Demography or selection on linked cultural traits or genes? Investigating the driver of low mtDNA diversity in the sperm whale using complementary mitochondrial and nuclear genome analyses

Phillip A. Morin[1], Andrew D. Foote[2], Charles Scott Baker[3,4], Brittany L. Hancock-Hanser[1], Kristin Kaschner[5], Bruce R. Mate[3,4], Sarah L. Mesnick[1], Victoria L. Pease[1], Patricia E. Rosel[6], Alana Alexander[7]

[1]Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, California
[2]Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Bangor University, Bangor, Gwynedd, UK
[3]Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, Newport, Oregon
[4]Department of Fisheries and Wildlife, College of Agricultural Sciences, Corvallis, Oregon
[5]Department of Biometry and Environmental System Analysis, Albert-Ludwigs-University of Freiburg, Freiburg, Germany
[6]Southeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Lafayette, Louisiana
[7]Biodiversity Institute, University of Kansas, Lawrence, Kansas

Correspondence
Phillip A. Morin, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA.
Email: phillip.morin@noaa.gov

Mitochondrial DNA has been heavily utilized in phylogeography studies for several decades. However, underlying patterns of demography and phylogeography may be misrepresented due to coalescence stochasticity, selection, variation in mutation rates and cultural hitchhiking (linkage of genetic variation to culturally-transmitted traits affecting fitness). Cultural hitchhiking has been suggested as an explanation for low genetic diversity in species with strong social structures, counteracting even high mobility, abundance and limited barriers to dispersal. One such species is the sperm whale, which shows very limited phylogeographic structure and low mtDNA diversity despite a worldwide distribution and large population. Here, we use analyses of 175 globally distributed mitogenomes and three nuclear genomes to evaluate hypotheses of a population bottleneck/expansion vs. a selective sweep due to cultural hitchhiking or selection on mtDNA as the mechanism contributing to low worldwide

mitochondrial diversity in sperm whales. In contrast to mtDNA control region (CR) data, mitogenome haplotypes are largely ocean-specific, with only one of 80 shared between the Atlantic and Pacific. Demographic analyses of nuclear genomes suggest low mtDNA diversity is consistent with a global reduction in population size that ended approximately 125,000 years ago, correlated with the Eemian interglacial. Phylogeographic analysis suggests that extant sperm whales descend from maternal lineages endemic to the Pacific during the period of reduced abundance and have subsequently colonized the Atlantic several times. Results highlight the apparent impact of past climate change, and suggest selection and hitchhiking are not the sole processes responsible for low mtDNA diversity in this highly social species.

Keywords
cetacean, pairwise sequentially Markovian coalescent, Pleistocene, population structure

Introduction

There is a growing body of literature indicating that the commonly held tenet of population genetics, that genetic diversity should be correlated with population size, is often violated, particularly for mitochondrial DNA (mtDNA) (Bazin, Glemin, & Galtier, 2006). When mtDNA diversity is found to be low in currently abundant populations, several hypotheses have been invoked to explain the discord, including population bottlenecks (e.g., Hoelzel et al., 2002; O'Brien, 1994), selection on mtDNA (e.g., Finch, Zhao, Korkin, Frederick, & Eggert, 2014; Foote et al., 2011), variation in mutation rates (Lyrholm, Leimar, & Gyllensten, 1996) and cultural hitchhiking (linkage of genetic variation to culturally-transmitted traits affecting fitness; Kopps et al., 2014; Premo & Hublin, 2009; Whitehead, 1998, 2005).

Among globally-distributed large whales, most baleen whales exhibit high mtDNA diversity relative to toothed whales and are divided into multiple subspecies and genetically distinct populations (e.g., fin, humpback, grey and blue whales; Archer et al., 2013; Baker et al., 2013; Jackson et al., 2014; Lang et al., 2014; Leduc et al., 2016). Among toothed whales, however, unusually low mitochondrial DNA (mtDNA) diversity in some of the social odontocetes (e.g., sperm, pilot, killer and false-killer whales; Alexander et al., 2016, 2013; Hoelzel et al., 2002; Martien et al., 2014; Van Cise et al., 2016) has limited power to infer population structure, phylogeography and historical demography using traditional genetic tools. The sperm whale

(*Physeter macrocephalus*) is particularly enigmatic in this respect, as it is one of the most cosmopolitan and abundant of the large odontocetes, and known to move over large ranges of up to thousands of kilometres over annual or longer time periods (Mizroch & Rice, 2013; Steiner et al., 2012; Straley et al., 2014), yet it exhibits low mtDNA diversity and evidence of female philopatry (Alexander et al., 2016; Lyrholm & Gyllensten, 1998; Lyrholm, Leimar, Johanneson, & Gyllensten, 1999; Mesnick et al., 2011).

In addition to the sperm whale's broad distribution throughout the world's oceans, it is considered relatively abundant even post industrial whaling, with the global population size estimated in 2002 at 360,000, approximately one-third of its prewhaling population size (Whitehead, 2002). Previous genetic studies based on mtDNA control region (CR) sequences showed haplotype frequency differences among oceans, major within-ocean geographic regions and female-led social groups, but limited phylogeographic signal, largely due to the low CR diversity and occurrence of three common haplotypes in all large ocean basins (Alexander et al., 2016; Engelhaupt et al., 2009; Lyrholm & Gyllensten, 1998; Mesnick et al., 2011; Whitehead, 1998).

In other species with low levels of CR diversity, complete mitochondrial genome (mitogenome) sequences have been used to detect phylogeographic structure and estimate divergence times and historical demography with higher precision (e.g., Archer et al., 2013; Buddhakosai et al., 2016; Morin et al., 2010, 2015; Shamblin et al., 2012). Several hypotheses have been proposed to explain the extremely low mtDNA diversity in sperm whales, including a population bottleneck (Lyrholm & Gyllensten, 1998; Lyrholm et al., 1996), low mutation rate (Lyrholm et al., 1996; Whitehead, 1998), stochastic variation in maternal lineage survival (Amos, 1999; Tiedemann & Milinkovitch, 1999), cultural hitchhiking (Whitehead, 1998, 2005; Whitehead, Vachon, & Frasier, 2017) and selective constraints on CR sequences that limit accumulation of variation and lead to saturation of sites free to vary (Alexander et al., 2013). Alexander et al. (2013) used complete mitogenome sequences ($N = 17$) to evaluate support for some of these hypotheses, and concluded that the data were consistent with a bottleneck or selective sweep (involving selection on mitochondrial protein-coding regions or cultural hitchhiking), and found no evidence for selective constraint on the CR or slow substitution rates in sperm whale mtDNA relative to other cetaceans. In this study, we analyse complete mitogenomes from 175 globally distributed sperm whale samples to investigate phylogeography, selection and historical demography. We compare demographic inferences from mitogenomes and nuclear genomes to evaluate the previously proposed hypotheses of a

population bottleneck/expansion vs. a selective sweep due to cultural hitchhiking or selection on mtDNA as the mechanism contributing to low worldwide mitochondrial diversity in sperm whales.

Materials and Methods

DNA extraction and library preparation

Sperm whale tissue samples (*n* = 158) collected by live biopsy or from dead stranded animals were stored in salt-saturated 20% DMSO at −20°C in the US National Marine Fisheries Service (NMFS) Marine Mammal and Sea Turtle Research (MMASTR) Collection at the Southwest Fisheries Science Center (SWFSC), or in 70% ethanol at −20°C at the Oregon State University Cetacean Conservation and Genomics Laboratory (CCGL). Sample information is in Table S1. DNA from SWFSC samples was extracted from tissue samples using either a silica-membrane method (Qiaxtractor® DX reagents, Qiagen, Valencia, CA, USA) or a simple salt-precipitation procedure (Miller, Dykes, & Polesky, 1988). DNA libraries for these samples were constructed and enriched for mitochondrial DNA according to Hancock-Hanser et al. (2013) and sequenced in three pools of 49–66 samples with Illumina GAII, HiSeq (100 bp) and NextSeq (75 bp) single-end reads. Eleven samples were repeated in two pools to increase read depth of coverage. CCGL samples were extracted using a standard phenol/chloroform method (Sambrook, Fritsch, & Maniatis, 1989), modified for smaller samples (Baker et al., 1994). Libraries for these samples were constructed from long-range PCR products following Alexander et al. (2013), individually barcoded and prepared for sequencing using a Nextera XT DNA Sample Preparation Kit (Illumina, La Jolla, CA, USA). Sample libraries were pooled and sequenced in three Illumina MiSeq paired-end runs (two 250 bp, one 75 bp).

Sequence read quality control and assembly

Assembly of sequence reads to the reference mitogenome (KC312603) was performed using custom scripts in R (R Core Team 2014) and publicly available programs as previously described (Hancock-Hanser et al., 2013; Dryad data repository https://doi.org/10.5061/dryad.cv35b). The reference mitogenome was modified to improve assembly coverage at the "ends" of the linearized mitogenome by adding 40 bp from each end of the sequence to the opposite end (so that reads could map across the artificial break point of the linearized sequence). Nucleotide sites in the consensus sequence for each sample were called "N" if there were <3

reads, <9 reads where there was nucleotide variation at the site among reads, or >9 reads where a single nucleotide did not represent >70% of the reads. All sequences were aligned and visually inspected in the program GENEIOUS (V. 6.0.5; Biomatters, Auckland, New Zealand), with indels and unique variants identified in GENEIOUS then verified by visual inspection of the read alignments of individual sample assemblies in the BAM files. New and 17 previously published sequences (Alexander et al., 2013) were aligned and checked for frameshift indels and coding sequence start and stop codons within protein-coding regions, based on published annotation of the reference sperm whale mitogenome.

Genetic diversity analyses and diagnosability

The R package strataG (Archer, Adams, & Schneiders, 2017) was used to calculate haplotype and nucleotide diversity by geographically defined population (stratum). To test whether haplotypic diversity and nucleotide diversity within strata were significantly different from one another, we estimated the variance of each measure via a stratified bootstrap (Supporting information Methods). In each iteration of the bootstrap, individuals were randomly chosen with replacement from each stratum, with sample sizes determined by the empirical data set. At each of the 1,000 bootstrap replicates, differences in nucleotide and haplotype diversity were calculated between each stratum (stratum 1 diversity - stratum 2 diversity, for each measurement). Strata were considered to have significantly different haplotype and/or nucleotide diversity if the observed difference was in the lower 5% of the distribution (strata 2 significantly greater diversity than stratum 1) or in the upper 5% of the distribution (strata 1 significantly greater than stratum 2).

Diagnosability is a measure of the ability to correctly determine whether a specimen of unknown origin can be correctly assigned to a group based on a trait or traits (Archer, Martien, & Taylor, 2017). To test the diagnosability of mitogenome sequences between Atlantic and Pacific sperm whales, we conducted a Random Forest analysis (Breiman, 2001) to create a model to classify samples to ocean basins following Archer, Martien et al. (2017). The Random Forest was built with 50,000 trees. Each tree was built using a random draw of 25 samples (half of the smallest oceanic sample size of n = 50 for the Atlantic, including the Gulf of Mexico and Mediterranean Sea) from each ocean basin to avoid biasing the model due to differences in sample size. The upper and lower confidence intervals of the correct classification score as well as the expected classification score (prior) were calculated as described in Archer, Martien et al. (2017).

Phylogenetic analyses

Time-calibrated phylogenetic analysis was performed using two methods: creating a prior distribution for the time to most recent ancestor (TMRCA) using a two-phase process (described below) and estimating the TMRCA using previously published substitution rates (see *Demographic reconstruction* section). The two-phase method utilized BEAST (v 1.8; Drummond, Suchard, Xie, & Rambaut, 2012). Aligned coding sequences were extracted based on published start/ stop codons of the reference sequence. Optimal substitution models for individual loci were determined based on Bayesian information content (BIC) using PARTITIONFINDER (v 1.1.1; Lanfear, Calcott, Ho, & Guindon, 2012). All loci were constrained to having the same underlying topology and clock rate (as loci on the mtDNA are fully linked without recombination). Phase 1 estimated the TMRCA for all sperm whale mitogenomes based on the coding partitions from four sperm whale haplotypes (mtGen11, 22, 30, 47) selected to represent the four major clades identified in a maximum-likelihood tree of all full mitogenome unique haplotypes (PHYML v. 2.2.0; Guindon, Delsuc, Dufayard, & Gascuel, 2009; implemented in Geneious), along with aligned coding region sequences from the pygmy sperm whale (*Kogia breviceps*, Accession no. NC005272) and four beaked whale species (*Berardius bairdii*, NC005274; *Ziphius cavirostris*, KC776696; *Mesoplodon densirostris*, KF032860; *Hyperodon ampulatus*, NC005273). The Bayesian phylogeny was calibrated using lognormal distributions on two calibration nodes: Odontoceti (34.67 Myr: 95% CI 29.93– 40.17 Myr) (McGowen, Spaulding, & Gatesy, 2009) and Ziphiidae (22 Myr: 14.86–32.56 Myr) (Dornburg, Brandley, McGowen, & Near, 2012; McGowen et al., 2009). The ucld prior was set to a uniform distribution (min. 1E-6, max. 1), the tree prior was speciation, Yule process, and the chain was 100M MCMC steps, logged every 10k steps. Convergence of two replicates and mixing were checked using
TRACER (v1.5; Rambaut, Suchard, Xie, & Drummond, 2014) and RWTY
(Warren, Geneva, & Lanfear, 2017). The maximum clade credibility tree was generated with TREEANNOTATOR (v1.8.1) in the BEAST software cluster (Drummond et al., 2012) after removal of the first 10% of trees.

The phylogenetic analysis of all sperm whale mitogenomes (Phase II) was conducted as above, separately for all complete mitogenome unique haplotypes ($N = 80$) and unique haplotypes of concatenated coding loci ($N = 60$), with TMRCA calibration based on the lognormal distribution (log(stdev = 0.2)) of the median TMRCA estimate of sperm whales from

Phase I (136.7 thousand years ago (kya), 95% CI 85.2–201.1 kya) following previously described methods (Morin et al., 2010, 2015). A strict clock was used with a constant size coalescent tree model. 100M MCMC steps and a single mutation model (TN93), selected based on the AIC in JMODELTEST v2.1 (Darriba, Taboada, Doallo, & Posada, 2012), were used for the full-length mitogenome sequences. For the concatenated protein-coding loci, we used 10M MCMC steps and 3rd position sites only, with two mutation models applied to *NADH6* and all other genes combined (HKY and TrN, respectively), based on analysis of individual loci in PARTITIONFINDER. For both analyses, convergence was checked based on four replicate runs using TRACER v1.5 and RWTY.

A haplotype median-joining network (MJN: Bandelt, Forster, & Ro€hl, 1999) was created using the program POPART (Leigh & Bryant, 2015) with default settings.

Demographic reconstruction

We used the codon-partitioned concatenated protein-coding regions of the sperm whale mitogenomes (excluding *NADH6* following Alexander et al., 2013; Ho & Lanfear, 2010) in a skygrid analysis (implemented in BEAST v1.8.3; Drummond et al., 2012) and a skyline analysis (implemented in BEAST v2.3.0; Bouckaert et al., 2014). All 175 samples were included in both analyses to approximate a "balanced" sampling strategy (multiple samples from multiple populations) found to result in the least bias in inferring demographic change (Heller, Chikhi, & Siegismund, 2013), and a separate GTR + G + I substitution model was used for each codon partition. Prior shapes for parameters within the substitution models and the relative rates for each codon position were derived from Alexander et al. (2013) (summarized in Table S2). For the skygrid analysis, two chains of 100M states, sampling every 100,000 states, were run. For the skyline analysis, two chains of 50 million states with sampling every 10,000 states were carried out. Convergence was assessed for the two analyses by comparing posterior distributions between chains in TRACER v1.6, and for topologies through RWTY. Demographic reconstructions were generated through TRACER. Model comparison was performed in TRACER using AICM with 1,000 bootstraps (Baele et al., 2012).

For demographic analysis based on the nuclear genome, we employed pairwise sequentially Markovian coalescent (PSMC; Li & Durbin, 2011) analysis of three sperm whale high-coverage genomes to infer changes in effective population size (*Ne*). We aligned published short read data from three sperm whales (GMX-SRS38925; BioSample SAMN01906698 (reference genome, Gulf of Mexico); read files SRR680161, SRR680169, SRR674482; SC991024-177-1;

BioSample SAMN06187412 (Pacific); read files SRR5136496, SRR5136506, SRR5136508, SRR5146847; BioSample SAMN06187413 (Indian); read files SRR5136491, SRR5136493, SRR5136497, SRR5146843), to the repeat-masked sperm whale reference genome v2.0.2 (Accession no. GCA_000472045.1; Warren, Kuderna et al., 2017). Details of read quality filtering and assembly are provided with Figure S1. The PSMC plot was scaled to an autosomal mutation rate (u_A) of 2.9×10^{-8} substitutions per nucleotide per generation (Dornburg et al., 2012) and a generation time of 31.9 years (Taylor, Chivers, Larese, & Perrin, 2007), and we conducted 100 bootstrap resamplings on all PSMC analyses. Sensitivity testing for different mutation rates and generation times was conducted by rescaling the PSMC plots (see Figure S1). While PSMC can provide useful insights into demographic change inferred from changes in coalescent rates (Li & Durbin, 2011), it is also sensitive to changes in population structure (Mazet, Rodriguez, Grusea, Boitard, & Chikhi, 2016). Therefore, we extended the PSMC analyses presented by Warren, Kuderna et al. (2017) by also carrying out PSMC analysis of a pseudodiploid genome made by randomly sampling an allele at each site from each of the individual genome assemblies using seqtk (https://github.com/lh3/seqtk; see Figure S1 for details). The pseudodiploid analysis provides information on changes in the rate of coalescence between two individual genomes through time, and therefore the timing of changes in population structure relative to the changes in *Ne* inferred by the single-genome PSMC analyses (see Cahill, Soares, Green, & Shapiro, 2016; Chikhi et al., 2018).

Ancestral range reconstruction

For biogeographic ancestral range reconstruction analyses, the full mitogenome phylogeny was pruned to twelve clades that had high posterior support. A custom R-script was then used to prune the tree topology to one branch per clade (N. Matzke; http://phylo.wikid ot.com/example-biogeobears-scripts#pruning_a_tree). The terminal branches were assigned to ocean basin(s) based on the origins of samples found within the nonpruned clades, and the tree and geographic information were used for biogeographic analysis using the R package BIOGEOBEARS (Matzke, 2012, 2013). The package was used to compare three basic models (DEC, DIVALIKE, BAYARELIKE), and two forms of each model (with and without founder-event cladogenesis; Matzke, 2014). These models allowed for different combinations of range expansion, range splitting (vicariance) and extinction, with the model conferring the highest likelihood on the tip data selected based on AIC. Likelihood-ratio tests were used to compare the nested models, with and without founder-event cladogenesis. Clades were subsampled to

test for the effect of uneven sampling (see methods in the legend of Figure S2).

Suitable habitat maps were generated with the AQUAMAPS approach to species distribution modelling (Kaschner, Tittensor, Ready, Gerrodette, & Worm, 2011; Kaschner et al., 2016; Ready et al., 2010) based on physical and oceanographic parameters for the last glacial maximum (LGM, ~20 kya), current and future conditions (year 2100, based on IPCC A2 emissions scenario). As adult male and female sperm whales have different habitat use patterns, with males migrating to higher latitudes to feed, we altered the global sperm whale suitable habitat model (males and females) to generate female-specific models by setting the minimum mean sea surface temperature (minSST, preferred minSST) to approximate the observed summer distributions of adult female and mixed family groups in the North Pacific identified from commercial whaling records (Ivashchenko, Brownell, & Clapham, 2014). The adjusted mean SST does not perfectly reflect the summer SST patterns, but allows approximate habitat models for the different time periods. All models exclude the primary production envelope because this information is not available for the LGM, but comparison of the presentday distribution maps with and without primary production indicated little effect of this parameter (Figure S3A–D, www.aquamaps.org, supported by previous research that suggested little correlation of primary production with sperm whale distributional patterns: Jaquet & Whitehead, 1996). Total sizes of distributions were subsequently calculated for both sexes in different ocean basins for all three time periods. We calculated areas using both a probability threshold of
0.0 (approximating the mean annual native range of a species including some potentially lesser utilized habitat) and ≥0.6 (shown to correspond to the core habitat of a species; Kaschner et al., 2011).

Selection analyses on mitogenome data

We aligned the codon-partitioned concatenated protein-coding regions of newly generated sperm whale mitogenomes with the multispecies mitogenome alignment of Alexander et al. (2013) and updated this with more recently published cetacean mitogenomes/ refseq versions of cetacean mitogenomes (see Table S3 for all accession numbers). Using the priors detailed in Table S4, and a Yule model of speciation, we ran two chains of 100 M generations sampling every 1,000 states in BEAST v2.3.0. After checking for convergence using TRACER and AWTY (Nylander, Wilgenbusch, Warren, & Swofford, 2008; Wilgenbusch, Warren, & Swofford, 2004), we used the tree we obtained for tree-based analyses of selection using TREE-SAAP v3.2 (Woolley,

Johnson, Smith, Crandall, & McClellan, 2003) and PAML v4.9 (Yang, 2007). In addition to these methods, we estimated selection for each codon using HYPHY (Pond, Frost, & Muse, 2005) as implemented in MEGA v6.06 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013). We further examined selection using the MEME (Murrell et al., 2012) and FUBAR methods (Murrell et al., 2013), as implemented on the DATAMONKEY webserver (Delport, Poon, Frost, & Kosakovsky Pond, 2010; Pond & Frost, 2005). MEME is designed to detect sites that have experienced episodic diversification, while FUBAR is designed to detect sites that have experienced pervasive diversification. We used both methods to detect any sites that have experienced positive selection in the sperm whale. We conducted all of these analyses using just one representative sperm whale haplotype (mtGen01) and GTR models of nucleotide substitution (or REV models where GTR was not available). For investigating specific sites inferred to be under positive selection, we used a = .05 as the threshold for statistical significance in HYPHY, MEME and FUBAR. Within PAML M8, we used Bayes empirical Bayes (BEB) (Yang, Wong, & Nielsen, 2005), with a threshold of $p > 95\%$. TREESAAP results were restricted to putative sites under selection where at least one property had a magnitude category of six or more in every pairwise comparison between the representative sperm whale and other cetaceans.

Following Caballero, Duchene, Garavito, Slikas, and Baker (2015), 3D homology models of proteins for genes where positive selection was detected in sperm whales (*ND1, ND2, ATP8, COX3, ND4L, COX3, ND4, ND5, CYTB*) were generated by the SWISS-MODEL server (Schwede, Kopp, Guex, & Peitsch, 2003), using the templates specified in Figure S4 (best-fitting model based out of the top four templates for each gene region), and visualized using UCSF CHIMERA v1.11 (Pettersen et al., 2004). This 3D model was annotated with domains based on alignment to the best-fitting template. We annotated a 2D model constructed with PROTTER (Omasits, Ahrens, Muller, & Wollscheid, 2014) with the domains and transmembrane topology based on alignment with the template, as well as additional secondary structure from the 3D model of the sperm whale constructed using DPSS (Kabsch & Sander, 1983).

Results

Mitogenome data set assembly

We generated mitogenome sequences for 158 new samples for this study. After combining

these with 17 previously published sequences from Alexander et al. (2013), our data set consisted of 175 globally distributed mitogenomes (Figure 1). We replicated three sequences from Alexander et al. (2013) using our capture enrichment methods and verified consistency between our approach and the long-range PCR approach of Alexander et al. (2013). Mean depth of coverage was 1269 (range 18–170), and all but 13 of the newly generated sequences contained ≤10 unresolved or missing nucleotides (maximum = 65). After alignment and verification of unique indels and variant sites, there were 80 unique haplotype sequences (Table S5).

Genetic diversity analyses and relationship of haplotypes

Mitogenome nucleotide diversity (p = 0.093%) was very similar to that reported previously for samples predominantly obtained near New Zealand (0.096%, Alexander et al., 2013) despite a 10-fold increase in sample size and global sample distribution. This estimate may still be slightly inflated due to selection of some specimens to maximize coverage of CR haplotype diversity both in this study and in Alexander et al. (2013). Sperm whales have some of the lowest documented mitogenomic diversity among cetaceans (see Table 1 in Alexander et al., 2013), but haplotype diversity was high overall (0.975) and differed significantly by ocean basin in all comparisons except "main basin" Atlantic vs. Gulf of Mexico (GoMx: both haplotype and nucleotide diversity) and Pacific vs. GoMx (nucleotide diversity) (Table 1; Table S6). Differences in diversity remained significant for all comparisons except GoMx vs. Mediterranean (haplotype diversity) even after removal of replicate haplotypes collected from the same social group to control for oversampling of close relatives (Table S6).

Phylogenetic analyses

The time-calibrated phylogeny for TMRCA of the four divergent sperm whale haplotypes with outgroups (Phase I) is in Figure S5. The mean substitution rate estimate for all codons was 7.596E-3 substitutions/site/Myr (95% CI = 6.357E-3–8.860E-3), lower than that estimated previously for the sperm whale (1.04E-2) and odontocetes (1.00E-2) based on different parameters (Alexander et al., 2013). The median TMRCA estimate of 136.7 kya (95% CI = 85.1– 201.1 kya) was used as the prior for analysis of the full set of sperm whale haplotypes (Phase II). The input xml, log and output trees files are available in the Dryad digital archive (https://doi.org/10.5061/ dryad.57271). The sperm whale phylogeny based on the full set of

complete mitogenome haplotypes is shown in Figure 2. All equivalent sample size (ESS) values were >200 and RWTY indicated convergence of topology among separate runs. The median TMRCA estimates from the full-length mitogenomes and the 3rd positions of concatenated coding loci were nearly identical: 126.4 kya (95% CI = 81.04–180.8 kya) and 126.5 kya (80.066–178.4), respectively. These estimates were older than the TMRCA estimated based on 17 mitogenomes from the Pacific alone (103 kya (95% HPD 72.8– 137.4); Alexander et al., 2013), likely due to the lower median substitution rate inferred in the current analysis (7.034E-3 substitutions per site per million years: 95% CI = 4.068E-3–1.078E-2). The input xml, log and output trees files are available in the Dryad digital archive (https://doi.org/10.5061/dryad.57271).

Full-length mitogenome haplotypes tended to be separated by only a few nucleotide substitutions (Figure 3). Despite this low diversity, a high degree of phylogeographic structure was evident: 65 of the 80 haplotypes were found only in the Pacific, 14 were found only in the Atlantic, and only one haplotype (mt03) was found in both of these ocean basins (1x N. Pacific, 9x N. Atlantic). Sampling in the Indian Ocean was limited to one sample from the Maldives that had a unique haplotype (mt33) that differed by 1 bp from a haplotype found in Tasmania (mt04). One additional Indian Ocean mitogenome (from the Seychelles) assembled from the genome data of Warren, Kuderna et al. (2017) was haplotyped as mt54, also found in the Tasmania. The three most common CR haplotypes globally (A, B, C based on 394 bp; Table S7; Alexander et al., 2016), which are shared between the Pacific and Atlantic Oceans (and the Mediterranean Sea), constituted 67% of the samples used in this study. These three CR haplotypes were further divided into 17, 16 and 14 mitogenome haplotypes, respectively, and the four samples from the Mediterranean Sea, all CR haplotype C, were split into two unique haplotypes (Table S7). Apart from mt03 (CR haplotype A), mentioned above, all of the 48 mitogenome haplotypes corresponding to these three CR haplotypes were ocean-basin specific.

We conducted Random Forest analysis to determine the probability of assigning known and newly discovered mitogenome haplotypes to ocean basin. When the GoMx and Mediterranean (Med.) samples were collapsed into the Atlantic population for purposes of assignment, results indicated 100% probability of correct assignment to the Pacific, but only 78% probability for the Atlantic Ocean (Table 2). Without the GoMx and Med. samples, the assignment probability was slightly lower for the Atlantic (71%). The large number of ocean-specific haplotypes (all but one shared Pacific/Atlantic haplotype) could suggest high diagnosability, but the lower success is due to both the single shared haplotype (mt03) and high similarity of

several Atlantic haplotypes to Pacific haplotypes (Figures 2 and 3).

Demographic reconstruction

The two skygrid chains showed convergence, and all skygrid parameters had ESS values above 200. The estimated TMRCA was 87 kya (95% HPD: 60.3–119.4). The skyline analysis also showed convergence between both chains, and all parameters had ESS values >200 after combining the chains. A TMRCA very similar to the skygrid analysis was recovered (Median: 86 kya; 95% HPD: 59–118.8 kya). The optimal model based on AICM was the skyline. Both analyses showed patterns consistent with a population expansion (Figure 4), but the timing appeared to differ, with the skyline suggesting the expansion began 30–35 kya and the skygrid analysis suggesting a more gradual increase from >80 kya (Figure 4). The 95% confidence intervals for the two methods indicate little resolution in the timing of the population expansion. As these analyses used all samples from both ocean-basin populations, assumptions of the models were violated. However, when limiting the sample set to only the Atlantic or Pacific Oceans (Figure S6), the pattern of population expansion is still recovered for the Pacific, the location of the inferred mitogenome MRCA (see ancestral reconstruction section). In contrast, the skyline plot is flat for the Atlantic, as might be expected where mitogenome clusters do not coalesce within the ocean basin. Input xml and output log and tree files are available in the Dryad digital archive (https://doi.org/10.5061/dryad.57271).

Although the nuclear genome PSMC analysis also recovered evidence of a recent population expansion, it indicated that this expansion was just one in a series of population fluctuations through time. Plots from analysis of all three ocean-basin samples declined starting about 2–3 million years ago (mya) to a low around 1 mya, with some fluctuation through the early and mid-Pleistocene climate cycles (Figure 4). In the late Pleistocene, the low effective population size (*Ne*), approximately 120–150 kya, appears to have increased to a peak *Ne* roughly 50 kya, then declined again to a low at about 20–30 kya just before the last glacial maximum (Figure 4). This pattern is nearly identical between samples from the Atlantic, Pacific and Indian Oceans, differing only during the last glacial cycle, where the estimated population size estimate in the Atlantic was lower (although variation is higher, so differences are less certain; Figure S1). These results are concordant with previous PSMC analyses of sperm whales (Warren, Kuderna et al., 2017). However, the *y*-axis of a PSMC plot is most accurately interpreted as the inverse of the coalescent rate, and changes in population structure can

wrongly be interpreted as changes in effective population size using this method (Foote et al., 2016; Mazet, Rodriguez, & Chikhi, 2015; Mazet et al., 2016). To further tease apart changes in structure from changes in effective population size, we generated pseudohybrids of pairs of individuals from different populations (Cahill et al., 2016). As populations undergo a gradual fission, the rate of coalescence between the two haploid genomes that comprise the pseudodiploid will decline forward in time, and should cease altogether when the two populations are completely isolated (Cahill et al., 2016). PSMC will thus infer a gradual increase in *Ne* as the coalescence rate decreases, and an infinite increase in effective population size at the point in time when the lineages become completely isolated.

PSMC analyses of the pseudodiploids indicate high uncertainty at >1 mya, but closely resemble the shared *Ne* of the three ocean-specific samples 0.15–1 mya, indicative of a stable rate of coalescence between lineages (although possibly still structured, see Chikhi et al., 2018) during this period (Figure 4). During the Eemian interglacial, at around 120 kya, the pseudodiploid estimates of *Ne* increase rapidly and abruptly rise to infinity for all three population-pair pseudodiploids. We interpret this as indicative of divergence between sperm whale populations associated with colonization of ocean basins at this time. Immediately following this period, estimates of *Ne* fall in both the Atlantic and Pacific (and to a lesser extent, the Indian Ocean) individual PSMC plots. We interpret these combined inferences as being indicative that the population split started in the Eemian, likely with some ongoing gene flow (resulting in additional coalescence events leading to the PSMC inferring a larger *Ne* in all samples), and that gene flow (and hence coalescence) then ceased altogether after the Eemian, leading to the inferred decline in *Ne* by PSMC in the individual samples. Thus, the increase and then decline in inferred *Ne* for each ocean between ~125,000 and 20,000 years ago is likely to be an artefact of population differentiation, masking the true effective population sizes.

Ancestral range reconstruction and suitable habitat models

Biogeographic ancestral range reconstruction analyses with BIO-GEOBEARS found dispersal–extinction–cladogenesis (DEC) as the most likely model, with the nested form of the model allowing for founder-event cladogenesis (DEC + J) significantly more likely than the traditional DEC model (p = .0057). The resulting phylogeographic inference for the pruned tree indicates that the root of the tree was in the Pacific, with multiple colonizations of the Atlantic around

approximately 20,000 and 60,000 years ago (Figure 5). The majority of our samples originated from the Pacific, which could potentially bias our ancestral area reconstruction (Moyle et al., 2016). We examined the effect of this discrepancy in sample size by downsampling the Pacific to equal that of the Atlantic (see Figure S2). Across n = 1,000 replicates downsampling the Pacific to 35 individuals, the root of the tree was inferred as Pacific in 99.9% of the replicates, indicating that uneven sampling is not driving this pattern.

Suitable habitat models for females in particular show a striking change between the present-day and the LGM, especially in the Atlantic Ocean (Figure 6), where core suitable habitat for females was reduced by 50% at the LGM (Table S8). The latitudinal shift also indicates that the Atlantic and Pacific habitats were likely completely separated by land masses, with only marginal potential for female dispersal between ocean basins even in the current warmer period. Projections of future habitat at the beginning of the next century suggest that dispersal between ocean basins may be more likely, but that Northern and Southern Hemisphere populations may become separated (Figure S3).

Selection analyses on mitogenome data

PAML detected a pervasive pattern of purifying selection across the cetacean mitogenomes (Table S9), with the ratio of nonsynonymous to synonymous changes estimated as x = .093. However, all methods except FUBAR detected at least one site putatively under positive selection in the lineage leading to sperm whales (after restricting sites inferred to be under positive selection to those where all sperm whales showed a fixed amino acid substitution in comparison with the rest of the cetacean species, Table S10). In total, 19 amino acids spanning eight mitochondrial-encoded proteins were found to show signatures of positive selection in the sperm whale (Table S10, Figure S4). Of these 19 amino acid changes, 12 of them occurred in or adjacent to transmembrane regions, contrasting with the overall pattern of mammalian evolution of most adaptive variation being restricted to loop regions (da Fonseca, Johnson, O'Brien, Ramos, & Antunes, 2008).

Discussion

Despite the high abundance and global extent of the sperm whale's distribution, mitogenomic diversity within this species is markedly low (this study; Alexander et al., 2013; Lyrholm & Gyllensten, 1998; Whitehead, 2005). However, without detailed examination of demographic

patterns indicated by the nuclear genome, previous research was unable to distinguish between demographic causes and selective sweeps as the most likely cause of this low diversity. Here, we showed that both the mitogenome and nuclear genome analyses provide evidence of a population expansion and ocean-basin divergence since the last interglacial period. The presence of this pattern across both genomes is consistent with a historically small effective population size (suggested by the PSMC plots for all three ocean basins) rather than selective sweeps on the mitogenome or cultural hitchhiking of haplotypes as the primary cause of low mtDNA diversity in sperm whales, despite our detection of sites under positive selection in the sperm whale mitogenome. In contrast, the expectation proposed under the mitochondrial selective sweep hypotheses (due to either direct selection on the mitogenome or a matrilineal cultural trait, and hitchhiking of linked neutral mtDNA diversity) is discordant patterns between the nuclear and mitochondrial genomes, with a decline in mitochondrial diversity over time (Whitehead, 2005) but limited or no concomitant decline in nuclear DNA diversity. This expectation is at least partially due to the high rates of sex-specific dispersal of males in sperm whales (Alexander et al., 2016), and recombination in the nuclear genome, unlinking the nuclear and mitochondrial genomes in these processes. However, it is important to note that given the strong influence of social structure on sperm whales (Whitehead et al., 2017), and the detection of positive selection on the protein-coding regions, we cannot rule out other forces, including cultural hitchhiking, having further reduced mitogenome diversity in the sperm whale. Nevertheless, the consistency of the inferred reduction in population size based on the nuclear genome occurring at approximately the same time as the TMRCA of the mitogenomes suggests demographic processes as the primary cause of low present-day sperm whale mtDNA diversity.

Our mitogenomic analyses suggest that the current global distribution of sperm whales results from a relatively recent expansion (20–40 kya). The inferred mitogenomic MRCA suggests expansion from a single, refugial population most likely located in the Pacific Ocean (though we cannot rule out an Indian Ocean refugium due to lack of sampling). “Ice-house” conditions are believed to have developed approximately 3 mya, characterized by the development of perennial Arctic Ocean sea ice and continental ice sheets in North America and Eurasia (Greene, Pershing, Cronin, & Ceci, 2008). This global change has been implicated in a global extinction event among marine megafauna (Pimiento et al., 2017) and coincides with an inferred decline in the global sperm whale population (revealed by PSMC analyses of individual sperm whale genomes), and an evident restriction in the distribution of sperm whales

(revealed by pseudodiploid PSMC analysis of the nuclear genome, Figure 4), potentially reducing the distribution to a single refugial population from which present-day sperm whale lineages descended. The global abundance appears to have increased from a low of <10,000 breeding individuals (*Ne*) since the last long cold period (Saale glaciation, ~80 kyr) that ended approximately 125 kya. Since then, there have been several shorter cold periods, lasting <10 kyr each (Figure 4), which appear to correspond approximately with coalescence nodes in the mitogenome tree (Figure 2), after which haplotypes diversified and mitochondrial lineages dispersed between oceans. Although timing estimates are approximate, the pattern suggests effects of climate on population size, distribution and persistence of specific maternal lineages through time in the sperm whale. While male sperm whales are known to travel to and feed in cold temperate/polar waters, females are generally restricted to warmer tropical and temperate waters, and mtDNA data suggest that they may only be able to disperse between the Pacific and Atlantic oceans during warm periods that allow them to extend their latitudinal range. However, the pseudodiploid analysis and divergence of the nuclear genome PSMC plots from samples located in different oceans over the last ~100,000 years (Figure 4) both suggest that males also rarely disperse between ocean basins, resulting in genetic divergence between the Atlantic and Pacific (see also Alexander et al., 2016; Lyrholm & Gyllensten, 1998; Lyrholm et al., 1999). Nuclear genomewide SNP analysis of more samples from each ocean basin—particularly the undersampled Indian Ocean—will be needed to infer degree of divergence and levels of current and/or historical gene flow.

The use of complete mitogenomes instead of CR sequences allows for substantially greater power to detect phylogeographic structure in low-diversity species such as the sperm whale and can also improve power in population genetic studies. The sampling design for this study was not appropriate for traditional frequency-based population genetic analyses, but comparative analysis of mitogenome and CR haplotypes for large geographic regions (Figure S7) indicated mitogenomes were more sensitive at detecting significant differentiation among strata based on F_{ST}, Φ_{ST} and chi-square statistics. Although patterns of differentiation were largely concordant between these data sets, the lack of power from CR data indicates caution must be used when genetic structure among potential strata is not detected solely based on CR data.

Sperm whales are among the deepest diving marine mammals, and it is likely that their mitochondrially encoded proteins have been under intense selection for oxygen efficiency (Janik, 2001) and robustness to pressure changes (Somero, 1992; Warren, Kuderna et al., 2017). Changes in transmembrane regions could more strongly affect protein structure and function than

in the loop regions (Saier, 1994), potentially directly addressing these selective pressures. Future comparisons to patterns of positive selection on other deep-diving species would be needed to see whether this was a pervasive pattern across other taxa.

Sperm whales are not considered rare, but have been depleted to approximately one-third of their prewhaling population sizes and are protected by various national and international laws and treaties (e.g., the US Endangered Species Act, IUCN Red List, International Whaling Commission). Given evidence of population structure within and between ocean basins (Alexander et al., 2016; Engelhaupt et al., 2009; Lyrholm & Gyllensten, 1998; Mesnick et al., 2011), the likelihood that populations were unevenly depleted by whaling (Ivashchenko et al., 2014), and lack of information on population recovery globally, it is highly probable that some sperm whale populations are still endangered or at risk (Carroll, Hedley, Bannister, Ensor, & Harcourt, 2014; Gero & Whitehead, 2016; Notarbartolo-Di-Sciara, 2014). Our results provide support for current isolation of populations located in different ocean basins, with episodic dispersal of females restricted to only warm climate periods. Although the current warming trends predict expansion of sperm whale habitat that allows both growth and opportunities for interocean dispersal, we cannot predict how rapid climate change may affect the ecosystems on which sperm whales depend (see, e.g., Dawe, Hendrickson, Colbourne, Drinkwater, & Showell, 2007; Ibanez et al., 2011; Jaquet, Gendron, & Coakes, 2003; Pecl & Jackson, 2008). In particular, climate change can affect the Atlantic and Pacific oceans quite differently (Boyle & Keigwin, 1985; Cheng et al., 2009; Greene et al., 2008; Howard, 1997), which may explain the apparent refugial population of sperm whales in the Pacific Ocean during the last long glacial period. Further research on trends in abundance and health of sperm whale populations and protection from other anthropogenic effects such as pollution (e.g., de Stephanis, Gimenez, Carpinelli, Gutierrez-Exposito, & Canadas, 2013; Savery, Wise, Falank et al., 2014; Savery, Wise, Wise et al., 2014; Unger et al., 2016), ocean noise (Mate, Stafford, & Ljungblad, 1994), ship strikes (Jensen & Silber, 2003), entanglement (Barlow & Cameron, 2003) and prey competition (Hucke-Gaete, Moreno, Arata, & Ctr, 2004) should be top priorities to prevent loss of already depleted populations.

Acknowledgements

We are grateful to all who collected and contributed samples used in this study: Marie Hill of the PIFSC Cetacean Research Program and Joint Institute for Marine and Atmospheric

Research, Erin Oleson (PIFSC), Allan Ligon, Adam U€, Mark Deakos, the Commander, U.S. Pacific Fleet, Office of Naval Research and Living Marine Resources Programs, Diane Claridge, John Durban, Kim Parsons, Luke Rendell, Barb Taylor, Louella Dolar, William Perrin, Deborah Thiele, Karen Evans, Robert Pitman, Tim Gerrodette, Tom Jefferson, Samual Hung, Liz Slooten, Sophie Laran, Emilie Praca, Lisa Ballance, Paul Wade, William McLellan, Ken Balcomb, Aleta Hohn, Jay Barlow, Dan Odell, Hisao Azuma, Christine Adkins, Wayne Haggard, Lars Kleivane, Manuel Furtado Neto, Sue Barco, Wayne McFee, K. Mullin and the Southeast US marine mammal stranding network. All samples have been imported to the US under CITES permit, and samples from the United States were collected under Marine Mammal Protection Act permits. We are grateful for technical and analytical assistance from The Scripps Research Institute core sequencing laboratory, Sebastian Duchene, Simon Ho, D. Steel, K. Hoekzema, C. Sislak and Milaja Nyk€anen. Cristina Garilao generously provided customized distribution maps from AQUAMAPS. We are very grateful for helpful discussion and comments on the manuscript provided by Hal Whitehead, Kim Parsons, Karen Chambers and four anonymous reviewers. Funding was provided by the National Marine Fisheries Service Southwest Region; the Marine Mammal and Turtle Division, Southwest Fisheries Science Center; U.S. Navy Pacific Fleet and NMFS Pacific Islands Science Center. Financial support for ADF was provided by the Welsh Government and Higher Education Funding Council for Wales through the Ser Cymru National Research Network for Low Carbon, Energy and Environment, and from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 663830.

Data Accessibility

Mitogenome haplotype sequences generated for this study have been submitted to GenBank (Accession nos. KU891329–KU891394). Sequence alignment input and output xml file from Beast phylogenetic and skyline/skygrid analyses are available from the Dryad Digital Repository https://doi.org/10.5061/dryad.57271.

References

Alexander, A., Steel, D., Hoekzema, K., Mesnick, S. L., Engelhaupt, D., Kerr, I., … Baker, C. S. (2016). What influences the worldwide genetic structure of sperm whales (*Physeter macrocephalus*)? *Molecular Ecology*, *25*, 2754–2772. https://doi.org/10.1111/mec.13638

Alexander, A., Steel, D., Slikas, B., Hoekzema, K., Carraher, C., Parks, M., .. Baker, C. S. (2013). Low diversity in the mitogenome of sperm whales revealed by next-

generation sequencing. *Genome Biology and Evolution*, *5*, 113–129. https://doi.org/10.1093/gbe/evs126

Amos, W. (1999). Culture and genetic evolution in whales. *Science*, *284*, 2055a.

Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). strataG: An R package for manipulating, summarizing, and analyzing population genetic data. *Molecular Ecology Resources*, *17*, 5–11. https://doi.org/10.1111/ 1755-0998.12559

Archer, F. I., Martien, K. K., & Taylor, B. L. (2017). Diagnosability of mtDNA with random forests: Using sequence data to delimit sub-species. *Marine Mammal Science*, *33*, 101–131. https://doi.org/10. 1111/mms.12414

Archer, F. I., Morin, P. A., Hancock-Hanser, B. L., Robertson, K. M., Leslie,
M. S., Berube, M., … Taylor, B. L. (2013). Mitogenomic phylogenetics of fin whales (*Balaenoptera physalus* spp.): Genetic evidence for revision of subspecies. *PLoS ONE*, *8*, e63396. https://doi.org/10.1371/ journal.pone.0063396

Baele, G., Lemey, P., Bedford, T., Rambaut, A., Suchard, M. A., & Alekseyenko, A. V. (2012). Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, *29*, 2157–2167. https://doi.org/10.1093/molbev/mss084

Baker, C. S., Slade, R. W., Bannister, J. L., Abernethy, R. B., Weinrich, M. T., Lien, J., … Palumbi, S. R. (1994). Hierarchical structure of mitochondrial DNA gene flow among humpback whales Megaptera novaeangliae, world-wide. *Molecular Ecology*, *3*, 313–327. https://doi. org/10.1111/j.1365-294X.1994.tb00071.x

Baker, C. S., Steel, D., Calambokidis, J., Falcone, E., Gonzalez-Peral, U., Barlow, J., … Yamaguchi, M. (2013). Strong maternal fidelity and natal philopatry shape genetic structure in North Pacific humpback whales. *Marine Ecology Progress Series*, *494*, 291–306. https://doi.org/ 10.3354/meps10508

Bandelt, H.-J., Forster, P., & Ro€hl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barlow, J., & Cameron, G. A. (2003). Field experiments show that acoustic pingers reduce marine mammal bycatch in the California drift gill net fishery. *Marine Mammal Science*, *19*, 265–283. https://doi.org/10. 1111/j.1748-7692.2003.tb01108.x

Bazin, E., Glemin, S., & Galtier, N. (2006). Population size does not influence mitochondrial genetic diversity in animals. *Science*, *312*, 570–572. https://doi.org/10.1126/science.1122033

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537. https://doi.org/10.1371/journal.pcbi.1003537

Boyle, E. A., & Keigwin, L. D. (1985). Comparison of Atlantic and Pacific paleochemical records for the last 215,000 years – changes in deep ocean circulation and chemical inventories. *Earth and Planetary Science Letters*, *76*, 135–150. https://doi.org/10.1016/0012-821X(85)90154-2

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Buddhakosai, W., Klinsawat, W., Smith, O., Sukmak, M., Kaolim, N., Duangchantrasiri, S., … Wajjwalku, W. (2016). Mitogenome analysis reveals a complex phylogeographic relationship within the wild tiger population of Thailand. *Endangered Species Research*, *30*, 125–131. https://doi.org/10.3354/esr00729

Caballero, S., Duchene, S., Garavito, M. F., Slikas, B., & Baker, C. S. (2015). Initial evidence for adaptive selection on the NADH subunit two of freshwater dolphins by analyses of mitochondrial genomes. *PLoS ONE*, *10*, e0123543. https://doi.org/10.1371/journal.pone.0123543

Cahill, J. A., Soares, A. E., Green, R. E., & Shapiro, B. (2016). Inferring species divergence times using pairwise sequential Markovian coalescent modelling and low-coverage genomic data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*, pii: 20150138. https://doi.org/10.1098/rstb.2015.0138

Carroll, G., Hedley, S., Bannister, J., Ensor, P., & Harcourt, R. (2014). No evidence for recovery in the population of sperm whale bulls off Western Australia, 30 years post-whaling. *Endangered Species Research*, *24*, 33–43. https://doi.org/10.3354/esr00584

Cheng, H., Edwards, R. L., Broecker, W. S., Denton, G. H., Kong, X. G., Wang, Y. J., … Wang, X. F. (2009). Ice age terminations. *Science*, *326*, 248–252. https://doi.org/10.1126/science.1177840

Chikhi, L., Rodriguez, W., Grusea, S., Santos, P., Boitard, S., & Mazet, O. (2018). The IICR (inverse instantaneous coalescence rate) as a summary of genomic diversity: Insights into demographic inference and model choice. *Heredity*, *120*, 13–24. https://doi.org/10.1038/s41437-017-0005-6

da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J., & Antunes, A. (2008). The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*, *9*, 119. https://doi.org/10.1186/1471-2164-9-119

Darriba, D., Taboada, G., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, *9*, 772. https://doi.org/10.1038/nmeth.2109

Dawe, E. G., Hendrickson, L. C., Colbourne, E. B., Drinkwater, K. F., & Showell, M. A. (2007). Ocean climate effects on the relative abundance of short-finned (*Illex illecebrosus*) and long-finned (*Loligo pealeii*) squid in the northwest Atlantic Ocean. *Fisheries Oceanography*, *16*, 303–316. https://doi.org/10.1111/j.1365-2419.2007.00431.x

de Stephanis, R., Gimenez, J., Carpinelli, E., Gutierrez-Exposito, C., & Canadas, A. (2013). As main meal for sperm whales: Plastics debris. *Marine Pollution Bulletin*, *69*, 206–214. https://doi.org/10.1016/j.mar polbul.2013.01.033

Delport, W., Poon, A. F., Frost, S. D., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, *26*, 2455–2457. https://doi.org/10. 1093/bioinformatics/btq429

Dornburg, A., Brandley, M. C., McGowen, M. R., & Near, T. J. (2012). Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution*, *29*, 721–736. https://doi.org/10.1093/molbev/msr228

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology & Evolution*, *29*, 1969–1973. https://doi.org/10.1093/molbev/ mss075

Engelhaupt, D., Hoelzel, A. R., Nicholson, C., Frantzis, A., Mesnick, S., Gero, S., … Mignucci-Giannoni, A. A. (2009). Female philopatry in coastal basins and male dispersion across the North Atlantic in a highly mobile marine species, the sperm whale (*Physeter macrocephalus*). *Molecular Ecology*, *18*, 4193–4205. https://doi.org/10. 1111/j.1365-294X.2009.04355.x

Finch, T. M., Zhao, N., Korkin, D., Frederick, K. H., & Eggert, L. S. (2014). Evidence of positive

selection in mitochondrial complexes I and V of the African elephant. *PLoS ONE*, *9*, e92587. https://doi.org/10.1371/ journal.pone.0092587
Foote, A. D., Morin, P. A., Durban, J. W., Pitman, R. L., Wade, P., Willerslev, E., … da Fonseca, R. R. (2011). Positive selection on the killer whale mitogenome. *Biology Letters*, *7*, 116–118. https://doi.org/10. 1098/rsbl.2010.0638
Foote, A. D., Vijay, N., Avila-Arcos, M. C., Baird, R. W., Durban, J. W., Fumagalli, M., … Wolf, J. B. (2016). Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications*, *7*, 11693. https://doi.org/10.1038/ncomms11693
Gero, S., & Whitehead, H. (2016). Critical decline of the eastern Caribbean sperm whale population. *PLoS ONE*, *11*, e0162019. https://doi. org/10.1371/journal.pone.0162019
Greene, C. H., Pershing, A. J., Cronin, T. M., & Ceci, N. (2008). Arctic climate change and its impacts on the ecology of the North Atlantic. *Ecology*, *89*, S24–S38. https://doi.org/10.1890/07-0550.1
Guindon, S., Delsuc, F., Dufayard, J. F., & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods in Molecular Biology*, *537*, 113–137. https://doi.org/10.1007/978-1-59745-251-9
Hancock-Hanser, B., Frey, A., Leslie, M., Dutton, P. H., Archer, E. I., & Morin, P. A. (2013). Targeted multiplex next-generation sequencing: Advances in techniques of mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology Resources*, *13*, 254–268. https://doi.org/10.1111/1755-0998.12059
Heller, R., Chikhi, L., & Siegismund, H. R. (2013). The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS ONE*, *8*, e62992. https://doi.org/10.1371/journa l.pone.0062992
Ho, S. Y., & Lanfear, R. (2010). Improved characterisation of among-lineage rate variation in cetacean mitogenomes using codon-partitioned relaxed clocks. *Mitochondrial DNA*, *21*, 138–146. https://doi.org/10. 3109/19401736.2010.494727
Hoelzel, A. R., Natoli, A., Dahlheim, M. E., Olavarria, C., Baird, R. W., & Black, N. A. (2002). Low worldwide genetic diversity in the killer whale (*Orcinus orca*): Implications for demographic history. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *269*, 1467–1473. https://doi.org/10.1098/rspb.2002.2033
Howard, W. R. (1997). Palaeoclimatology – A warm future in the past. *Nature*, *388*, 418–419. https://doi.org/10.1038/41201
Hucke-Gaete, R., Moreno, C. A., Arata, J., & Ctr, B. W. (2004). Operational interactions of sperm whales and killer whales with the Patagonian toothfish industrial fishery off southern Chile. *Ccamlr Science*, *11*, 127–140.
Ibanez, C. M., Cubillos, L. A., Tafur, R., Arguelles, J., Yamashiro, C., & Poulin, E. (2011). Genetic diversity and demographic history of *Dosidicus gigas* (Cephalopoda: Ommastrephidae) in the Humboldt Current System. *Marine Ecology Progress Series*, *431*, 163–171. https://doi.org/10. 3354/meps09133
Ivashchenko, Y. V., Brownell, R. L., & Clapham, P. J. (2014). Distribution of Soviet catches of sperm whales *Physeter macrocephalus* in the North Pacific. *Endangered Species Research*, *25*, 249–263. https://doi. org/10.3354/esr00641
Jackson, J. A., Steel, D. J., Beerli, P., Congdon, B. C., Olavarria, C., Leslie, M. S., … Baker, C. S. (2014). Global diversity and oceanic divergence of humpback whales (*Megaptera novaeangliae*). *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *281*, pii: 20133222. https://doi.org/10.1098/rspb.2013.3222

Janik, V. M. (2001). Is cetacean social learning unique? *Behavioral and Brain Sciences*, *24*, 337–338. https://doi.org/10.1017/ S0140525X0137396X

Jaquet, N., Gendron, D., & Coakes, A. (2003). Sperm whales in the Gulf of California: Residency, movements, behavior, and the possible influence of variation in food supply. *Marine Mammal Science*, *19*, 545–562. https://doi.org/10.1111/j.1748-7692.2003.tb01320.x

Jaquet, N., & Whitehead, H. (1996). Scale-dependent correlation of sperm whale distribution with environmental features and productivity in the South Pacific. *Marine Ecology Progress Series*, *135*, 1–9.

Jensen, A. S., & Silber, G. K. (2003). Large Whale Ship Strike Database, p.37. National Marine Fisheries Service, NOAA Technical Memorandum NMFS-OPR-25, Silver Spring, MD.

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*, 2577–2637. https://doi.org/10.1002/(ISSN)1097-0282

Kaschner, K., Kesner-Reyes, K., Garilao, C., Rius-Barile, J., Rees, T., & Froese, R. (2016). *AquaMaps: Predicted range maps for aquatic species*. Retrieved from www.aquamaps.org, Version 08/2016.

Kaschner, K., Tittensor, D. P., Ready, J., Gerrodette, T., & Worm, B. (2011). Current and future patterns of global marine mammal biodiversity. *PLoS ONE*, *6*, e19653. https://doi.org/10.1371/journal.pone. 0019653

Kopps, A. M., Ackermann, C. Y., Sherwin, W. B., Allen, S. J., Bejder, L., & Kru€tzen, M. (2014). Cultural transmission of tool use combined with habitat specializations leads to fine-scale genetic structure in bottlenose dolphins. *Proceedings of the Royal Society of London, Series B, Biological Sciences*, *281*, 20133245. https://doi.org/10.1098/rspb.2013.3245

Lanfear, R., Calcott, B., Ho, S., & Guindon, S. (2012). PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, *29*, 1695– 1710. https://doi.org/10.1093/molbev/mss020

Lang, A. R., Calambokidis, J., Scordino, J., Pease, V. L., Klimek, A., Burkanov, V. N., … Taylor, B. L. (2014). Assessment of genetic structure among eastern North Pacific gray whales on their feeding grounds. *Marine Mammal Science*, *30*, 1473–1493. https://doi.org/10.1111/ mms.12129

Leduc, R. G., Archer, F. I., Lang, A. R., Martien, K. K., Hancock-Hanser, B., Torres-Florez, J. P., … Taylor, B. L. (2016). Genetic variation in blue whales in the eastern pacific: Implication for taxonomy and use of common wintering grounds. *Molecular Ecology*, *26*, 740–751.

Leigh, J. W., & Bryant, D. (2015). POPART: Fullfeature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116. https://doi.org/10.1111/2041-210X.12410

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*, 493–496. https://d oi.org/10.1038/nature10231

Lyrholm, T., & Gyllensten, U. (1998). Global matrilineal population structure in sperm whales as indicated by mitochondrial DNA sequences. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *265*, 1679–1684. https://doi.org/10.1098/rspb.1998.0488

Lyrholm, T., Leimar, O., & Gyllensten, U. (1996). Low diversity and biased substitution patterns in the mitochondrial DNA control region of sperm whales: Implications for estimates of time since common ancestry. *Molecular Biology and Evolution*, *13*, 1318–

1326. https://doi. org/10.1093/oxfordjournals.molbev.a025578
Lyrholm, T., Leimar, O., Johanneson, B., & Gyllensten, U. (1999). Sex-biased dispersal in sperm whales: Contrasting mitochondrial and nuclear genetic structure of global populations. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, *266*, 347–354. https://doi.org/10.1098/rspb.1999.0644
Martien, K. K., Chivers, S. J., Baird, R. W., Archer, F. I., Gorgone, A. M., Hancock-Hanser, B. L., … Taylor, B. L. (2014). Nuclear and mitochondrial patterns of population structure in north pacific false killer whales (*Pseudorca crassidens*). *Journal of Heredity*, *105*, 611–626. https://doi.org/10.1093/jhered/esu029
Mate, B. R., Stafford, K. M., & Ljungblad, D. K. (1994). A change in sperm whale (*Physeter macrocephalus*) distribution correlated to seismic surveys in the Gulf of Mexico. *Journal of the Acoustical Society of America*, *96*, 3268–3269. https://doi.org/10.1121/1.410971
Matzke, N. J. (2012). Founder-event speciation in BioGeoBEARS package dramatically improves likelihoods and alters parameter inference in dispersal-extinction-cladogenesis (DEC) analyses. *Frontiers of Biogeography*, *4*(suppl. 1), 210.
Matzke, N. J. (2013). *BioGeoBEARS: BioGeography with Bayesian (and Likelihood) evolutionary analysis in R scripts*. Berkeley, CA: University of California.
Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology*, *63*, 951–970. https://doi.org/10.1093/sysbio/ syu056
Mazet, O., Rodriguez, W., & Chikhi, L. (2015). Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theoretical Population Biology*, *104*, 46–58. https://doi.org/10.1016/j.tpb.2015.06.003
Mazet, O., Rodriguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution–lessons for ancestral population size inference? *Heredity*, *116*, 362–371. https://doi.org/10.1038/hdy.2015.104
McGowen, M. R., Spaulding, M., & Gatesy, J. (2009). Divergence date estimation and a comprehensive molecular tree of extant cetaceans. *Molecular Phylogenetics and Evolution*, *53*, 891–906. https://doi.org/ 10.1016/j.ympev.2009.08.018
Mesnick, S., Taylor, B., Archer, E. I., Martien, K., Escorza Trevin~o, S., Hancock-Hanser, B. L., … Morin, P. A. (2011). Sperm whale population structure in the eastern North Pacific inferred by the use of single nucleotide polymorphisms (SNPs), microsatellites and mitochondrial DNA. *Molecular Ecology Resources*, *11*(suppl. 1), 278–298. https://doi. org/10.1111/j.1755-0998.2010.02973.x
Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*, 1215. https://doi.org/10.1093/nar/16.3.1215
Mizroch, S. A., & Rice, D. W. (2013). Ocean nomads: Distribution and movements of sperm whales in the North Pacific shown by whaling data and Discovery marks. *Marine Mammal Science*, *29*, E136–E165. https://doi.org/10.1111/j.1748-7692.2012.00601.x
Morin, P. A., Archer, F. I., Foote, A. D., Vilstrup, J., Allen, E. E., Wade, P., … Harkins, T. (2010). Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Research*, *20*, 908–916. https://doi.org/10.1101/gr.102954.109
Morin, P. A., Parsons, K. M., Archer, F. I., Avila-Arcos, M. C., Barrett-Lennard, L. G., Dalla Rosa, L., … Foote, A. D. (2015). Geographic and temporal dynamics of a global radiation and

diversification in the killer whale. *Molecular Ecology*, *24*, 3964–3979. https://doi.org/10. 1111/mec.13284
Moyle, R. G., Oliveros, C. H., Andersen, M. J., Hosner, P. A., Benz, B. W., Manthey, J. D., … Faircloth, B. C. (2016). Tectonic collision and uplift of Wallacea triggered the global songbird radiation. *Nature Communications*, *7*, 12709. https://doi.org/10.1038/ncomms12709
Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*, 1196–1205. https://doi.org/10.1093/molbev/mst030
Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*, e1002764. https://doi. org/10.1371/journal.pgen.1002764
Notarbartolo-Di-Sciara, G. (2014). Sperm whales, *Physeter macrocephalus*, in the Mediterranean Sea: A summary of status, threats, and conservation recommendations. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *24*, 4–10. https://doi.org/10.1002/aqc.2409
Nylander, J. A., Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L. (2008). AWTY (are we there yet?): A system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, *24*, 581–583. https://doi.org/10.1093/bioinformatics/btm388
O'Brien, S. J. (1994). A role for molecular genetics in biological conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *91*, 5748–5755. https://doi.org/10.1073/pnas.91.13.5748
Omasits, U., Ahrens, C. H., Muller, S., & Wollscheid, B. (2014). Protter: Interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, *30*, 884–886. https://doi.org/ 10.1093/bioinformatics/btt607
Pecl, G. T., & Jackson, G. D. (2008). The potential impacts of climate change on inshore squid: Biology, ecology and fisheries. *Reviews in Fish Biology and Fisheries*, *18*, 373–385. https://doi.org/10.1007/ s11160-007-9077-3
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*, 1605–1612. https://doi.org/10.1002/(ISSN) 1096-987X
Pimiento, C., Griffin, J. N., Clements, C. F., Silvestro, D., Varela, S., Uhen, M. D., & Jaramillo, C. (2017). The Pliocene marine megafauna extinction and its impact on functional diversity. *Nature Ecology & Evolution*, *1*, 1100–1106. https://doi.org/10.1038/s41559-017-0223-6
Pond, S. L., & Frost, S. D. (2005). Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, *21*, 2531–2533. https://doi.org/10.1093/bioinformatics/bti320
Pond, S. L., Frost, S. D., & Muse, S. V. (2005). HyPhy: Hypothesis testing using phylogenies. *Bioinformatics*, *21*, 676–679. https://doi.org/10. 1093/bioinformatics/bti079
Premo, L. S., & Hublin, J. J. (2009). Culture, population structure, and low genetic diversity in Pleistocene hominins. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 33–37. https://doi.org/10.1073/pnas.0809194105
R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org
Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2014). *TRACER v1.6*. Retrieved from http://tree.bio.ed.ac.uk/software/tracer/

Ready, J., Kaschner, K., South, A. B., Eastwood, P. D., Rees, T., Rius, J., … Froese, R. (2010). Predicting the distributions of marine organisms at the global scale. *Ecological Modelling*, *221*, 467–478. https://doi. org/10.1016/j.ecolmodel.2009.10.025

Saier, M. H. (1994). Computer-aided analyses of transport protein sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. *Microbiological Reviews*, *58*, 71–93.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Savery, L. C., Wise, S. S., Falank, C., Wise, J., Gianios Jr, C., Douglas Thompson, W., … Wise Sr, J. P. (2014). Global assessment of oceanic lead pollution using sperm whales (*Physeter macrocephalus*) as an indicator species. *Marine Pollution Bulletin*, *79*, 236–244. https://doi. org/10.1016/j.marpolbul.2013.12.007

Savery, L. C., Wise, J. T., Wise, S. S., Falank, C., Gianios Jr, C., Thompson, W. D., ... Wise Sr, J. P. (2014). Global assessment of arsenic pollution using sperm whales (*Physeter macrocephalus*) as an emerging aquatic model organism. *Comparative Biochemistry and Physiology. Toxicology and Pharmacology*, *163*, 55–63. https://doi.org/10.1016/j.cbpc.2014.01.003

Schwede, T., Kopp, J., Guex, N., & Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Research*, *31*, 3381–3385. https://doi.org/10.1093/nar/gkg520

Shamblin, B. M., Bjorndal, K. A., Bolten, A. B., Hillis-Starr, Z. M., Lundgren, I., Naro-Maciel, E., & Nairn, C. J. (2012). Mitogenomic sequences better resolve stock structure of southern Greater Caribbean green turtle rookeries. *Molecular Ecology*, *21*, 2330–2340. https://doi.org/10.1111/j.1365-294X.2012.05530.x

Somero, G. N. (1992). Biochemical ecology of deep-sea animals. *Experientia*, *48*, 537–543. https://doi.org/10.1007/BF01920236

Steiner, L., Lamoni, L., Plata, M. A., Jensen, S. K., Lettevall, E., & Gordon, J. (2012). A link between male sperm whales, *Physeter macrocephalus*, of the Azores and Norway. *Journal of the Marine Biological Association of the United Kingdom*, *92*, 1751–1756. https://doi.org/10.1017/S0025315412000793

Straley, J. M., Schorr, G. S., Thode, A. M., Calambokidis, J., Lunsford, C. R., Chenoweth, E. M., … Andrews, R. D. (2014). Depredating sperm whales in the Gulf of Alaska: Local habitat use and long distance movements across putative population boundaries. *Endangered Species Research*, *24*, 125–135. https://doi.org/10.3354/esr00595

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*, 2725–2729. https://doi.org/10. 1093/molbev/mst197

Taylor, B. L., Chivers, S. J., Larese, J., & Perrin, W. F. (2007). *Generation length and percent mature estimates for IUCN assessments of cetaceans*. Administrative Report, Southwest Fisheries Science Center, 8604 La Jolla Shores Blvd., La Jolla, CA 92038, USA.

Tiedemann, R., & Milinkovitch, M. C. (1999). Culture and genetic evolution in whales. *Science*, *284*, 2055a.

Unger, B., Rebolledo, E. L., Deaville, R., Grone, A., IJsseldijk, L. L., Leopold, M. F., … Herr, H. (2016). Large amounts of marine debris found in sperm whales stranded along the North Sea coast in early 2016. *Marine Pollution Bulletin*, *112*, 134–141. https://doi.org/10.1016/j.mar polbul.2016.08.027

Van Cise, A., Morin, P. A., Baird, R. W., Lang, A. R., Robertson, K. M., Chivers, S. J., …

Martien, K. K. (2016). Redrawing the map: mtDNA provides new insights into the distribution and diversity of short-finned pilot whales in the Pacific Ocean. *Marine Mammal Science*, *32*, 1177– 1199. https://doi.org/10.1111/mms.12315

Warren, D. L., Geneva, A. J., & Lanfear, R. (2017). RWTY (R We There Yet): An R package for examining convergence of Bayesian phylogenetic analyses. *Molecular Biology and Evolution*, *34*, 1016– 1020.

Warren, W. C., Kuderna, L., Alexander, A., Catchen, J., Perez-Silva, J. G., Lopez-Otın, C., … Wise Sr, J. P. (2017). The novel evolution of the sperm whale genome. *Genome Biology and Evolution*, *9*, 3260–3264. https://doi.org/10.1093/gbe/evx187

Whitehead, H. (1998). Cultural selection and genetic diversity in matrilineal whales. *Science*, *282*, 1708–1711. https://doi.org/10.1126/scie nce.282.5394.1708

Whitehead, H. (2002). Estimates of the current global population size and historical trajectory for sperm whales. *Marine Ecology Progress Series*, *242*, 295–304. https://doi.org/10.3354/meps242295

Whitehead, H. (2005). Genetic diversity in the matrilineal whales: Models of cultural hitchhiking and group-specific non-heritable demographic variation. *Marine Mammal Science*, *21*, 58–79. https://doi.org/10. 1111/j.1748-7692.2005.tb01208.x

Whitehead, H., Vachon, F., & Frasier, T. R. (2017). Cultural hitchhiking in the matrilineal whales. *Behavior Genetics*, *47*, 324–334. https://doi. org/10.1007/s10519-017-9840-8

Wilgenbusch, J. C., Warren, D. L., & Swofford, D. L. (2004). *AWTY: a system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference*. Retrieved from http://king2.sc.fsu.edu/CEBProjects/ awty/awty_start.php

Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A., & McClellan, D. A. (2003). TreeSAAP: Selection on amino acid properties using phylogenetic trees. *Bioinformatics*, *19*, 671–672. https://doi.org/10.1093/ bioinformatics/btg043

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*, 1586–1591. https://doi.org/10. 1093/molbev/msm088

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, *22*, 1107–1118. https://doi.org/10.1093/molbev/msi097

Figure 1 Map of all samples used in this study. Black circles = new mitogenome data, grey triangles = previously published mitogenome sequences (Alexander et al., 2013)

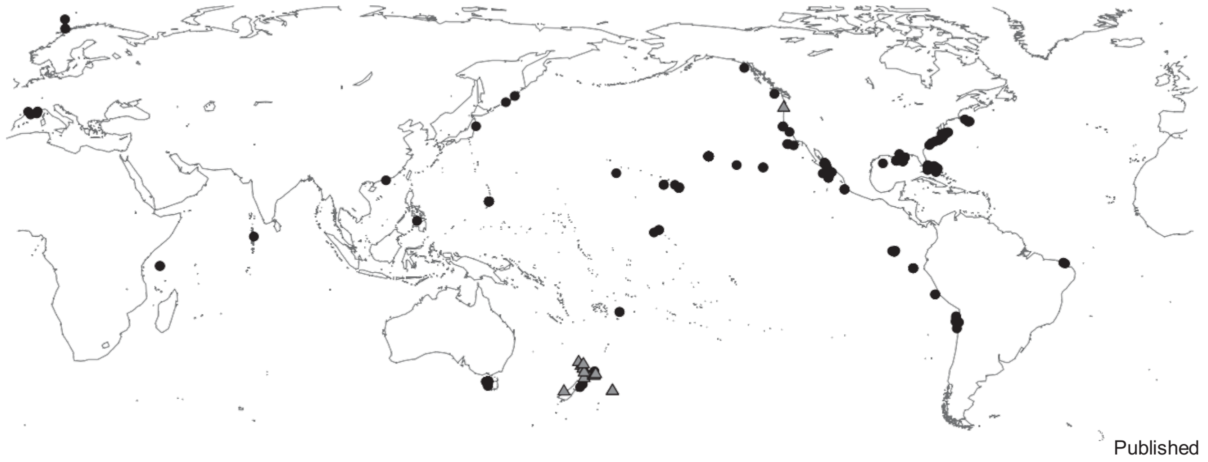


Figure 2
Time-calibrated (millions of years) phylogeny of all unique sperm whale mitochondrial genomes. Posterior probability support >0.5 is shown at nodes. Haplotype ID's are provided as "mt##" (see supplemental Table S1 for samples associated with haplotypes) followed by abbreviated locations (see supplemental Table S1). The clades to the right correspond to the pruned tree clades used for phylogeographic analysis (Figure 5). Branches are color coded by ocean basin where haplotypes are found: Blue = Pacific, Red = Atlantic, Green = Indian, Orange = Mediterranean, Pink = Pacific/Atlantic (shared). *Mitogenome haplotypes assembled from SRA data (Warren et al. 2017b).

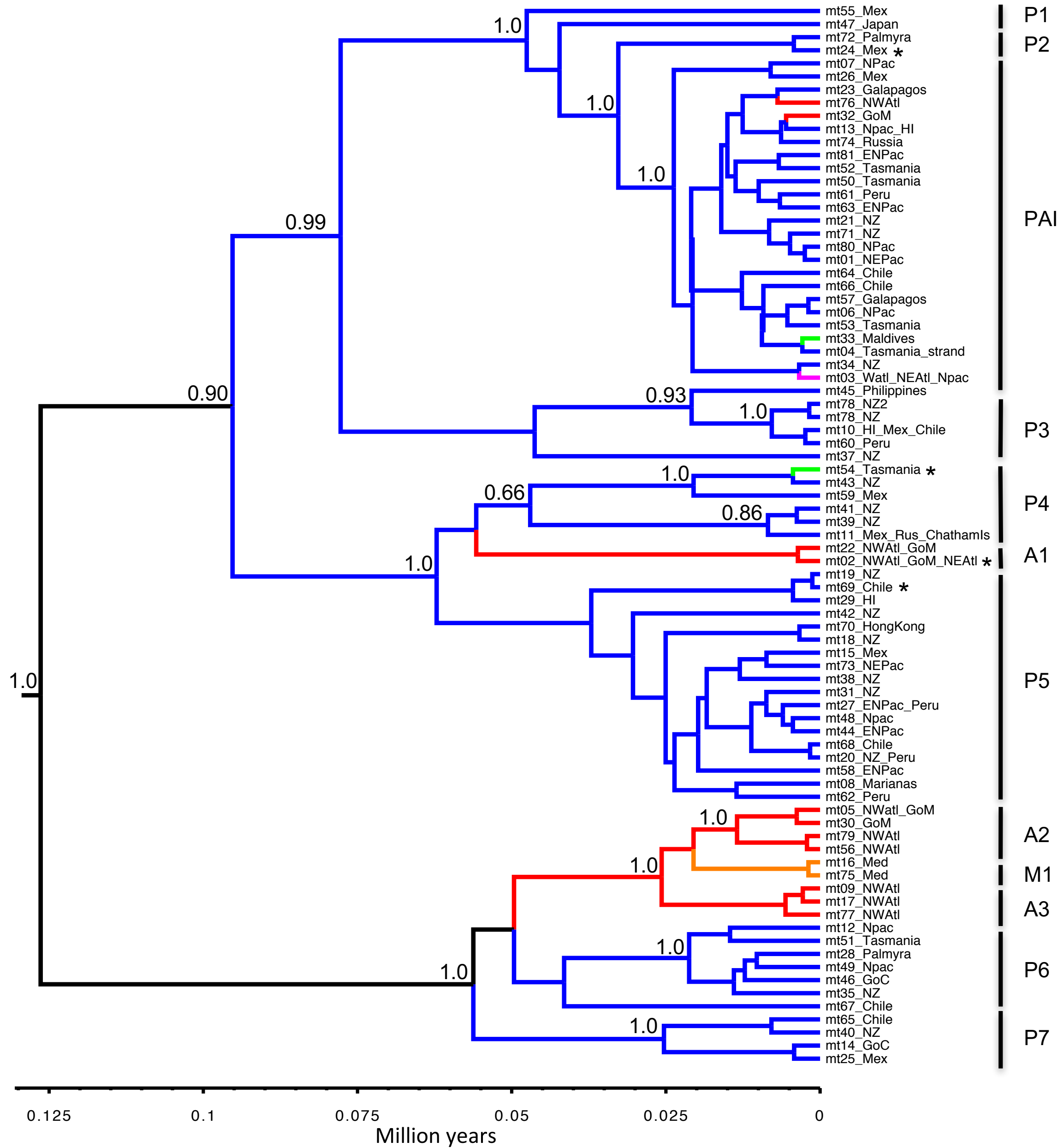


Figure 3
Median Joining Network of all unique haplotypes, colored by ocean. Tick marks on branches indicate the number of nucleotide differences. Not all 80 unique haplotypes are included in the network, as the algorithm collapses haplotypes with missing or ambiguous positions.

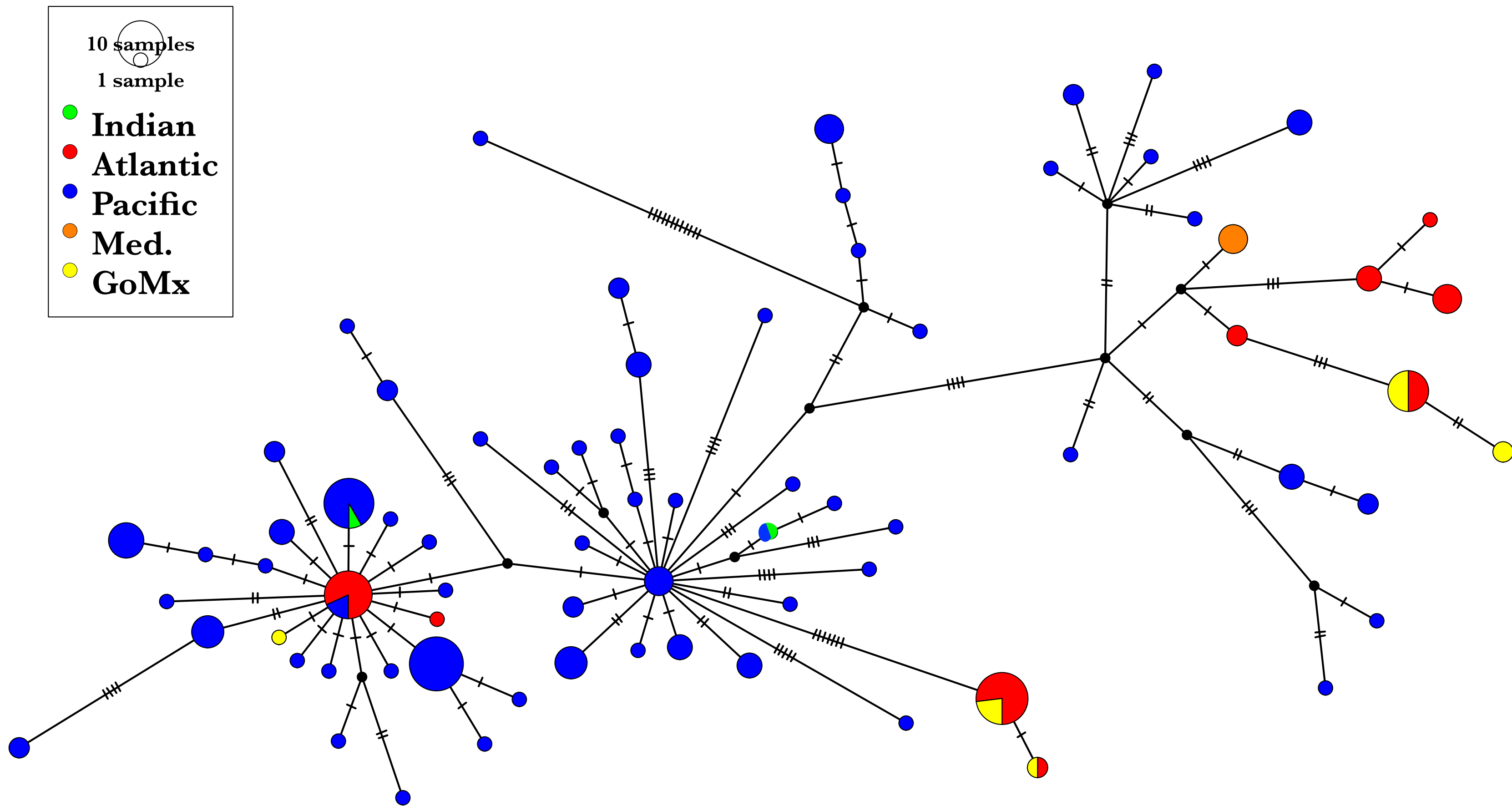


Figure 4
Demographic reconstructions based on the sperm whale mitogenomes and nuclear genome PSMC. The mitogenome skyline (black) and skygrid (gray) plots are shown with lower and upper 95% highest posterior density intervals shown by thinner lines and median values by thicker lines. Demographic estimates were converted to effective population size (females) by dividing by a τ of 31.9 years (Taylor *et al.* 2007). The nuclear PSMC plots are for three sperm whales from the Atlantic (red), Pacific (blue), and Indian (yellow) Oceans. Pseudo-diploid plots are shown for each ocean pair: Atlantic-Pacific (green), Atlantic-Indian (pink), and Pacific-Indian (purple). The X axis starts at 10,000 yr before present, and the Y axis is truncated at 3.5×10^4, as pseudo-diploid plots increase exponentially and bootstrap variation at <2×10^4 years was very large. Glacial maxima and Eemian warm period are shown with beige and yellow shading, respectively. Bootstrap plots for each PSMC analysis are presented in supplemental Figure S1.

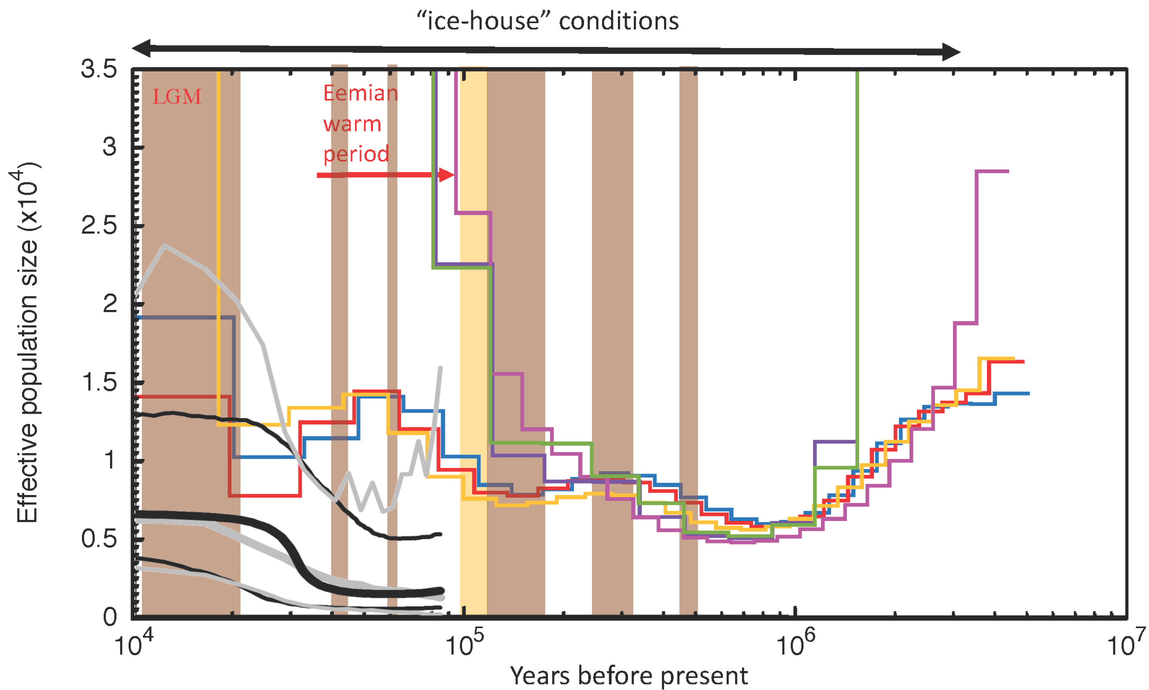


Figure 5
Most probable ancestral ranges for sperm whale populations estimated using the best fit model DEC+J from BioGeoBEARS and full mitogenomes. Clade IDs (tips) are as in Figure 2, and represent the tree topology pruned to one tip per clade. The boxes at nodes indicate the most likely geographic range immediately after cladogenesis. Oceanic regions are indicated by P = Pacific, A = Atlantic, I = Indian, M = Mediterranean.

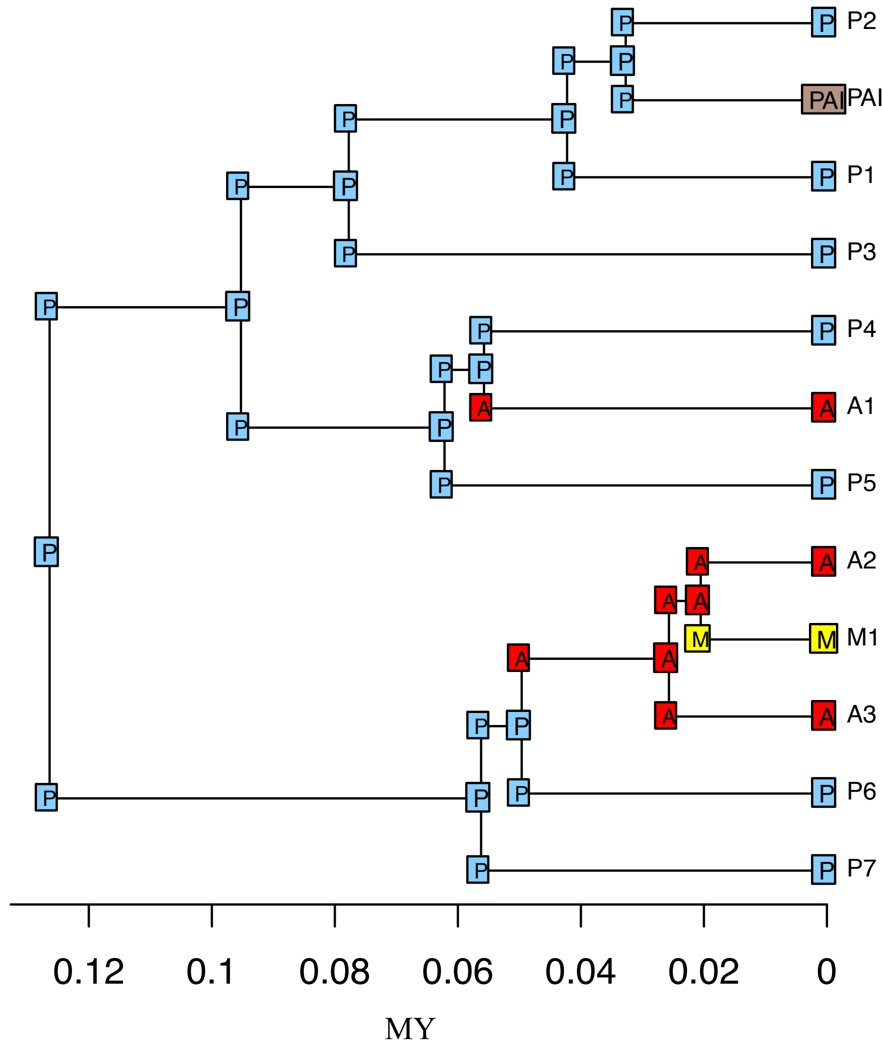
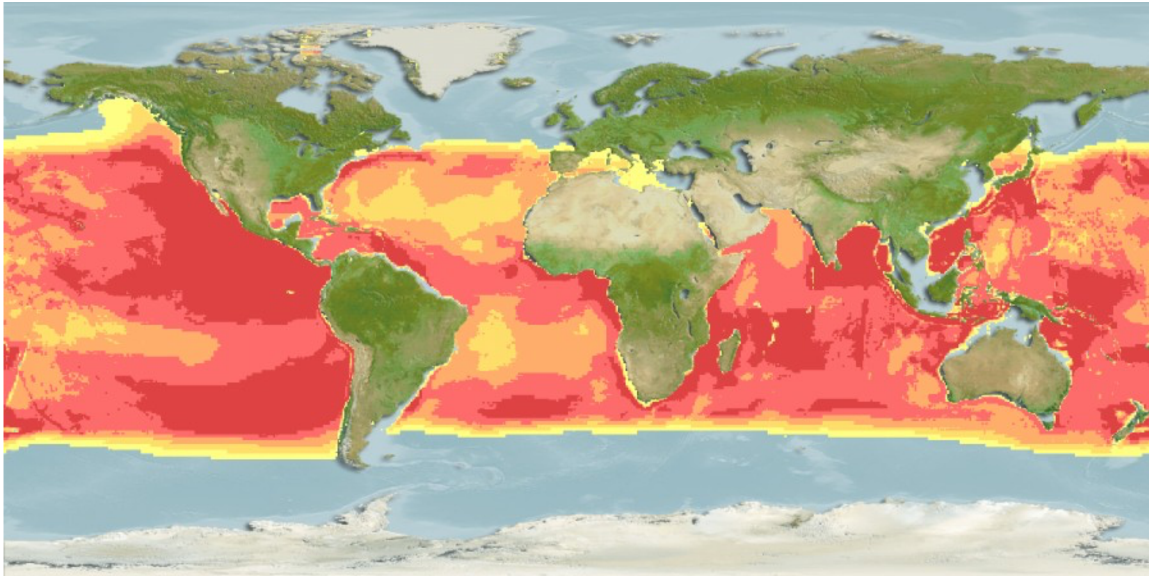


Figure 6.
AquaMaps environmental envelope models for distribution of female sperm whale habitat mapped to environmental conditions for a) last glacial maximum and b) current habitat. Dark red color indicates core suitable habitat; yellow indicates marginal habitat.

a)



b)

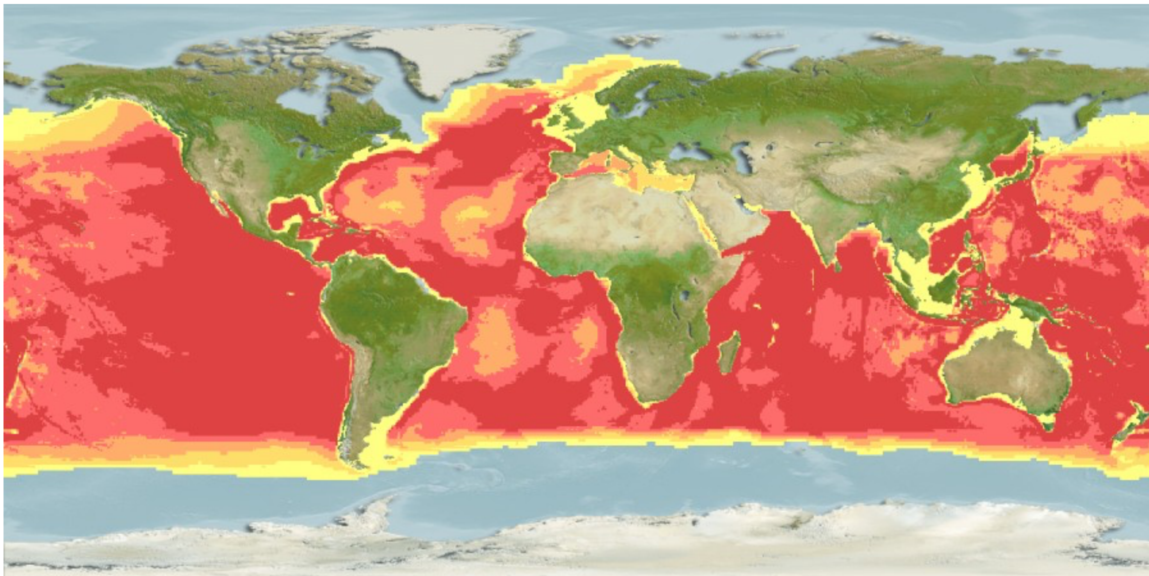


Table 1 Number of haplotypes, haplotypic diversity and nucleotide diversity by ocean basin

	No. of samples	No. of haplotypes	Haploytpe diversity	Nucleotide diversity
Atlantic[a]	35	10	0.839	0.000991
GoMx	11	5	0.818	0.001019
Mediterranean	4	2	0.500	0.000000
Pacific	124	66	0.972	0.000795
Global[b]	175	80	0.975	0.000934

Table 2 Random Forest assignment of samples to ocean basin, percent correctly assigned, and lower (LCI) and upper (UCI) 95% confidence limits. All sample origins were known (rows), but assigned to ocean basins based on the training subset of mitogenomes

	Atlantic	Pacific	Pct correct	LCI	UCI
Atlantic	39	11	78	64.0	88.5
Pacific	0	124	100	97.1	100