome haplotype	Washington State	Baja	MMex-CA/OR	Southern Mexico	CentAm-CA/OR	Southern Hemisphere
A3	MnMito01	0	0	2	1	0	0
A-	MnMito02	1	0	0	0	0	0
A+	MnMito03	2	0	2	1	0	0
	MnMito04	1	1	0	0	0	0
	MnMito06	0	0	1	0	0	0
E1	**MnMito08**	0	0	**7**	1	**5**	0
	MnMito09	0	0	0	1	4	0
	MnMito10	0	0	0	0	2	0
	MnMito12	0	0	0	0	1	0
	MnMito13N	0	0	**2**	0	**2**	0
	MnMito14	2	0	**1**	1	**1**	0
E2	MnMito20N	0	0	1	0	0	0
E3	MnMito21N	0	0	1	0	0	0
E4	**MnMito22**	1	2	**2**	0	**3**	0
E45	MnMito51	1	0	2	0	0	0
E5	**MnMito23**	0	0	**1**	1	**1**	0
	MnMito24	0	0	0	0	1	0
E6	**MnMito25**	0	0	**2**	1	**1**	0
E7	MnMito26	0	0	1	0	0	0
	MnMito27	1	0	0	0	0	0
E10	**MnMito15**	0	1	**3**	0	**1**	0
E13	MnMito16	0	0	0	1	4	0
	MnMito17	0	0	2	0	0	0
	z0068471	0	0	1	0	0	0
F1	**MnMito28N**	0	0	**2**	0	**1**	0
F2	MnMito18	0	0	0	0	0	2
	MnMito29	0	0	**7**	5	**16**	0
	MnMito32	0	0	**3**	0	**6**	0
	MnMito33	0	0	1	0	0	0
	MnMito34	0	0	**2**	0	**1**	0
	MnMito35	0	0	0	0	1	0
	MnMito36	0	0	0	0	0	1
	MnMito37	0	0	0	0	0	1
	MnMito38N	0	0	1	0	0	0
	MnMito39N	0	0	0	0	1	0
F3	MnMito41	0	0	0	0	7	0
F4	MnMito42	0	0	1	0	0	0
F6	**MnMito43**	0	1	**1**	0	**2**	0
F7	MnMito44	0	0	1	0	0	0
H35	MnMito45N	0	0	0	0	2	0
H36	MnMito46	0	0	0	0	1	0
H37	MnMito47	0	0	0	0	0	1
SP90	MnMito48	0	0	0	0	1	2
	MnMito49	0	0	0	0	0	1
	MnMito50N	0	0	0	0	0	1
Total		**9**	**5**	**50**	**13**	**65**	**9**

CCES data set

We Sanger sequenced 293 samples collected during the CCES survey and identified 20 haplotypes, 19 of which are also found in the migratory herd data set (Table 3). The new haplotype (E6.14) is assigned to an individual who is heteroplasmic for the single substitution that distinguishes haplotypes E6 and E14.

Table 3. Haplotype frequencies from samples collected during CCES survey. Samples are stratified geographically, with samples collected south of the Oregon/Washington border considered part of the CA/OR feeding aggregation. Haplotype frequencies from two historical data sets from CA/OR and from the mainland Mexico wintering aggregation are from Baker et al. (2013).

Control region haplotype	**CCES 2018**		**Historical data**		**Mainland Mexico Wintering Aggregation (n=62)**
	Washington State (n=66)	**CA/OR (n=227)**	**2004 CA/OR (n=123)**	**1988/9 CA/OR (n=49)**	
A3	0	0.004	0.04	0	0.03
A-	0.30	0.04	0.02	0	0.06
A+	0.41	0.12	0.07	0.08	0.24
E1/E11*	0.03	0.26	0.20	0.22	0.15
E10	0.02	0.02	0.02	0.04	0.03
E13	0	0.04	0.04	0.04	0.02
E15	0	0.01	0	0	0
E2	0.05	0.004	0	0	0
E3	0	0.01	0.02	0.02	0.03
E4	0.09	0.01	0.08	0	0.03
E4.5	0.02	0	0	0	0
E5	0	0.03	0.02	0.04	0.08
E6	0	0.03	0.02	0.06	0.08
E6.14	0	0.004	0	0	0
E7	0.02	0.05	0.02	0	0.03
E8	0	0.004	0	0	0
F1	0.02	0.02	0.04	0.04	0
F2	0.06	0.32	0.35	0.35	0.13
F3	0	0.02	0.03	0.04	0
F4	0	0	0.01	0	0.03
F5	0	0	0	0	0.02
F6	0	0.01	0.02	0.06	0.02
F7	0	0	0	0	0.02
F8	0	0	0.01	0	0

*Haplotypes E1 and E11 are distinguished by a single substitution that is outside the sequence fragmented generated by the primers we used (standard SWFSC primers). Thus, we cannot distinguish between these two haplotypes in our data set.

If the herd data sets were representative of the CA/OR feeding aggregation, the haplotype frequencies of the feeding aggregation should all be intermediate between those of the two herds, which is not the case. For instance, the combined frequency of the A+, A-, and A3 haplotypes in the CCES data set (16.7%; Table 3) is substantially higher than their combined frequency in either the MMex-CA/OR herd (10%) or CentAm-CA/OR herd (0%; Table 1). In the 2004

SPLASH data set, the combined frequencies of the A-type haplotypes is 12.2% - lower than in the CCES data set, but still higher than either of the herd data sets (Table 3). Only the 1988/89 CA/OR data set has a combined frequency of A-type haplotypes (8.2%; Table 3) that is in between those of the MMex-CA/OR and CentAm-CA/OR herds.

Genetic Differentiation

The MMex-CA/OR herd and CentAm-CA/OR herd are significantly genetically differentiated from each other in both their control region sequences and full mitogenomes (Table 4). The MMex-CA/OR herd is also significantly differentiated from the mainland Mexico wintering aggregation, of which it is a component (Table 4).

All of the samples in our CCES data set were collected in 2018. We compared our CCES data set to published data from the CA/OR feeding aggregation for samples collected in 1988/1989 and 2004 and found no significant genetic differentiation the data sets (Table 4).

Table 4. Pairwise estimates of genetic differentiation between strata. All tests are based on mitochondrial sequence data. 'Sequence type' indicates whether full mitogenomes or only control region sequence was used. P-values are based on 10,000 permutations. Haplotype frequencies for the mainland Mexico (MMex) wintering ground and 2004 and 1988-89 California/Oregon (CA/OR) feeding ground are taken from Baker et al. (2013).

Strata	Sequence type	Φ_{ST}	*p*-value
Herd comparisons			
CentAm-CA/OR (n=65) v. MMex-CA/OR (n=50)	Mitogenome	0.054	**0.024**
CentAm-CA/OR (n=65) v. MMex-CA/OR (n=50)	Control region	0.044	**0.035**
Herd vs. wintering aggregation comparison			
MMex-CA/OR (n=50) v. MMex wintering ground (n=62)	Control region	0.044	**0.039**
CentAm-CA/OR (n=65) v. MMex wintering ground (n=62)	Control region	0.192	**0.001**
CA/OR feeding ground temporal comparisons			
2018 (n=227) v. 1988-9 (n=49)	Control region	0.015	0.130
2018 (n=227) v. 2004 (n=123)	Control region	0.009	0.084
1988-9 (n=49) v. 2004 (n=123)	Control region	-0.010	0.745

RandomForest

Correct assignment (1 - OOB error rate) from our Random Forest analysis were 64.6% for the CentAm-CA/OR herd and 42.0% for the MMex-CA/OR herd. Assignment accuracy varied substantially across mitochondrial control region haplotypes (Table 5). In most cases,

haplotypes that were only detected in one herd had high assignment accuracy (e.g., haplotypes A3, A+, and E2 in the MMex-CA/OR herd, haplotypes F3, H35, and SP90 in the CentAm-CA/OR herd). Exceptions to this pattern were haplotypes that were represented by a single individual (haplotypes E3, E7, F4, F7, and H36). Among haplotypes that were shared by the herds, most had low assignment accuracy in both herds. However, individuals with haplotypes E10 and E6 all assigned most strongly to MMex-CA/OR, resulting in very low accuracy for CentAm-CA/OR. Similarly, all individuals with haplotypes F2 and F6 assigned to CentAm-CA/OR herd, including individuals with those haplotypes that were actually part of the MMex-CA/OR herd.

Assignment accuracy was higher in the simulations designed to examine performance in the absence of mis-stratification errors (Table 6). In the simulations, accuracy steadily increased as the proportional representation of CentAm-CA/OR in the CA/OR feeding aggregation increased.

Table 5. Assignment accuracy for the Random Forest classification model as a function of control region haplotype. Accuracy is calculated as one minus the out-of-bag (OOB) error rate. 'na' indicates that there were no samples in the specified herd that possessed the haplotype.

CR haplotype	Mmex	CentAm
A3	0.854	na
Aplus	0.810	na
E1	0.479	0.410
E10	0.707	0.063
E13	0.321	0.430
E2	0.788	na
E3	0.591	na
E4	0.298	0.443
E45	0.934	na
E5	0.351	0.374
E6	0.735	0.050
E7	0.414	na
F1	0.372	0.380
F2	0.151	0.772
F3	na	0.984
F4	0.192	na
F6	0.076	0.695
F7	0.260	na
H35	na	0.751
H36	na	0.593
SP90	na	0.762

Table 6. Average OOB error rates from Random Forest when using the empirical CentAm-CA/OR herd data set and simulated MMex-CA/OR data set. Each error rate is averaged over 1,000 simulated data sets. Error rates steadily decline as the proportional representation of CentAm animals assumed in the simulation increases. Average error rates for all simulated data sets were lower than that obtained using the empirical MMex-CA/OR herd data set (bottom row).

Proportion of CentAm herd	Average Assignment Accuracy	
	CentAm-CA/OR	MMex-CA/OR
10	0.670	0.589
20	0.674	0.625
30	0.688	0.651
40	0.698	0.688
50	0.706	0.710
Empirical	0.646	0.420

Discussion

Our data revealed significant genetic differentiation between herds that share a feeding ground but utilize different wintering grounds, as well as between a herd and the wintering aggregation of which it is a part. The magnitude of differentiation between the MMex-CA/OR herd and its wintering aggregation ($\Phi_{ST} = 0.044$) is exactly the same as the magnitude of differentiation between the two CA/OR herds (Table 4), indicating that, at least in this case, a migratory herd is no more genetically similar to other herds with which it shares a wintering ground than it is to a herd from a different wintering ground. This result confirms that the strong maternally-inherited feeding ground fidelity observed in north Pacific humpback whales results in genetic differentiation among herds that share a wintering ground but migrate to different feeding grounds. It also indicates that haplotype frequencies from wintering aggregations cannot be used as proxies for the haplotype frequencies of individual herds.

The magnitude of differentiation between the MMex-CA/OR and CentAm-CA/OR herds ($\Phi_{ST} = 0.044$) is half that between the Central America wintering ground and the mainland Mexico wintering ground ($\Phi_{ST} = 0.087$)(Baker *et al.* 2013). The reduced differentiation between the herds could be the result of higher dispersal between herds that share a feeding ground than between wintering grounds overall. This would occur if, for instance, animals show higher fidelity to feeding grounds than they do to wintering grounds. However, the apparent similarity of the two CA/OR herds compared to their wintering grounds could also result from chance through random genetic drift. Data from nuclear markers would be useful in distinguishing between these two possibilities.

Our mitogenome data set revealed a surprising lack of phylogeographic signal within the two migratory herds we examined. Though sequencing the full mitogenome did result in some control region haplotypes resolving into multiple mitogenome haplotypes, the proportion of individuals that possessed a haplotype that is shared between herds remained high in both data sets. This result contrasts with previous studies of large whales (e.g., Morin *et al.* 2010, Archer *et al.* 2013).

The haploid, non-recombining nature of mitochondrial DNA generally precludes its use in frequency-based assignment tests due to the low statistical power associated with assigning

based on a single locus. If two strata have no shared haplotypes, assignments can be made with certainty, though individuals with novel haplotypes still cannot be assigned. Random Forests provide an alternative to frequency-based methods. Because Random Forests operates at the level of individual nucleotides, it makes use of the phylogenetic signal that results from mutations that occur within a given stratum and are therefore not shared with other strata. However, the lack of phylogeographically concordant substitutions between the two migratory herds means that in this case, Random Forests does not perform better than a traditional frequency-based assignment algorithm.

Traditional frequency-based assignment methods can be effective for assigning unknown samples to breeding population using nuclear genetic data. The ability to combine data across multiple unlinked loci increases statistical power compared to use of a single locus, such as mtDNA, with overall power directly related to the number of loci. Future efforts should therefore examine the power of a large panel of nuclear markers to assign individuals to herds. Published estimates of nuclear genetic differentiation between the Central America and mainland Mexico wintering grounds are very low (F_{ST} = 0.0026)(Baker *et al.* 2013), suggesting high assignment accuracy may be difficult to achieve, that even with nuclear markers. Power analysis will be valuable in determining the expected assignment accuracy as a function of the number of loci, number of samples, and possible magnitude of differentiation between the herds.

The low frequency of A-type animals in our MMex-CA/OR herd compared to the CCES samples clearly indicates that the samples in our mitogenome data set are not representative of the animals that currently feed off of CA/OR. There are several possible explanations for this discrepancy. First, the high frequency of A-type animals off of CA/OR could result from changes in the genetic composition of the CA/OR feeding aggregation over time. While the CCES samples were all collected in a single year (2018), the MMex-CA/OR herd samples were collected over a span of 19 years, from 1998 to 2016, though nearly three-quarters of them were collected during 2004 to 2006, the years of the SPLASH study. However, given our finding of no significant genetic differentiation between the CCES samples and the SPLASH samples from CA/OR, it is unlikely that the discrepancy between the CCES haplotype frequencies and the herd haplotype frequencies is due to changing haplotype composition over time.

It is possible that some of the animals in our MMex-CA/OR herd data set are actually members of the CentAm-CA/OR that were photographed as they migrated past mainland Mexico. The resulting mis-stratification would obscure the signal of differentiation that Random Forests looks for, leading to lower assignment accuracy. A second possibility is that the identification of members of the MMex-CA/OR herd is somehow biased. For instance, temporal or spatial segregation of individuals within the herd during the migration or on the wintering grounds could result in part of the herd being missed during photo-identification efforts. In order for such a bias to lead to the discrepancy we found in the frequency of A-type individuals, the temporal or spatial segregation would have to correlate with haplotype.

The discrepancy in haplotype frequencies between the CA/OR feeding aggregation data sets and our herd data sets has implications for all future work that relies on photographic assignment of individuals to wintering grounds, whether that work is aimed at stratifying genetic sample sets, estimating abundance, or apportioning takes. Mis-stratification of Central America animals as being from mainland Mexico or spatial/temporal segregation of mainland Mexico animals would both result in a bias in any analysis that relies on photographic assignment of CA/OR individuals to a wintering ground. Consequently, additional work is urgently needed to better understand this discrepancy and address any bias that may result from photographic

assignment of individuals to herds. Analysis of nuclear data from all of the migratory herds from the U.S. west coast feeding aggregations and the Central America, Mexico, and Hawai‘i DPSs could be useful in identifying individuals that are mis-stratified using photo-identification.

Acknowledgments

We would like to thank Vicki Pease, Kelly Robertson, and Gabriela Serra-Valente for their assistance with sample curation and laboratory analyses, and Aimee Lang and Phil Morin for their input into project planning. Alex Curtis, Jim Carretta, Jeff Moore, and Jay Barlow provided valuable insights that greatly aided the analysis and interpretation of our data. Funding for this project was provided by the National Marine Fisheries West Coast Regional Office and Southwest Fisheries Science Center. This project would not have been possible without the hard work and dedication of all of the army of people in the SPLASH and CCES projects.

References cited

Amos, W. 1997. Marine mammal tissue sample collection and preservation for genetic analyses. Pages 107-113 *in* A.E. Dizon, S.J. Chivers and W.F. Perrin, eds. Molecular Genetics of Marine Mammals. Society of Marine Mammalogy, Lawrence, KS.

Angliss, R. and P.R. Wade. 1997. Guidelines for assessing marine mammal stocks: report of the GAMMS Workshop, April 3-5, 1996, Seattle, WA.

Archer, E. 2016. strataG: summaries and population structure analyses of haplotypic and genotypic data. R package version 0.9.4.

Archer, F.I., K.K. Martien and B.L. Taylor. 2017. Diagnosability of mtDNA with Random Forests: Using sequence data to delimit subspecies. Marine Mammal Science.

Archer, F.I., P.A. Morin, B.L. Hancock-Hanser, *et al.* 2013. Mitogenomic phylogenetics of fin whales (Balaenoptera physalus spp.): genetic evidence for revision of subspecies. PLoS One 8:e63396.

Baker, C.S., D. Steel, J. Calambokidis, *et al.* 2013. Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. Marine Ecology Progress Series 494:291-306.

Barlow, J., J. Calambokidis, E.A. Falcone, *et al.* 2011. Humpback whale abundance in the North Pacific estimated by photographic capture-recapture with bias correction from simulation studies. Marine Mammal Science 27:793-818.

Bettridge, S., C.S. Baker, J. Barlow, *et al.* 2015. Status review of the humpback whale (*Megaptera novaeangliae*) under the Endangered Species Act. NOAA Technical Memorandum NMFS-SWFSC-540. 263 pp.

Calambokidis, J., E.A. Falcone, T.J. Quinn, *et al.* 2008. SPLASH: Structure of populations, levels of abundance and status of humpback whales in the north Pacific. Cascadia Research. Final report for contract AB133F-03-RP-00078. 57 pp.

Carretta, J.V., K.A. Forney, E.M. Oleson, *et al.* 2019. U.S. Pacific marine mammal stock assessments: 2018. NOAA Technical Memorandum NMFS-SWFSC-617. 382 pp.

Hancock-Hanser, B.L., A. Frey, M.S. Leslie, P.H. Dutton, E.I. Archer and P.A. Morin. 2013. Targeted multiplex next-generation sequencing: advances in techniques of mitochondrial and nuclear sequencing for population genomics. Molecular Ecology Resources 13:254-268.

Henry, A., J. Moore, J.V. Carretta, *et al.* 2020. Report on the California Current Ecosystem Survey (CCES): cetacean and seabird data collection efforts June 26 - December 4, 2018. Background document PSRG-2020-B09 submitted to the Pacific Scientific Review Group, March 2020. 38 pp.

Hodges, E., M. Rooks, Z. Xuan, *et al.* 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature Protocols 4:960-974.

Jackson, J.A., D.J. Steel, P. Beerli, *et al.* 2014. Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). Proceedings of the Royal Society B: Biological Sciences 281:20133222-20133222.

Jun, G., M. Flickinger, K.N. Hetrick, *et al.* 2012. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. American Journal of Human Genetics 91:839-848.

Kircher, M., S. Sawyer and M. Meyer. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40:e3.

Li, H. and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transformation. Bioinformatics 25:1754-1760.

Liaw, A. and M. Wiener. 2002. Classification and regression by randomForest. R News 2:18-22.

Meirmans, P.G. and P.W. Hedrick. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5-18.

Miller, S.A., D.D. Dykes and H.F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Research 16:1215.

Moore, J.E. and R. Merrick. 2011. Guidelines for Assessing Marine Mammal Stocks: Report of the GAMMS III Workshop, February 15-18, 2011, La Jolla, California. NOAA Technical Memorandum NMFS-OPR-47. 107 pp.

Morin, P.A., F.I. Archer, A.D. Foote, *et al.* 2010. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. Genome Res 20:908-916.

Morin, P.A., A. Nestler, N.T. Rubio-Cisneros, K.M. Robertson and S.L. Mesnick. 2005. Interfamilial characterization of a region of the ZFX and ZFY genes facilitates sex determination in cetaceans and other mammals. Molecular Ecology 14:3275-3286.

Rockwood, R.C., J. Calambokidis and J. Jahncke. 2017. High mortality of blue, humpback and fin whales from modeling of vessel collisions on the U.S. West Coast suggests population impacts and insufficient protection. PLoS One 12:e0183052.

Rosel, P.E., A.E. Dizon and J.E. Heyning. 1994. Genetic analysis of sympatric morphotypes of common dolphins (genus *Delphinus*). Marine Biology 119:159-167.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sasaki, T., M. Nikaido, H. Hamilton, *et al.* 2005. Mitochondrial phylogenetics and evolution of mysticete whales. Systematic Biology 54:77-90.

Team, R.D.C. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Van Cise, A.M., R.W. Baird, C.S. Baker, *et al.* 2019. Oceanographic barriers, divergence, and admixture: Phylogeography and taxonomy of two putative subspecies of short-finned pilot whale. Molecular Ecology 28:2886-2902.