Oceanographic barriers, divergence, and admixture: Phylogeography and taxonomy of two putative subspecies of short-finned pilot whale

Amy M. Van Cise[1,2], Robin W. Baird[3], C. Scott Baker[4], Salvatore Cerchio[5], Diane Claridge[6], Russell Fielding[7], Brittany Hancock-Hanser[8], Jacobo Marrero[9,10], Karen K. Martien[8], Antonio A. Mignucci-Giannoni[11], Erin M. Oleson[12], Marc Oremus[4], M. Michael Poole[13], Patricia E. Rosel[14], Barbara L. Taylor[8], Phillip A. Morin[8]

1. Scripps Institution of Oceanography, 8622 Kennel Way, La Jolla, CA 92037
2. Woods Hole Oceanographic Institution, 266 Woods Hole Road, Woods Hole, MA 02453
3. Cascadia Research Collective, 218 1/2 4th Ave W, Olympia, WA 98501
4. Marine Mammal Institute, Hatfield Marine Science Center, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365, USA
5. New England Aquarium, 1 Central Wharf, Boston, MA 02110
6. Bahamas Marine Mammal Research Organisation, PO Box AB-20714, Marsh Harbour, Abaco, Bahamas
7. Department of Earth & Environmental Systems, University of the South, 735 University Avenue, Sewanee, TN 37383
8. Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 8901 La Jolla Shores Drive, La Jolla CA 92037
9. Asociación Tonina, Investigación y Divulgación del medio natural Marino. Calle Dr. Antonio González, Nº5, Planta 4º, CP: 38204, San Cristóbal de La Laguna, Tenerife (Islas Canarias), España
10. BIOECOMAC. Department of Animal Biology, La Laguna University, La Laguna 38206, Tenerife, Canary Islands, Spain
11. Centro de Conservación de Manatíes de Puerto Rico, Universidad Interamericana, 500 Carr. Dr. John Will Harris, Bayamón, 00957, Puerto Rico
12. Pacific Islands Fisheries Science Center, National Marine Fisheries Service, NOAA, 1845 Wasp Boulevard, Building 176, Honolulu, HI 96818
13. Marine Mammal Research Program, BP 698 Maharepa, Moorea, 98728, French Polynesia
14. Southeast Fisheries Science Center, National Marine Fisheries Service, NOAA, 646 Cajundome Blvd., Lafayette, LA, USA

Corresponding Author: Amy Van Cise
Woods Hole Oceanographic Institution
266 Woods Hole Road, MS#50
Woods Hole, MA 02453
508-289-3419
avancise@gmail.com

Running title: Global short-finned pilot whale phylogeography

Abstract

Genomic phylogeography plays an important role in describing evolutionary processes and their geographic, ecological, or cultural drivers. These drivers are often poorly understood in marine environments, which have fewer obvious barriers to mixing than terrestrial environments. Taxonomic uncertainty of some taxa (e.g. cetaceans), due to the difficulty in obtaining morphological data, can hamper our understanding of these processes. One such taxon, the short-finned pilot whale, is recognized as a single global species but includes at least two distinct morphological forms described from stranding and drive hunting in Japan, the 'Naisa' and 'Shiho' forms. Using samples (n = 735) collected throughout their global range, we examine phylogeographic patterns of divergence by comparing mitogenomes and nuclear SNP loci. Our results suggest three types within the species: an Atlantic Ocean type, a western/central Pacific and Indian Ocean (Naisa) type, and an eastern Pacific Ocean and northern Japan (Shiho) type. mtDNA control region differentiation indicates these three types form two subspecies, separated by the East Pacific Barrier: Shiho short-finned pilot whale, in the eastern Pacific Ocean and northern Japan, and Naisa short-finned pilot whale, throughout the remainder of the species' distribution. Our data further indicate two diverging populations within the Naisa subspecies, in the Atlantic Ocean and western/central Pacific and Indian Oceans, separated by the Benguela Barrier off South Africa. This study reveals a process of divergence and speciation within a globally-distributed, mobile marine predator, and indicates the importance of the East Pacific Barrier to this evolutionary process.

Keywords: Phylogeography, taxonomy, population structure, *Globicephala macrorhynchus*, cetacean

Introduction

Genomic phylogeography is the modern continuation of classic taxonomic disciplines, and as such has an important role in the description of evolutionary processes such as isolation, selection, and speciation (Bowen et al., 2016). Marine phylogeography attempts to evaluate these processes in an environment with few obvious barriers and many widely distributed species, some with large home ranges and long migratory routes. However, meta-analyses of phylogeography across several marine taxa reveal some common, large-scale barriers between what may be considered biogeographic provinces. These include the Isthmus of Panama, separating the Pacific and Atlantic Oceans; the East Pacific Barrier, which refers to the large, oligotrophic, deep open ocean that limits the dispersal of many tropical species between the eastern Pacific Ocean and the central/western Pacific Ocean; the Indo-Pacific Barrier (the Indo-Malay Archipelago), separating the western Pacific and Indian Oceans; the Benguela Barrier, separating the Indian and Atlantic Oceans; and the equatorial tropics, separating temperate species in the northern and southern hemispheres (Bowen et al., 2016; Davies, 1963; Gaither, Bowen, Rocha, & Briggs, 2016; Lessios, 2008; Perrin, 2007).

These barriers often cause genetic divergence that gives rise to populations, subspecies, or species. However, taxonomic under-classification can limit our understanding of evolutionary processes such as isolation and divergence, inhibit our understanding of the ecological drivers of species evolution, and undermine conservation efforts (Bowen et al., 2016; Taylor, Archer, et al., 2017; Taylor, Perrin, et al., 2017). Taxonomic species delineation based on morphological characteristics sometimes overlooks the existence of cryptic species, subspecies, or evolutionarily important population structure (Rosel et al., 2017; Taylor, Perrin, et al., 2017). Recent advances in

genomic techniques allow researchers to describe species' taxonomy and population structure with higher resolution than was previously possible (Cammen et al., 2016), oftentimes revealing cryptic speciation in the absence of physical barriers to dispersal, driven by local adaptation and/or social behavior (Leslie & Morin, 2016; Morin et al., 2015; Pazmiño et al., 2018; Podos, 2010; Rendell, Mesnick, Dalebout, Burtenshaw, & Whitehead, 2012; Rocha, Craig, & Bowen, 2007; Smith & Friesen, 2007; Yoshino, Armstrong, Izawa, Yokoyama, & Kawata, 2008)

This is especially true for some cetaceans, which, despite being highly mobile, often exhibit high site fidelity and adaptation to local environments (Andrews et al., 2010; Bowen et al., 2016; Foote et al., 2016; Hamner et al., 2012; Mahaffy, Baird, McSweeney, Webster, & Schorr, 2015). Others have ranges that cover entire ocean basins, yet exhibit socially-driven population structure (Balcazar et al., 2015; Carroll et al., 2015; Rendell et al., 2012; Witteveen et al., 2011). In this study, we use genetic data to understand the evolutionary phylogeography and propose taxonomic revision of a data-deficient cetacean species, the short-finned pilot whale, in order to improve our ability to understand evolutionary processes within this taxonomic unit.

Short-finned pilot whales (*Globicephala macrorhynchus*) are recognized as a single widely distributed species with a pan-tropical and pan-temperate distribution, strong social structure (Alves et al., 2013; Mahaffy et al., 2015; Whitehead, 1998), site fidelity (Mahaffy et al., 2015), and low mtDNA diversity, with widely distributed mtDNA control region haplotypes (Oremus et al., 2009; Van Cise et al., 2016). Two morphologically and genetically distinct types, originally described off Japan (Kasuya, Miyashita, & Kasamatsu, 1988; Oremus et al., 2009; Yamase, 1760), have largely non-overlapping distributions throughout the Pacific Ocean based on samples examined to date (Van Cise et al., 2016), as well as distinct vocal repertoires in tested regions (Van

Cise, Roch, Baird, Mooney, & Barlow, 2017). These two types, called ‘Naisa’ and ‘Shiho’ types, were originally described in 1760 based on morphological characteristics (Yamase, 1760). Their parapatric distributions around Japan remain segregated due to differing habitat preferences associated with thermally differentiated currents (Kasuya et al., 1988).

The Naisa and Shiho types differ in body size, melon (and skull) shape, color pattern (specifically the brightness of the saddle patch), and number of teeth (Kasuya et al., 1988; Miyazaki & Amano, 1994; Polisini, 1980; Yonekura, Matsui, & Kasuya, 1980). Naisa-type individuals are the smaller of the two types (females 316-405 cm, males 422-525 cm (Chivers, Perryman, Lynn, West, & Brownell, 2018)), with square-shaped melons and a dark, barely visible saddle patch. Shiho-type individuals are larger by one-to-two meters, with rounded melons and a bright saddle patch.

Based on morphological data and mtDNA control region sequences from outside Japan, the Shiho type has been found in the eastern Pacific Ocean from the northern to southern extent of the short-finned pilot whale range, while the Naisa type has been found in the central/western Pacific Ocean and in the Indian Ocean (Chen et al., 2014; Chivers, Perryman, Lynn, West, & Brownell, 2018; Oremus et al., 2009; Polisini, 1980; Van Cise et al., 2016). A single skull collected from Alaska indicates that the historical range of the Shiho type may have extended between the eastern North Pacific Ocean and northern Japan. Nuclear sequences from samples collected in Hawai‘i, the Mariana Islands, and the eastern tropical Pacific Ocean suggested that the two types may be genetically distinct in their nuclear DNA, with no male-mediated gene flow between them (Van Cise, Morin, Baird, Oleson, & Martien, 2016). This evidence suggests that Naisa- and Shiho-type short-finned pilot whales may be subspecies or species, but further genetic sampling and

132 analyses from throughout the global range of the species is needed to determine the correct
133 taxonomic delimitation of these two types.

134 To date, a global taxonomic study of short-finned pilot whales has been inhibited by a lack
135 of samples from the Indian and Atlantic Oceans. The limited data that are published suggest that
136 the distribution of the Naisa type may extend into the Indian Ocean (Van Cise et al., 2016), while
137 samples from the Atlantic Ocean haven't yet been classified. These data are insufficient to resolve
138 how short-finned pilot whales from these two ocean basins are related to the two types described
139 in the Pacific Ocean. In this study we examine the global phylogeography of short-finned pilot
140 whales, and present a formal proposal to recognize two subspecies of short-finned pilot whale.

141

142 **Methods**

143 *Sample collection and sequencing*

144 Samples were obtained from NOAA's SWFSC Marine Mammal and Sea Turtle Research
145 (MMASTR) Collection (n = 268) and from other contributors and collections throughout the world
146 (n = 53). The majority of tissue samples were collected by dart-biopsy of free-ranging whales,
147 using an 8 mm diameter biopsy dart deployed from a crossbow. All samples were collected under
148 permit and according to protocol to minimize disturbance to the animals. When possible, sampling
149 was limited to 1-2 individuals per encounter in order to minimize the effect of related individuals
150 on population structure. Additional samples were obtained from stranded animals. Sampling
151 locations are shown in Figure 1, and sample details in supplemental Table S1.

152 DNA was extracted from tissue samples using the methods described by Martien *et al*.
153 (2014). Genomic libraries were prepared and pooled for separate capture enrichment of

mitogenome and nuclear SNP loci according to the methods described in Hancock-Hanser *et al*. (2013) with minor modifications. The libraries used for nuclear locus enrichment were prepared using 400 ng of DNA per sample, pooled in equimolar amounts prior to capture enrichment on capture arrays containing nuclear loci only. The nuclear SNP capture array was modified from Van Cise, Martien, *et al.* (2017) to include 54 targeted loci, rather than the original 78 targeted loci, based on results from capture arrays used in Van Cise, Martien, *et al.* (2017). Single-end 100 bp reads were sequenced on an Illumina HiSeq500.

Mitogenome assembly

Mitogenome sequences were assembled using custom R scripts (R Core Team, 2016) written at SWFSC (Dryad data repository doi:10.5061/dryad.cv35b), which call on the Burrows-Wheeler alignment program (BWA; Li, Durban 2009). The reference sequence used to assemble mitogenome sequences (GenBank Accession No. JF339976) was modified at the ends to include 40 bp from the opposite end, in order to improve coverage in these regions across the artificial break point in the linearized sequence.

In most cases, nucleotides were called at a locus if there were at least 10 reads and that nucleotide was called in >80% of the reads, or if a locus had at least 5 reads with 100% agreement in nucleotide calling. Due to the possibility of "index-hopping" during library amplification and by Illumina sequencers (Jun et al., 2012; Kircher, Sawyer, & Meyer, 2012), the R scripts were modified to include additional quality control steps. If the 80% threshold was not met, an additional filtering step was used to probabilistically call the nucleotide at that locus. First, we determined the "common" nucleotide at each locus across all samples in the dataset, which was defined as the

nucleotide that was represented in >50% of the samples at the locus. We also determined the "rare" nucleotide at each locus, or the nucleotides(s) represented in <50% of the samples at that locus. Next, for each locus in each sample we calculated the number of reads that matched the common nucleotide, and the number of reads that matched one or more rare nucleotides. The common nucleotide was called if the proportion of common reads at that locus in the sample (common read proportion, c*rp*) was greater than the common read proportion at that locus across the entire sample pool (pooled common read proportion, *pcrp*). In other words, if *crp* > *pcrp*, the final call for that site went to the common nucleotide. Finally, rare nucleotides were called using a conservative, two-step approach that required 1) a high ratio of rare reads at that position in that sample versus the pooled dataset, and 2) a high binomial probability of the rare nucleotide at that site. If the proportion of rare nucleotide reads at that locus in the sample (rare read proportion, *rrp*) was greater than the proportion of the rare read at that site across all samples (pooled rare read proportion, *prrp*) by a ratio of at least $rrp = (prrp + 0.25)/1.25$, then the locus in question was passed to the binomial probability test based on the frequency of each nucleotide at that site across the entire sample set. The rare read at each locus had to pass the binomial probability test with a binomial probability greater than 95%. If the rare nucleotide passed each of these tests at a given locus for a given sample, the final call for that site went to the rare nucleotide. The modified R script is included as Supplementary File S1. Finally, consensus sequences for each individual were aligned and visually inspected in Geneious (V. 7.1.5, Biomatters, Auckland, New Zealand), and unique variants were verified by visual comparison with the BAM files.

Nuclear assembly and SNP genotyping

198 Nuclear sequences containing 112 SNPs in 54 previously selected and quality controlled
199 loci (Van Cise, Martien, *et al.*, 2017) were assembled as in Morin et al. (2015). Sequences were
200 assembled using custom scripts (Dryad data repository doi:10.5061/dryad.cv35b) and reference
201 sequences previously obtained from a draft genome of the common bottlenose dolphin (*Tursiops*
202 *truncatus*; assembly turTru1, Jul 2008; database version 69.1) as described elsewhere (Hancock-
203 Hanser et al., 2013; Van Cise et al., 2017). For each individual, SNP genotypes were called only
204 if there were a minimum of 10 reads at each position, to minimize genotyping error (Fountain,
205 Pauli, Reid, Palsbøll, & Peery, 2016). SNPs within the same locus were combined into multi-SNP
206 genotypes using PHASE (Morin et al., 2012; Stephens & Donnelly, 2003). Phasing was based on
207 allele frequencies across all samples, with a cutoff threshold of 0.5 to minimize bias against rare
208 heterozygotes (Garrick, Sunnucks, & Dyer, 2010), and the MCMC was run with a burn-in of
209 10,000, followed by 10,000 iterations, and thinned by 100 iterations.
210
211 *Data analysis: Phylogeography*
212 The published literature uses the terms 'Naisa type' and 'Shiho type' to refer to two groups
213 identified using mitochondrial and morphological data. Because multiple genetic datasets are used
214 in this study (mitogenomes, nuclear SNPs, and control region sequences), we use specific
215 nomenclature to orient the reader to the dataset being used in each analysis. When discussing
216 structure derived using mitogenome sequences, we refer to mitogenomic "clades". Similarly, when
217 discussing nuclear SNP data we refer to nuclear "groups". Finally, we combine the mitogenome
218 and nuclear SNP results to form a hypothesis of phylogeographic structure within short-finned
219 pilot whales based on geographically defined strata, and refer to these as "strata", which we test

using control region sequences. Control region sequences are used to test strata, rather than mitogenomes, in order to include a larger number of samples from the geographic range of the species, and to allow for comparison with published guidelines on taxonomic delimitation in marine mammals (Taylor, Archer, et al., 2017).

Mitogenome sequences were assigned haplotype labels using the strataG package (Archer, Adams, & Schneiders, 2017) implemented in the R computing environment. Tree topology was determined based on those haplotypes using a Bayesian maximum-clade-credibility (MCC) phylogenetic approach implemented in BEAST v. 1.8.4 (Drummond, Suchard, Xie, & Rambaut, 2012), rooted using a long-finned pilot whale mitogenome as the outgroup (GenBank Accession #HM060334.1). We used an HKY substitution model (Hasegawa, Kishino, & Yano, 1985) with gamma + invariant sites, which was selected using jModelTest (Darriba, Taboada, Doallo, & Posada, 2012; Guindon & Gascuel, 2003) for the complete mitochondrial genome haplotype alignment. We used an average substitution rate of 6.24×10^{-9} substitutions/site/year, based on recent analyses of killer whales (*Orcinus orca*) (Morin et al., 2015). Low mtDNA control region diversity (A M Van Cise et al., 2016) indicates a shallow tree; therefore, we do not expect variability in the substitution rate among branches. We therefore used a strict clock with a normal distribution and a standard deviation of 1×10^{-7}. Finally, we used a constant-size coalescent tree prior (Kingman, 1982) and 10 million MCMC steps sampled every 1000 steps. Convergence of four replicate runs was checked using TRACER v1.6 (Rambout, Suchard, Xie, & Drummond, 2014) and RWTY (Warren, Geneva, & Lanfear, 2017). TreeAnnotator (v1.8.1) in the BEAST software cluster (Drummond et al. 2012) was used to generate the maximum clade credibility tree after removal of the first 10% of trees. Additionally, a haplotype median joining network (MJN;

242 Bandelt *et al.* 1999) was generated using the program PopArt with default parameter settings
243 (Leigh & Bryant, 2015).

244 We examined population structure in the nuclear genotypes using STRUCTURE
245 (Pritchard, Stephens, & Donnelly, 2000), implemented in R using the strataG package. We used
246 settings for correlated allele frequencies, 10,000 MCMC steps with a burn-in of 1,000 steps, and a
247 k-range of 1 to 6 with 5 run for each k. For each value of k, runs were combined into a single
248 output using CLUMPP (Jakobsson & Rosenberg, 2007), and the optimum k value was selected by
249 calculating the modal Δk using Evanno metrics (Evanno, Regnaut, & Goudet, 2005; Verity et al.,
250 2016). In order to ensure that strong signals did not hide more localized population structure,
251 additional STRUCTURE analyses were performed within mitogenome clades and within the
252 geographic regions defined in the mitogenome population structure analyses below. We further
253 used strataG to calculate the number of alleles in each of the nuclear groups when K = 2 (Naisa
254 and Shiho groups), as well as the proportion of private alleles in each. We also analyzed nuclear
255 differentiation using a Discriminant Analysis of Principal Components (DAPC), implemented in
256 R using the adegenet package (Jombart, 2008; Jombart & Ahmed, 2011). We first ran an
257 unsupervised k-means DAPC to determine the number of clusters in the data by minimizing BIC.
258 We then ran a supervised DAPC, choosing the optimum number of PCs with a 10-fold cross-
259 validation test using a random selection of 90% of the data for training and 1,000 repetitions, to
260 minimize RMSE in classification when compared with STRUCTURE classifications.

261 *Data analysis: Taxonomy and the subspecies hypothesis*

262 We tested the subspecies hypothesis by calculating divergence and diagnosibility
263 according to guidelines established by Taylor *et al.* (2017). According to these guidelines, when

using mtDNA control region sequences to quantify differentiation, taxonomic categories for cetaceans (populations, subspecies, and species) are best characterized using Nei's d_A (subspecies: >0.004, species: > 0.02) and Percent Diagnosibility (PD) using a bootstrapped Random Forest analysis (subspecies and species: > 95%). Morphological data were not available for all samples, and therefore could not be used to stratify control region sequences for hypothesis testing. Therefore, we used the independent, unsupervised clustering of SNP data into nuclear groups to stratify samples into three groups, which were then tested for divergence and diagnosibility using control region sequences. Using control region sequences for this analysis allows us to integrate a larger number of samples into this test, and to compare our results with established guidelines for taxonomic delimitation in marine mammals. This test was conducted in two steps: in the first, only samples that were grouped *a priori* based on STRUCTURE analysis of SNP data were included (n = 105); in the second, all samples from the current study and previous studies for which we had mtDNA control region sequences (Hill et al., 2015; Martien, Hill, et al., 2014; Oremus et al., 2009; Van Cise et al., 2016; Van Cise et al., 2017) were included, resulting in a total of 725 samples. The additional samples were assigned to a type based on sampling location, corresponding with the STRUCTURE-derived nuclear group stratification. In order to account for the potential effect of social structure on pairwise estimates of divergence and diagnosibility, we subsampled the full control region dataset to include no more than three samples from each encounter with a group of pilot whales (n = 619), then re-ran these analyses and included the results in the Supplemental Materials.

Molecular diversity indices were calculated for all samples, and for each stratum, for mitogenomes (Theta (θ_H), mean nucleotide diversity (π), haplotype diversity, and number of

haplotypes) and SNP genotypes (average number of alleles per locus, expected and observed heterozygosity (H_e, H_o)), using the strataG package in R.

Pairwise differentiation was calculated to test the hypotheses that the STRUCTURE-derived stratification represents distinct populations, subspecies, or species, according to guidelines for subspecies delimitation presented by Taylor, Archer, *et al.* (2017), Archer, Martien, *et al.* (2017), and Rosel *et al.* (2017). Because each pairwise comparison tested a unique hypothesis, corrections for multiple pairwise tests are inappropriate for this analysis and were therefore not conducted (Armstrong, 2014; Perneger, 1998). In order to compare our results with guidelines on subspecies and species (Taylor, Archer, et al., 2017), we extracted 345 bp of the mitochondrial control region sequence from all mitogenomes, and combined these with previously published control region sequences throughout the global distribution of short-finned pilot whales (Oremus et al., 2009; Van Cise et al., 2016). We estimated Φ_{ST}; net nucleotide divergence, or d_A (Nei, 1987); and percent diagnosibility (PD) based on a random forest classification algorithm following Archer, Martien, *et al.* (2017). Φ_{ST} was calculated using a Tamura-Nei model with invariant sites and without a gamma correction (Tamura & Nei, 1993), which was identified by jModelTest2 (Darriba et al., 2012) as the substitution model that best fit the data; p-values were calculated based on 1,000 permutations. d_A was calculated using p-distance without a correction factor, with pairwise deletion of sites with indels, using a bootstrap approach with 1,000 replications. We estimated the magnitude of nuclear F_{ST} among STRUCTURE-assigned nuclear groups using SNP genotypes, but did not test the significance of this estimate because the nuclear groups are not *a priori* hypotheses. Mitochondrial F_{ST} was not calculated because it has been found

to be an unreliable indicator of taxonomic classification, exhibiting broad overlap in values among taxonomic classes within the order Cetartiodactyla (Rosel et al., 2017).

Results

Full mitogenome sequences (16,390 bp) were successfully assembled for 181 samples. Ninety-seven unique mitogenome haplotypes were found. Sixty-two haplotypes had no unknown nucleotides, and all but four mitogenome sequences had 10 or fewer Ns; haplotypes mtGen13, mtGen80, mtGen81, mtGen89 had 21, 111, 38, and 12 Ns, respectively. Control region sequences were extracted from these samples and aligned to previously sequenced control regions, resulting in 725 control region sequences from throughout the global distribution of short-finned pilot whales. Genotypes from 112 SNPs were generated at 47 unique loci for 245 samples (Supplemental Tables S2 and S3); samples were only included if they had genotypes for at least 70% of the 112 SNPs in the dataset. A total of 105 samples had both mitogenome sequences and SNP genotypes. Supplemental Figure S1 shows the number of samples in each dataset and overlap between datasets.

Phylogeography

Mitochondrial nucleotide diversity and number of haplotypes was greatest within the Naisa nuclear group and the Naisa geographic stratum, while haplotype diversity was greatest within the Atlantic nuclear group and Atlantic stratum (Table 1). Expected and observed heterozygosity, and average number of alleles, were also greatest in the Naisa nuclear group, followed by the Atlantic nuclear group.

329 The global phylogenetic tree (Figure 2) and mitogenome haplotype MJN (Figure 3) both
330 identify four mitogenomic clades; the previously described Shiho and Naisa types are contained
331 within two of those clades. Mitogenome haplotype frequencies within each clade can be seen in
332 Supplemental Table S4. The Shiho clade is distinguished from the other three mitogenome clades
333 by 15 substitutions (Figure 3). Shiho-clade short-finned pilot whales are primarily found in the
334 eastern Pacific Ocean and northern Japan, largely separated from the Naisa clade (Figure 4a). The
335 third Pacific clade has a geographic distribution that largely overlaps the Naisa clade in the Pacific
336 and Indian Oceans, and is referred to as Clade 3 in this study. Clade 3 extends into the eastern
337 Pacific, where it overlaps the distribution of the Shiho clade (Figure 4a). The fourth clade is found
338 only in the Atlantic Ocean, and will therefore be called the Atlantic clade, although there are also
339 three Atlantic Ocean sample haplotypes that were placed within the Naisa clade. Based on the
340 combined BEAST analysis log files, we estimate a mean rate of 7.88×10^{-8} (95% HPD: 5.48×10^{-10} – 2.0×10^{-7}) substitutions/site/year across the entire short-finned pilot whale mitogenome, and
342 the median divergence time of the Shiho clade from all other short-finned pilot whales to be
343 approximately 17.5 Kya (95% HPD: 3 – 176 Kya). Posterior support for each mitogenomic clade
344 are shown in Figure 2; mitogenome haplotype labels and coalescent time distributions are shown
345 in Supplemental Figure S2.
346 Based on ΔK (Evanno et al., 2005; Verity et al., 2016), the STRUCTURE analysis of
347 nuclear SNPs indicated k = 3 as the best supported number of groups (Supplemental Figure S3),
348 and differentiated Naisa, Shiho, and Atlantic nuclear groups, corresponding closely with Naisa,
349 Shiho, and Atlantic mitogenome clades, but did not support the differentiation of mitogenome
350 Clade 3; all but one of the Clade 3 whales grouped within the Naisa nuclear group (Figure 5a).

351 The unsupervised DAPC also returned an optimum group size of k = 3. Using 20 PCs (optimized
352 by minimizing RMSE in classification), we achieved 100% classification agreement with the
353 STRUCTURE analysis (Figure 5b).

354 Some disparity was found between nuclear classifications and mitogenomic clades. One
355 sample from the eastern tropical Pacific Ocean (ETP) had a Clade 3 mitogenome but clustered
356 with the Shiho group in its nuclear DNA. Seven samples had Naisa clade mitogenomes but
357 clustered with the Atlantic group in their nuclear DNA; of these, three were collected in the
358 Atlantic Ocean and four were collected in the western/central Pacific and Indian Oceans. One
359 sample, collected in the Bahamas, had an Atlantic clade mitogenome but clustered with the Naisa
360 group in its nuclear DNA (Figure 4b). Additional STRUCTURE analyses (results not shown) did
361 not indicate differentiation within the Shiho or Naisa nuclear groups, or mitogenome Clade 3.

362 Based on the concordance between mitogenomic clades and nuclear groups (Figures 2 and
363 4a and b), we define three distinct geographic strata within the short-finned pilot whale species
364 (Figures 4a and b): an Atlantic Ocean stratum, a Naisa stratum (encompassing the western/central
365 Pacific and Indian Oceans), and a Shiho stratum (occupying the eastern Pacific Ocean with a
366 potentially relic population in northern Japan). Stratifying individual samples based on this
367 hypothesis allows us to use geography as a proxy for genetic assignment in the absence of full
368 mitogenome or nuclear SNP data, so that much larger data sets of mtDNA control region sequences
369 can be used for phylogeographic and taxonomic analysis.

370

371 *Taxonomy and the subspecies hypothesis*

We estimated divergence and diagnosibility among the three geographic strata using 345 bp of the mtDNA control region, based on guidelines established to improve subspecies delineation using genetic data (Taylor, Archer, et al., 2017). Those guidelines suggested lower limits for two measures of mtDNA control region differentiation at the subspecies and species boundaries: d_A (0.004, 0.02) and Random Forest Percent Diagnosibility, or PD (95% for both). In addition to these two metrics, we also report Φ_{ST} (Table 2), for which subspecies were found to generally fall between 0.2 and 0.6, but this measure is not recommended for use in the guidelines because it can result in overclassification (Rosel *et al*., 2017), therefore we did not use this metric in our taxonomic evaluation.

We examined pairwise net divergence (d_A) and PD of control region sequences for both the nuclear groups (N = 105) and the geographic strata (N = 725, Table 2). For the smaller data set based on the nuclear groups, the Naisa versus Shiho comparison met the subspecies threshold proposed in Taylor, Archer, *et al.*'s (2017) guidelines for both metrics (d_A > 0.004, PD > 95%). The Atlantic-Shiho nuclear group comparison met the threshold for PD but not dA, while the Atlantic-Naisa comparison did not meet either threshold. When using the larger data set from the geographically defined strata, both the Naisa vs. Shiho geographic strata comparison and the Atlantic vs. Shiho geographic strata comparison met the subspecies thresholds for both PD and dA, while the Atlantic vs. Naisa geographic strata comparison met the dA threshold but not the PD threshold. Using a subsampled dataset to control for the potential effects of social structure did not appreciably change the results (Supplemental Table S5).

Within the STRUCTURE-defined, K = 2 nuclear groups (Supplemental Figure S4), the Naisa/Atlantic nuclear group had 104 private alleles (50%), while the Shiho nuclear group had 7

394 private alleles (4%). In total, 54% of the alleles in the dataset were private to either the Naisa or
395 Shiho nuclear group.

396 Although not supported by the nuclear SNP analyses, two mitogenomic clades (Naisa and
397 Clade 3; Figures 2, 3) were found in the western/central Pacific Ocean, with Clade 3 extending
398 into the eastern Pacific Ocean. Using only the control region sequences, differentiation between
399 these two mitogenome clades within the western/central Pacific/IO region met the threshold for
400 subspecies (d_A = 0.01, Table 3). We additionally examined control region differentiation between
401 the eastern Pacific Ocean and the western/central Pacific/IO within Clade 3, which was the only
402 clade to span multiple geographic regions, and found significant differentiation with Φ_{ST} with high
403 diagnosability (96%), but net divergence did not meet the subspecies threshold.
404
405 **Discussion**
406 *Phylogeography*
407 Our results indicate that there are at least three divergent types of short-finned pilot whales
408 throughout their global distribution (Figures 2, 4, 5). Rather than directly conforming with ocean
409 basins, the three types are distributed predominantly in the Atlantic Ocean (Atlantic type),
410 western/central Pacific and Indian Oceans (Naisa type), and eastern Pacific Ocean (Shiho type).
411 The Shiho clade diverged earliest, approximately 17.5 Kya, corresponding with the last glacial
412 maximum (~18 Kya). The Atlantic clade was the next to diverge from Naisa/Clade 3. Although
413 the timing of this split had little support in the mitogenome tree (0.30), it was further supported by
414 nuclear DNA, which did not support a split between the Naisa mitogenome clade and Clade 3.

415 The distributions of these three types correlate with geographic and oceanographic
416 boundaries that are found among globally-distributed species (Figure 4). The Eastern Pacific
417 Barrier is a known barrier to many shallow, coastally-distributed fishes, corals, and other
418 invertebrates (Bowen et al., 2016; Chow et al., 2011; Rocha et al., 2007). Increasing evidence
419 suggests it may also be an important barrier separating mobile trans-Pacific or globally-distributed
420 species, such as Galapagos sharks (*Carcharhinus galapagensis*) (Pazmiño et al., 2018), tope sharks
421 (*Galeorhinus galeus*) (Chabot, 2015), Risso's dolphins (*Grampus griseus*) (Chen et al., 2018), and
422 spinner dolphins (*Stenella longirostris*) (Leslie & Morin, 2018). Some evidence suggests that the
423 barrier may be semi-permeable to some species (Lessios & Robertson, 2006), allowing occasional
424 migration and mixing between the eastern and western Pacific Ocean. Similarly, the Eastern
425 Pacific Barrier seems to be a semi-permeable barrier between Shiho- and Naisa-type short-finned
426 pilot whales. The exact boundaries of their ranges in this area remain undefined due to lack of data,
427 and it may be a region of occasional geographic overlap or temporal segregation. The species has
428 been found distributed throughout this oligotrophic region (Hamilton et al., 2009), though their
429 density is lower there. The concordance across taxa of isolation and divergence along the East
430 Pacific Barrier indicates this barrier, although not formed by a land mass, may be as important as
431 more obvious barriers in driving evolutionary processes within marine taxa.
432 The Atlantic type is bordered to the west by the Isthmus of Panama, which separated the
433 Atlantic Ocean from the eastern Pacific Ocean approximately 3.5 Mya (Lessios, 2008), long before
434 the estimated radiation of short-finned pilot whales began. To the east, mixing between the Atlantic
435 and Naisa types is limited by the cold Benguela Current on the southwest side of Africa. The long-
436 finned pilot whale, the sub-arctic sister species to the short-finned pilot whale, inhabits the

437 Benguela Current, while short-finned pilot whales prefer the warmer Agulhas Current, based on
438 stranding records (Findlay, Best, Ross, & Cockcroft, 1992; van Bree, Best, & Ross, 1978). The
439 Benguela Barrier has limited dispersal between the Atlantic and Indian Ocean species for
440 approximately 2.5 My (Dwyer et al., 1995). Similar to the East Pacific Barrier, the Benguela
441 Barrier is semi-permeable, preventing dispersal in temperate mobile species such as whale sharks
442 (Castro et al., 2007), sailfish (Graves & McDowell, 1995) and blue marlins (Buonaccorsi,
443 Mcdowell, & Graves, 2001), but permitting occasional migrations of tunas and other pelagic fishes
444 (e.g. Viñas, Alvarado Bremer, & Pla, 2004). These occasional dispersal events are likely driven
445 by southward incursions of the warm Agulhas Current from the southeast, providing a potential
446 warm-water route for sporadic gene flow between the Atlantic Ocean and western/central Pacific
447 and Indian Oceans (Hutchings et al., 2009). In most species, mixing tends to be unidirectional,
448 following the prevailing current westward from the Indian to the Atlantic Ocean (Bowen et al.,
449 2016). Large, globally distributed whales also exhibit restricted gene flow between ocean basins,
450 although they are not restricted to tropical or temperate waters (e.g. Baker et al., 1993).

451 The Indo-Pacific Barrier is a common barrier for many tropical and coastally-distributed
452 species (Bowen et al., 2016), but our data do not show that this is a barrier to gene flow in short-
453 finned pilot whales. It is possible, rather, that the complicated bathymetry in the region provides a
454 rich prey base and habitat for short-finned pilot whales, which are often found along shelf breaks
455 and slopes where they are thought to hunt deep water squid species. Similarly, we found no
456 evidence of differentiation across the equatorial tropics.

457 There is evidence of limited historical or continued gene flow between the Atlantic and
458 Naisa types, as well as the Naisa and Shiho types (Figures 2, 5). Our data support the migration of

459 males and females from the Pacific or Indian Oceans to the Atlantic Ocean; it is possible that males
460 and females migrated separately, but due to their social nature it is likely these animals migrated
461 as social units across the Benguela Barrier. We also found evidence for male migration in the other
462 direction, from the Atlantic Ocean to the western/central Pacific and Indian Oceans. Finally, we
463 found evidence of female migration from the western/central Pacific and Indian Oceans to the
464 eastern Pacific Ocean. Although these samples suggest the potential for historical or ongoing gene
465 flow across barriers, it is important to note that there are several artifacts that might affect genetic
466 clustering. We were able to rule out missing SNPs, high homozygosity, or large numbers of
467 unknown mitogenome nucleotides as potential drivers of non-geographic clustering among
468 samples. Genotype errors may also be caused by potential cross-contamination among samples
469 sequenced in the same lane (Jun et al., 2012), miscalled genotypes, errors introduced during the
470 phasing of genotypes into haplotypes, or errors introduced by the STRUCTURE algorithm (e.g.,
471 violation of model assumptions). Because the number of samples indicating gene flow between
472 these types is small, we caution against drawing specific conclusions about gene flow among
473 geographic regions without additional sampling.

474 Mitogenome Clade 3 was not supported in the nuclear DNA (Table 3 and Figure 5), and
475 its distribution overlaps the Naisa clade throughout its range. In the Mariana Islands, Clade 3 and
476 Naisa individuals have been found in the same social groups (Hill et al., 2018). Within Clade 3,
477 the eastern Pacific Ocean regions and western/central Pacific and Indian Oceans regions were
478 significantly differentiated, mimicking the patterns seen in the Naisa and Shiho types. These
479 patterns could be caused by historically divergent clades with recent mixing or by lineage sorting

480 within a widely distributed population, and may be better understood with additional nuclear and
481 morphological data from Clade 3.

482 Genetic phylogeography has often been based on mitochondrial (mtDNA) control region
483 diversity. Yet this single locus can, under certain conditions (e.g. low genetic diversity, large
484 effective population size), misrepresent underlying patterns of isolation, divergence or speciation.
485 Low mtDNA diversity may arise for a number of reasons, including recent population bottlenecks
486 (Hoelzel, Fleischer, Campagna, Alvord, & Le Boeuf, 2002; Morin et al., 2018; Weber, Stewart, &
487 Lehman, 2004), variation in mutation rates across the mitogenome (Aris-Brosou & Excoffier,
488 1996; Nabholz, Glémin, & Galtier, 2009; Nabholz, Glémin, Galtier, Glemin, & Galtier, 2008), or
489 selection on mtDNA (e.g. Foote *et al.* 2011; Finch *et al.* 2014). In some social species, low mtDNA
490 diversity may be caused by cultural hitchhiking, a phenomenon in which mtDNA variation
491 changes through selection on maternally-transmitted cultural traits (Whitehead, 1998). Due to low
492 mtDNA diversity in short-finned pilot whales, the use of additional lines of data (mitogenomes
493 and SNPs) has improved our understanding of phylogeographic patterns and evolutionary
494 divergence within the species.

495

496 *Taxonomy and the subspecies hypothesis*

497 Following the guidelines for subspecies delineation summarized in Figure 3 of Taylor et
498 al. (2017a), we find support for two subspecies: a Shiho subspecies in the eastern Pacific Ocean,
499 and a Naisa subspecies encompassing the central/western Pacific and Indian Oceans as well as the
500 Atlantic Ocean. We propose a nominate subspecies, *Globicephala macrorhynchus*
501 *macrorhynchus*, with the common name "Naisa short-finned pilot whale" (currently called "ma-

502 gondo" in Japan), distributed throughout the central/western Pacific, Indian, and Atlantic Oceans,
503 and an unnamed subspecies with the common name "Shiho short-finned pilot whale" (currently
504 "tappa-naga" in Japan), found in the eastern Pacific Ocean and northern Japan (Supplemental
505 Figure S5). See the Supplemental Materials for further considerations of the proposed common
506 names. The holotype for *Globicephala macrorhynchus macrorhynchus* would be that previously
507 designated for *Globicephala macrorhynchus.*

508 We recommend the unnamed subspecies be designated according to one of the previously
509 synonymized names for *Globicephala macrorhynchus* (see Supplemental Materials for further
510 consideration). *Globicephala scammonii* (Cope, 1869) is likely to be the earliest taxonomic
511 designation for this subspecies, based on morphology and sample location, but this should be
512 confirmed through genetic sequencing. We recommend that the holotypes for *G. macrorhynchus*
513 (U.K. Natural History Museum, Accession #1846.8.9.2), *G. scammonii* (U.S. Natural History
514 Museum, Accession #USNM A 9074), and *G. sieboldii* (Naturalis Biodiversity Centre, Accession
515 #RMNH.MAM.21648) be sequenced and compared with the Shiho and Naisa subspecies to
516 resolve their taxonomic nomenclature.

517 There are two known regions of potential sympatry and/or introgression between the two
518 subspecies: coastal Japan and the eastern tropical Pacific Ocean. Off the coast of Japan, the two
519 subspecies are spatio-temporally and ecologically isolated, with the Naisa subspecies using the
520 warmer Kuroshio Current and the Shiho subspecies using the colder Oyashio Current (Kasuya et
521 al., 1988). Less information is available from the eastern tropical Pacific Ocean. Because of this,
522 we tested mitochondrial control region differentiation between the Naisa and Shiho geographic
523 strata in two ways – we first stratified Clade 3 samples according to geography, to test the

hypothesis of geographically-separated subspecies with some degree of admixture (Table 2). Second, we stratified the Clade 3 samples with the Naisa samples according to their mitogenomic classification, to test the hypothesis of genetically-differentiated subspecies with an area of sympatric distribution in the ETP and no genetic exchange (Supplemental Table S6). Both stratification schemes support subspecies delimitation.

In addition to control region support, our analysis of nuclear SNPs independently clustered samples into Naisa and Shiho groups. The high proportion of private alleles (54%) indicates that contemporary gene flow between the two strata is very low, or possibly zero (Slatkin, 1985). Although this analysis is representative of the global range of short-finned pilot whales, there are geographic areas where sampling is scant or missing, such as the southern Atlantic Ocean, pelagic Indian Ocean, and the eastern central Pacific Ocean area of potential sympatry.

Additional data from other studies further support this recommendation. Morphological data collected off the coast of Japan show that the Shiho and Naisa types differ in skull shape, body length, and color pattern (Kasuya, 2017; Kasuya et al., 1988; Miyazaki & Amano, 1994). Skull size (length and width) may be considered a diagnostic difference between the two types (Kasuya, 2017; Miyazaki & Amano, 1994), but the limited sample size and geographic coverage outside Japan (e.g. Polisini, 1980) prohibit the use of this trait for taxonomic analysis. There is a considerably greater sample size for body length measurements, expanding the range of geographic coverage to include Hawai‘i and the eastern tropical Pacific Ocean (Chivers et al., 2018). However, there is some overlap between the two types in the range of body length measurements; therefore body length cannot be considered a diagnostic trait (e.g. Cracraft, 1983; Helbig *et al.*, 2002; Sites and Marshall, 2004; De Queiroz, 2007). However, length measurements

546 do indicate a high level of concordance in geographic distribution between morphologically
547 recognized forms and mitochondrial haplotypes (Chivers et al., 2018; Oremus et al., 2009; Polisini,
548 1980; Van Cise et al., 2016). Where genetic samples have been sequenced from individuals of
549 known morphological form (Japan, Hawai‘i, and the eastern Pacific Ocean), concordance is 100%
550 between the two (Oremus et al., 2009; Van Cise et al., 2016).

551 Finally, a few localized studies indicate the potential for additional differences between the
552 two proposed subspecies. Where the two subspecies have allopatric distributions off the coast of
553 Japan, there are differences in their life-history parameters, such as peak mating season (Kasuya,
554 2017). Similarly, a study of vocal repertoires in the eastern Pacific Shiho subspecies and the Naisa
555 subspecies found in Hawai‘i indicated acoustic differentiation between the two (Van Cise et al.,
556 2017).

557 A conservative taxonomic approach requires additional data, particularly from regions of
558 sympatry, supporting complete diagnosibility of the two subspecies in order to classify them as
559 distinct species. Although we adhere to this conservative approach to species delimitation within
560 this study, it is important to remember the risk involved in the under-classification of taxonomic
561 units, especially with regard to conservation implications, as well as our scientific understanding
562 of basic biological and evolutionary processes (Bowen et al., 2016; Daugherty, Cree, Hay, &
563 Thompson, 1990; Leslie, 2015; Mace, 2004). In the case of short-finned pilot whales, and many
564 other cetaceans, the difficulty in obtaining a sufficient dataset of morphological or genetic samples
565 covering the entire range of the species may perpetuate taxonomic under-classification, with
566 consequences that range from failing to properly characterize the evolutionary trajectory within a
567 specific taxon or failing to detect recent speciation events, to the extinction of under-classified

568 species and loss of the associated evolutionary potential of that species (Allendorf & Luikart, 2011;
569 Daugherty et al., 1990; Taylor, Perrin, et al., 2017; Wang, Frasier, Yang, & White, 2008). It is
570 therefore our responsibility to consider, as we characterize and classify diversity, the potential for
571 under- or over-classification of certain taxa due to logistical or biological constraints, as well as
572 the trade-offs and consequences that may occur if our classification is not correct. In the case of
573 Naisa and Shiho short-finned pilot whales, we recommend that priority be given to generating
574 nuclear sequence data from areas of potential sympatry or introgression (i.e. coastal Japan and the
575 eastern tropical Pacific Ocean, see Figure 4), which can be used to assess gene flow and migration
576 between the two taxa and determine whether there is support to formally elevate these two
577 subspecies to species. Alternative methods for collecting morphological data, for example using
578 drone photography to determine body length or melon shape, should also be explored.

579 Within the Naisa subspecies, two populations in the Atlantic Ocean and central/western
580 Pacific and Indian Oceans may also be sufficiently distinct to be considered subspecies with further
581 sampling. We suggest that delimitation of an Atlantic Ocean subspecies would require additional
582 mitogenomic and nuclear data, or expanded morphological analyses, from the Atlantic and Indian
583 Oceans.

584

585 **Acknowledgements**

586 We are grateful to the many people and organizations who collected and contributed
587 samples for this study, and specifically to: Jooke Robbins, Provincetown Center for Coastal
588 Studies; Jay Barlow, Susan Chivers, Tim Gerodette, Lisa Ballance, Robert Pitman, Louella Dolar,
589 Marine Mammal and Sea Turtle Research (MMASTR) Collection, SWFSC; Tony Martinez and

590 field crews of the NOAA Ship *Gordon Gunter*, SEFSC and *Delware II*, NEFSC; Jeff Jacobsen,
591 Humboldt State University; Natacha Aguilar Soto, Emma Carroll, Scottish Oceans Institute,
592 University of St. Andrews; Kenny Travouillon & Rebecca Bray, Western Australian Museum;
593 Charles Anderson, Marine Research Section, Ministry of Fisheries and Agriculture, Republic of
594 the Maldives; Isabel Beasley, James Cook University. Leigh Hickmott and John Durban collected
595 samples contributed by the Bahamas Marine Mammal Research Organization. Samples were
596 collected under the following permits: NMFS MMPA permit 774-1714 issued to the SWFSC and
597 CNMI-DFW permit, license no. 01721-10 (2010 samples); NMFS MMPA permit 14097 issued to
598 the SWFSC and CNMI-DFW permit, license no. 02260-11 (2011 samples); NMFS MMPA permit
599 15240 issued to PIFSC and CNMI-DFW permit, license nos. 02444-12 and 02694-13 (2012 and
600 2013 samples). We thank NOAA's SWFSC Marine Mammal Genetics Group, especially Kelly
601 Robertson, Gabriela Serra-Valente, and Vicki Pease for the curation and processing of samples
602 used in this study. Samples provided by C.S. Baker were curated by Debbie Steel in Oregon and
603 Rochelle Constantine in Auckland. We thank Bill Perrin and Bob Brownell for their expert advice
604 regarding the taxonomic interpretation of results from this study. Funding for this project was
605 provided by Commander, U.S. Pacific Fleet Environmental Readiness Division and NMFS Pacific
606 Islands Fisheries Science Center; NMFS West Coast Region; Scripps Institution of Oceanography
607 Edna Bailey Sussman Research Fellowship; and Woods Hole Oceanographic Institution.

608 **References**

609 Allendorf, F. W., & Luikart, G. (2011). *Conservation and the Genetics of Populations*. Malden,
610 MA: Blackwell Publishing Ltd.

611 Alves, F., Quérouil, S., Dinis, A., Nicolau, C., Ribeiro, C., Freitas, L., … Fortuna, C. (2013).
612 Population structure of short-finned pilot whales in the oceanic archipelago of Madeira
613 based on photo-identification and genetic analyses: implications for conservation. *Aquatic*
614 *Conservation: Marine and Freshwater Ecosystems*, *23*(5), 758–776. doi:10.1002/aqc.2332

615 Andrews, K. R., Karczmarski, L., Au, W. W. L., Rickards, S. H., Vanderlip, C. A., Bowen, B.
616 W., … Toonen, R. J. (2010). Rolling stones and stable homes: social structure, habitat
617 diversity and population genetics of the Hawaiian spinner dolphin (*Stenella longirostris*).
618 *Molecular Ecology*, *19*(4), 732–48. doi:10.1111/j.1365-294X.2010.04521.x

619 Archer, F. I., Adams, P. E., & Schneiders, B. B. (2017). stratag: An R package for manipulating,
620 summarizing and analysing population genetic data. *Molecular Ecology Resources*, *17*(1),
621 5–11. doi:10.1111/1755-0998.12559

622 Aris-Brosou, S., & Excoffier, L. (1996). The impact of population expansion DNA and mutation
623 rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution*,
624 *13*(3), 494–504.

625 Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic & Physiological*
626 *Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *34*(5),
627 502–508. doi:10.1111/opo.12131

628 Baker, C. S., Perry, A., Bannister, J. L., Weinrich, M. T., Abernethy, R. B., Calambokidis, J., …
629 Vasquez, O. (1993). Abundant mitochondrial DNA variation and world-wide population
630 structure in humpback whales. *Proceedings of the National Academy of Sciences of the*
631 *United States of America*, *90*(17), 8239–43. doi:10.1073/PNAS.90.17.8239

632 Balcazar, N. E., Tripovich, J. S., Klinck, H., Nieukirk, S. L., Mellinger, D. K., Dziak, R. P., &
633 Rogers, T. L. (2015). Calls reveal population structure of blue whales across the southeast
634 Indian Ocean and southwest Pacific Ocean. *Journal of Mammalogy*, *96*(6), 1184–1193.
635 doi:10.1093/jmammal/gyv126

636 Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific
637 phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48.

638 Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S., …
639 Briggs, J. C. (2016). Comparative phylogeography of the ocean planet. *Proceedings of the*
640 *National Academy of Sciences*, *113*(29), 7962–7969. doi:10.1073/pnas.1602404113

641 Buonaccorsi, V. P., Mcdowell, J. R., & Graves, J. E. (2001). Reconciling patterns of inter-ocean
642 molecular variance from four classes of molecular markers in blue marlin (Makaira
643 nigricans). *Molecular Ecology*, *10*(5), 1179–1196. doi:10.1046/j.1365-294X.2001.01270.x

644 Cammen, K. M., Andrews, K. R., Carroll, E. L., Foote, A. D., Humble, E., Khudyakov, J. I., …
645 Van Cise, A. M. (2016). Genomic methods take the plunge: recent advances in high-
646 throughput sequencing of marine mammals. *Journal of Heredity*, *107*(6), 481–495.
647 doi:10.1093/jhered/esw044

648 Carroll, E. L., Baker, C. S., Watson, M., Alderman, R., Bannister, J., Gaggiotti, O. E., …
649 Harcourt, R. (2015). Cultural traditions across a migratory network shape the genetic
650 structure of southern right whales around Australia and New Zealand. *Scientific Reports*,
651 *5*(1), 16182. doi:10.1038/srep16182

652 Castro, A. L. F., Stewart, B. S., Hueter, R. E., Meekan, M. G., Motta, P. J., Bowen, B. W., &
653 Karl, S. A. (2007). Population genetic structure of Earth's largest fish, the whale shark
654 (Rhincodon typus). *Molecular Ecology*, *16*(24), 5183–5192. doi:10.1111/j.1365-

294X.2007.03597.x

Chabot, C. L. (2015). Microsatellite loci confirm a lack of population connectivity among globally distributed populations of the tope shark *Galeorhinus galeus* (Triakidae). *Journal of Fish Biology*, *87*(2), 371–385. doi:10.1111/jfb.12727

Chen, I., Nishida, S., Chou, L.-S., Tajima, Y., Yang, W.-C., Isobe, T., … Hoelzel, A. R. (2018). Concordance between genetic diversity and marine biogeography in a highly mobile marine mammal, the Risso's dolphin. *Journal of Biogeography*. doi:10.1111/jbi.13360

Chen, I., Yu, H., Yang, W., Nishida, S., Isobe, T., & Tanabe, S. (2014). The "Southern form" of short-finned pilot whale (*Globicephala macrorhynchus*) in tropical west Pacific Ocean off Taiwan. *Raffles Bulletin of Zoology*, *62*, 188–199.

Chivers, S. J., Perryman, W. L., Lynn, M. S., West, K., & Brownell, R. L. (2018). "Northern" form short-finned pilot whales (Globicephala macrorhynchus) inhabit the eastern tropical Pacific Ocean. *Aquatic Mammals*, *44*(4), 357–366. doi:10.1578/AM.44.4.2018.357

Chow, S., Jeffs, A., Miyake, Y., Konishi, K., Okazaki, M., Suzuki, N., … Sakai, M. (2011). Genetic Isolation between the Western and Eastern Pacific Populations of Pronghorn Spiny Lobster Panulirus penicillatus. *PLoS ONE*, *6*(12), e29280. doi:10.1371/journal.pone.0029280

Cracraft, J. (1983). Species concepts and speciation analysis. In *Current Ornithology* (pp. 159–187). Boston, MA: Springer US. doi:10.1007/978-1-4615-6781-3_6

Darriba, D., Taboada, G., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.

Daugherty, C. H., Cree, A., Hay, J. M., & Thompson, M. B. (1990). Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature*, *347*(6289), 177–179. doi:10.1038/347177a0

Davies, J. L. (1963). The antitropical factor in cetacean speciation. *Evolution*, *17*(1), 107–116. doi:10.1016/0011-7471(64)90215-3

De Queiroz, K. (2007). Species Concepts and Species Delimitation. *Systematic Biology*, *56*(6), 879–886. doi:10.1080/10635150701701083

Drummond, A. J., Suchard, M. a., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. doi:10.1093/molbev/mss075

Dwyer, G. S., Cronin, T. M., Baker, P. A., Raymo, M. E., Buzas, J. S., & Corrège, T. (1995). North Atlantic Deepwater Temperature Change During Late Pliocene and Late Quaternary Climatic Cycles. *Science*, *270*(5240), 1347–1351. doi:10.1126/science.270.5240.1347

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–20. doi:10.1111/j.1365-294X.2005.02553.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*,

694 *9*(4). doi:10.1371/journal.pone.0092587

695 Findlay, K. P., Best, P. B., Ross, G. J. B., & Cockcroft, V. G. (1992). The distribution of small
696 odontocete cetaceans off the coasts of South Africa and Namibia. *South African Journal of*
697 *Marine Science*, *12*(1), 237–270. doi:10.2989/02577619209504706

698 Foote, A. D., Morin, P. A., Durban, J. W., Pitman, R. L., Wade, P., Willerslev, E., … da
699 Fonseca, R. R. (2011). Positive selection on the killer whale mitogenome. *Biology Letters*,
700 *7*(1), 116–118. doi:DOI 10.1098/rsbl.2010.0638

701 Foote, A. D., Vijay, N., Avila-Arcos, M., Baird, R., Durban, J., Morin, P., … Wolf, J. (2016).
702 Genome-culture coevolution promotes rapid divergence in the killer whale. *Nature*
703 *Communications*, *7*(11693), 1–12. doi:10.1101/040295

704 Fountain, E. D., Pauli, J. N., Reid, B. N., Palsbøll, P. J., & Peery, M. Z. (2016). Finding the right
705 coverage: the impact of coverage and sequence quality on single nucleotide polymorphism
706 genotyping error rates. *Molecular Ecology Resources*, *16*(4), 966–978. doi:10.1111/1755-
707 0998.12519

708 Gaither, M. R., Bowen, B. W., Rocha, L. A., & Briggs, J. C. (2016). Fishes that rule the world:
709 circumtropical distributions revisited. *Fish and Fisheries*, *17*(3), 664–679.
710 doi:10.1111/faf.12136

711 Garrick, R. C., Sunnucks, P., & Dyer, R. J. (2010). Nuclear gene phylogeography using PHASE:
712 dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation.
713 *BMC Evolutionary Biology*, *10*(1), 118. doi:10.1186/1471-2148-10-118

714 Graves, J. E., & McDowell, J. R. (1995). Inter-ocean genetic divergence of istiophorid billfishes.
715 *Marine Biology*, *122*(2), 193–203. doi:10.1007/BF00348932

716 Guindon, S., & Gascuel, O. (2003). A simple, fast and accurate method to estimate large
717 phylogenies by maximum-likelihood. *Systematic Biology*, *52*, 696–704.

718 Hamilton, T. A., Redfern, J. V, Barlow, J., Ballance, L. T., Gerrodette, T., Holt, R. S., … Taylor,
719 B. L. (2009). *Atlas of cetacean sightings for Southwest Fisheries Science Center cetacean*
720 *and ecosystem surveys: 1986 - 2005. NOAA Technical Memoradum NOAA-TM-NMFS-*
721 *SWFSC-440 U.S.* (NOAA-NMFS, Ed.).

722 Hamner, R. M., Oremus, M., Stanley, M., Brown, P., Constantine, R., & Baker, C. S. (2012).
723 *Estimating the abundance and effective population size of Maui's dolphins using*
724 *microsatellite genotypes in 2010–11, with retrospective matching to 2001–07*.

725 Hancock-Hanser, B. L., Frey, A., Leslie, M. S., Dutton, P. H., Archer, F. I., & Morin, P. A.
726 (2013). Targeted multiplex next-generation sequencing: advances in techniques of
727 mitochondrial and nuclear DNA sequencing for population genomics. *Molecular Ecology*
728 *Resources*, *13*(2), 254–68. doi:10.1111/1755-0998.12059

729 Hasegawa, M., Kishino, H., & Yano, T. A. (1985). Dating of the human-ape splitting by a
730 molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*(2), 160–174.
731 doi:10.1007/BF02101694

732 Helbig, A. J., Knox, A. G., Parkin, D. T., Sangster, G., & Collinson, M. (2002). Guidelines for
733 assigning species rank. *Ibis*, *144*(3), 518–525. doi:10.1046/j.1474-919X.2002.00091.x

734 Hill, M. C., Bendlin, A. R., Van Cise, A. M., Milette-Winfree, A., Ligon, A. D., Ü, A. C., …
735 Oleson, E. M. (2018). Short-finned pilot whales (Globicephala macrorhynchus) of the
736 Mariana Archipelago: Individual affiliations, movements, and spatial use. *Marine Mammal*
737 *Science*. doi:10.1111/mms.12567

738 Hill, M. C., Oleson, E. M., Ligon, A. D., Martien, K. K., Archer, F. I., Baumann-Pickering, S.,
739 … Woodman, S. M. (2015). *Cetacean Monitoring in the Mariana Islands Range Complex,*
740 *2014. Prepared for the U.S. Pacific Fleet Environmental Readiness Office. PIFSC Data*
741 *Report DR-15-003*.

742 Hoelzel, A. R., Fleischer, R. C., Campagna, C., Alvord, G., & Le Boeuf, B. J. (2002). Impact of
743 a population bottleneck on symmetry and genetic diversity in the northern elephant seal
744 (*Mirounga angustirostris*). *Journal of Evolutionary Biology*, *15*, 567–575.

745 Hutchings, L., van der Lingen, C. D., Shannon, L. J., Crawford, R. J. M., Verheye, H. M. S.,
746 Bartholomae, C. H., … Monteiro, P. M. S. (2009). The Benguela Current: An ecosystem of
747 four components. *Progress in Oceanography*, *83*(1–4), 15–32.
748 doi:10.1016/J.POCEAN.2009.07.046

749 Jakobsson, M., & Rosenberg, N. a. (2007). CLUMPP: a cluster matching and permutation
750 program for dealing with label switching and multimodality in analysis of population
751 structure. *Bioinformatics*, *23*(14), 1801–1806. doi:10.1093/bioinformatics/btm233

752 Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers.
753 *Bioinformatics*, *24*(11), 1403–1405. doi:10.1093/bioinformatics/btn129

754 Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide
755 SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi:10.1093/bioinformatics/btr521

756 Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., … Kang,
757 H. M. (2012). Detecting and Estimating Contamination of Human DNA Samples in
758 Sequencing and Array-Based Genotype Data. *The American Journal of Human Genetics*,
759 *91*(5), 839–848. doi:10.1016/j.ajhg.2012.09.004

760 Kasuya, T. (2017). *Small cetaceans of Japan: exploitation and biology*. CRC Press, Inc.

761 Kasuya, T., Miyashita, T., & Kasamatsu, F. (1988). Segregation of two forms of short-finned
762 pilot whales off the Pacific coast of Japan. *The Scientific Reports of the Whales Research*
763 *Intitute*, *39*, 77–90.

764 Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and Their Applications*, *13*(3),
765 235–248. doi:10.1016/0304-4149(82)90011-4

766 Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in
767 multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3.
768 doi:10.1093/nar/gkr771

769 Leigh, J. W., & Bryant, D. (2015). popart : full-feature software for haplotype network
770 construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. doi:10.1111/2041-
771 210X.12410

772 Leslie, M. S. (2015). Impacts of phylogenetic nomenclature on the efficacy of the U.S.
773 Endangered Species Act. *Conservation Biology*, *29*(1), 69–77. doi:10.1111/cobi.12375

774 Leslie, M. S., & Morin, P. A. (2016). Using Genome-Wide SNPs to Detect Structure in High-
775 Diversity and Low-Divergence Populations of Severely Impacted Eastern Tropical Pacific
776 Spinner (*Stenella longirostris*) and Pantropical Spotted Dolphins (*S. attenuata*). *Frontiers in*
777 *Marine Science*, *3*(December), 1–12. doi:10.3389/fmars.2016.00253

778 Leslie, M. S., & Morin, P. A. (2018). Structure and phylogeography of two tropical predators,
779 spinner (*Stenella longirostris*) and pantropical spotted (*S. attenuata*) dolphins, from SNP
780 data. *Royal Society Open Science*, *5*, 171615.

781 Lessios, H. A. (2008). The Great American Schism: Divergence of Marine Organisms After the
782 Rise of the Central American Isthmus. *Annual Review of Ecology, Evolution, and*
783 *Systematics*, *39*(1), 63–91. doi:10.1146/annurev.ecolsys.38.091206.095815

784 Lessios, H. A., & Robertson, D. R. (2006). Crossing the impassable: genetic connections in 20
785 reef fishes across the eastern Pacific barrier. *Proceedings. Biological Sciences*, *273*(1598),
786 2201–8. doi:10.1098/rspb.2006.3543

787 Mace, G. M. (2004). The role of taxonomy in species conservation. *Philosophical Transactions*
788 *of the Royal Society of London. Series B, Biological Sciences*, *359*(1444), 711–9.
789 doi:10.1098/rstb.2003.1454

790 Mahaffy, S. D., Baird, R. W., McSweeney, D. J., Webster, D. L., & Schorr, G. S. (2015). High
791 site fidelity, strong associations, and long-term bonds: Short-finned pilot whales off the
792 island of Hawai'i. *Marine Mammal Science*, *31*(4), 1427–1451. doi:10.1111/mms.12234

793 Martien, K. K., Chivers, S. J., Baird, R. W., Archer, F. I., Gorgone, A. M., Hancock-Hanser, B.
794 L., … Taylor, B. L. (2014). Nuclear and Mitochondrial Patterns of Population Structure in
795 North Pacific False Killer Whales (*Pseudorca crassidens*). *Journal of Heredity*, *105*, 611–
796 626. doi:10.5061/dryad.2pq32

797 Martien, K. K., Hill, M. C., Van Cise, A. M., Roberston, K. M., Woodman, S. M., Dolar, L., …
798 Oleson, E. M. (2014). *Genetic diversity and population structure in four species of*
799 *cetaceans around the Mariana Islands* (Vol. NMFS-SWFSC). La Jolla, CA.

800 Miyazaki, N., & Amano, M. (1994). Skull morphology of two forms of short-finned pilot whales
801 off the Pacific coast of Japan. *Reports of the International Whaling Commission*, *44*, 499–
802 508.

803 Morin, P. A., Archer, F. I., Pease, V. L., Hancock-hanser, B. L., Kelly, M., Huebinger, R. M., …
804 Taylor, B. L. (2012). An empirical comparison of SNPs and microsatellites for population
805 structure, assignment, and demographic analyses of bowhead whale populations.
806 *Endangered Species Research*, *19*(2), 1–27. doi:10.3354/esr00459

807 Morin, P. A., Foote, A. D., Baker, C. S., Hancock-Hanser, B. L., Kaschner, K., Mate, B. R., …
808 Alexander, A. (2018). Demography or selection on linked cultural traits or genes?
809 Investigating the driver of low mtDNA diversity in the sperm whale using complementary
810 mitochondrial and nuclear genome analyses. *Molecular Ecology*, *27*(11), 2604–2619.
811 doi:10.1111/mec.14698

812 Morin, P. A., Parsons, K. M., Archer, F. I., Ávila-Arcos, M. C., Barrett-Lennard, L. G., Dalla
813 Rosa, L., … Foote, A. D. (2015). Geographical and temporal dynamics of a global radiation

814 and diversification in the killer whale. *Molecular Ecology*, *24*(15), 3964–3979.
815 doi:10.1111/mec.13284

816 Nabholz, B., Glémin, S., & Galtier, N. (2009). The erratic mitochondrial clock: Variations of
817 mutation rate, not population size, affect mtDNA diversity across birds and mammals. *BMC*
818 *Evolutionary Biology*, *9*(1), 1–13. doi:10.1186/1471-2148-9-54

819 Nabholz, B., Glémin, S., Galtier, N., Glemin, S., & Galtier, N. (2008). Strong variations of
820 mitochondrial mutation rate across mammals - the longevity hypothesis. *Molecular Biology*
821 *and Evolution*, *25*(1), 120–130. doi:DOI 10.1093/molbev/msm248

822 Nei, M. (1987). *Molecular evolutionary genetics*. New York, NY: Columbia University Press.

823 Oremus, M., Gales, R., Dalebout, M. L., Funahashi, N., Endo, T., Kage, T., … Baker, S. C.
824 (2009). Worldwide mitochondrial DNA diversity and phylogeography of pilot whales
825 (*Globicephala spp.*). *Biological Journal of the Linnean Society*, *98*, 729–744.
826 doi:10.1111/j.1095-8312.2009.01325.x

827 Pazmiño, D. a., Maes, G. E., Green, M. E., Simpfendorfer, C. a., Hoyos-Padilla, E. M., Duffy, C.
828 J. A., … van Herwerden, L. (2018). Strong trans-Pacific break and local conservation units
829 in the Galapagos shark (Carcharhinus galapagensis) revealed by genome-wide cytonuclear
830 markers. *Heredity*, *120*(5), 407–421. doi:10.1038/s41437-017-0025-2

831 Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139), 1236–
832 1238. doi:10.1136/bmj.316.7139.1236

833 Perrin, W. F. (2007). *The South African Species Gate*. La Jolla, CA: NMFS Southwest Fisheries
834 Science Center.

835 Podos, J. (2010). Acoustic discrimination of sympatric morphs in Darwin's finches: a
836 behavioural mechanism for assortative mating? *Philosophical Transactions of the Royal*
837 *Society of London. Series B, Biological Sciences*, *365*(1543), 1031–1039.
838 doi:10.1098/rstb.2009.0289

839 Polisini, J. (1980). *A comparison of Globicephala macroryncha (Gray, 1846) with the pilot*
840 *whale of the North Pacific Ocean: An analysis of the skull of the broad-rostrum pilot*
841 *whales of the genus Globicephala. (PhD Thesis)*. University of Southern California.

842 Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using
843 multilocus genotype data. *Genetics*, *155*(2), 945–59.

844 R Core Team. (2016). R: A Language and Environment for Statistical Computing. Vienna,
845 Austria.

846 Rambout, A., Suchard, M., Xie, D., & Drummond, A. (2014). Tracer v1.6, , Available from
847 http://tree.bio.ed.ac.uk/software/tracer/.

848 Rendell, L., Mesnick, S. L., Dalebout, M. L., Burtenshaw, J., & Whitehead, H. (2012). Can
849 Genetic Differences Explain Vocal Dialect Variation in Sperm Whales, *Physeter*
850 *macrocephalus*? *Behavior Genetics*, *42*(2), 332–43. doi:10.1007/s10519-011-9513-y

851 Rocha, L. A., Craig, M. T., & Bowen, B. W. (2007). Phylogeography and the conservation of
852 coral reef fishes. *Coral Reefs*, *26*(3), 501–512. doi:10.1007/s00338-007-0261-7

Rosel, P. E., Hancock-Hanser, B. L., Archer, F. I., Robertson, K. M., Martien, K. K., Leslie, M. S., … Taylor, B. L. (2017). Examining metrics and magnitudes of molecular genetic differentiation used to delimit cetacean subspecies based on mitochondrial DNA control region sequences. *Marine Mammal Science*, *33*(June), 76–100. doi:10.1111/mms.12410

Sites, J. W., & Marshall, J. C. (2004). Operational Criteria for Delimiting Species. *Annual Review of Ecology, Evolution, and Systematics*, *35*(1), 199–227. doi:10.1146/annurev.ecolsys.35.112202.130128

Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, *39*(1), 53–65. doi:10.2307/2408516

Smith, A. L., & Friesen, V. L. (2007). Differentiation of sympatric populations of the band-rumped storm-petrel in the Galapagos Islands: an examination of genetics, morphology, and vocalizations. *Molecular Ecology*, *16*(8), 1593–603. doi:10.1111/j.1365-294X.2006.03154.x

Stephens, M., & Donnelly, P. (2003). A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet*, *73*(2002), 1162–1169. doi:10.1086/379378

Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, *10*(3), 512–26.

Taylor, B. L., Archer, F. I., Martien, K. K., Rosel, P. E., Hancock-Hanser, B. L., Lang, A. R., … Baker, C. S. (2017). Guidelines and quantitative standards to improve consistency in cetacean subspecies and species delimitation relying on molecular genetic data. *Marine Mammal Science*, *33*(June), 132–155. doi:10.1111/mms.12411

Taylor, B. L., Perrin, W. F., Reeves, R. R., Rosel, P. E., Wang, J. Y., Cipriano, F., … Brownell, R. L. (2017). Why we should develop guidelines and quantitative standards for using genetic data to delimit subspecies for data-poor organisms like cetaceans. *Marine Mammal Science*, *33*(S1), 12–26. doi:10.1111/mms.12413

van Bree, P. J. H., Best, B. B., & Ross, G. J. B. (1978). Occurrence of the two species of pilot whales (genus *Globicephala*) on the coast of South Africa. *Mammalia*, *42*(1872), 323–328.

Van Cise, A. M., Martien, K. K., Mahaffy, S. D., Baird, R. W., Webster, D. L., Fowler, J. H., … Morin, P. A. (2017). Familial social structure and socially driven genetic differentiation in Hawaiian short-finned pilot whales. *Molecular Ecology*, *26*(23), 6730–6741. doi:10.1111/mec.14397

Van Cise, A. M., 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 insight into the distribution and diversity of short-finned pilot whales in the Pacific Ocean. *Marine Mammal Science*, *32*(October), 1177–1199. doi:10.1111/mms.12315

Van Cise, A. M., Morin, P. A., Baird, R. W., Oleson, E. M., & Martien, K. K. (2016). *Relatedness and genetic structure in island-associated and pelagic short-finned pilot whales. Appendix in Hill et al. 2016. Cetacean Monitoring in the Mariana Islands Range*

Complex, 2015. Prepared for the U.S. Pacific Fleet Environmental Readiness Office.

Van Cise, A. M., Roch, M. A., Baird, R. W., Mooney, T. A., & Barlow, J. (2017). Acoustic differentiation of Shiho- and Naisa-type short-finned pilot whales in the Pacific Ocean. *The Journal of the Acoustical Society of America*, *141*(2), 737–748. doi:10.1121/1.4974858

Verity, R., Nichols, R. A., Baele, G., Lemey, P., Bedford, T., Rambaut, A., … Wright, S. (2016). Estimating the Number of Subpopulations (K) in Structured Populations. *Genetics*, *203*(4), 367–374. doi:10.1534/genetics.115.180992

Viñas, J., Alvarado Bremer, J. R., & Pla, C. (2004). Inter-oceanic genetic differentiation among albacore (Thunnus alalunga) populations. *Marine Biology*, *145*(2). doi:10.1007/s00227-004-1319-5

Wang, J. Y., Frasier, T. R., Yang, S. C., & White, B. N. (2008). Detecting recent speciation events: The case of the finless porpoise (genus *Neophocaena*). *Heredity*, *101*(2), 145–155. doi:10.1038/hdy.2008.40

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*(4), 1016–1020. doi:10.1093/molbev/msw279

Weber, D. S., Stewart, B. S., & Lehman, N. (2004). Genetic Consequences of a Severe Population Bottleneck in the Guadalupe fur Seal (*Arctocephalus townsendi*). *Journal of Heredity*, *95*(2), 144–153. doi:10.1093/jhered/esh018

Whitehead, H. (1998). Cultural Selection and Genetic Diversity in Matrilineal Whales. *Science*, *282*(5394), 1708–1711. doi:10.1126/science.282.5394.1708

Witteveen, B. A., Straley, J. M., Chenoweth, E., Baker, C. S., Barlow, J., Matkin, C., … Hirons, A. (2011). Using movements, genetics and trophic ecology to differentiate inshore from offshore aggregations of humpback whales in the Gulf of Alaska. *Endangered Species Research*, *14*, 217–226.

Yamase, H. (1760). *Geishi [Natural History of Whales]*. Osakashorin, Osaka.

Yonekura, M., Matsui, S., & Kasuya, T. (1980). On the external characters of *Globicephala macrorynchus* off the Taiji, Pacific coast of Japan. *Scientific Reports of the Whales Research Institute*, *32*(32), 67–95.

Yoshino, H., Armstrong, K. N., Izawa, M., Yokoyama, J., & Kawata, M. (2008). Genetic and acoustic population structuring in the Okinawa least horseshoe bat: are intercolony acoustic differences maintained by vertical maternal transmission? *Molecular Ecology*, *17*(23), 4978–91. doi:10.1111/j.1365-294X.2008.03975.x

Data Accessibility

All mitogenome, control region, and nuclear sequences are archived at NCBI GenBank. Mitogenome accession numbers are XXXX-XXXX. Control region accession numbers are XXXX-XXXX. Nuclear sequence accession numbers are XXXX-XXXX. SNP data are archived in the Dryad data repository: XXXX.

Author Contributions

AVC contributed to research design, lab analyses, data analysis, and paper writing. PAM contributed to research design and paper writing. Samples integral to the study were contributed by RWB, CSB, SC, DC, RF, JM, KKM, AAMG, EMO, MO, MP, PER, and BLT, all of whom also read and provided feedback on manuscript drafts. BHH contributed to lab analyses and manuscript editing. Significant editorial contributions were made by CSB, KKM, PER, and BLT.

939 Table 1. Molecular diversity indices for mitogenomes, mtDNA control region (345 bp), and nDNA
940 SNP (n=47) datasets. The SNP dataset includes 112 SNPs at 47 unique loci. *N* = mtDNA and SNP
941 sample size, θ_H = Theta, π = nucleotide diversity, H_o = observed heterozygosity, H_e = expected
942 heterozygosity.
943

Stratum	***N***	$\boldsymbol{\theta_H}$	$\boldsymbol{\pi}$	**Haplotype Diversity**	**No. Haplotypes**	**Ave. num alleles**	$\mathbf{H_o}$	$\mathbf{H_e}$
Mitogenome sequences								
All samples	181	0.73	0.002	0.98	97	-	-	-
Naisa clade	77	0.69	0.0008	0.92	35	-	-	-
Shiho clade	43	0.68	0.0003	0.91	26	-	-	-
Clade 3	36	0.71	0.0005	0.96	22	-	-	-
Atlantic clade	25	0.69	0.0007	0.92	14	-	-	-
Control Region sequences								
STRUCTURE-derived nuclear group samples only								
All samples	105	0.69	0.007	0.93	46	4.4	0.40	0.46
Naisa group	69	0.66	0.007	0.88	28	4.1	0.43	0.45
Shiho group	14	0.38	0.002	0.51	5	2.2	0.27	0.27
Atlantic group	22	0.72	0.003	0.97	17	2.7	0.36	0.37
Proposed type								
All samples	725	0.62	0.008	0.83	64	-	-	-
Naisa type	485	0.54	0.006	0.72	38	-	-	-
Shiho type	190	0.28	0.002	0.38	13	-	-	-
Atlantic type	50	0.60	0.002	.081	17	-	-	-

944

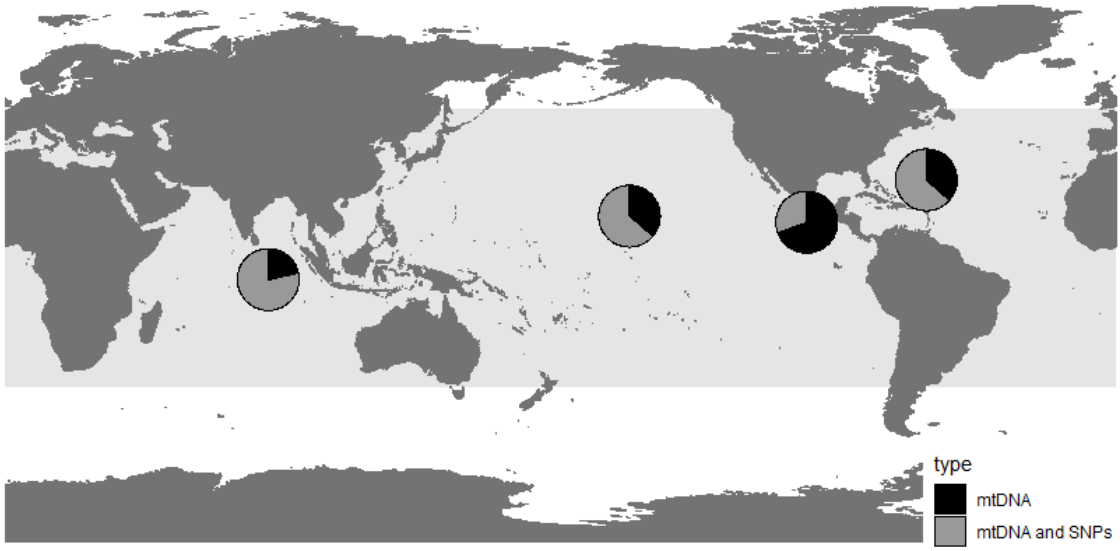
Table 2. Estimates of pairwise genetic differentiation between STRUCTURE-derived nuclear groups and geographically-defined hypothesized types. mtDNA control region differentiation was estimated using Φ_{ST}, net divergence (d_A), and percent diagnosibility (PD). The magnitude of nuclear SNP differentiation was estimated using F_{ST}. Sample sizes for each stratum are shown in parentheses. Significant *P*-values are shown in bold. The lower bounds of the subspecies threshold for d_A and PD are 0.004 and 95%, respectively. The lower bounds of the species threshold for d_A and PD are 0.02, and 95%, respectively.

Stratum	F_{ST}	Φ_{ST}	Φ_{ST} ***P*-value**	d_A	d_A **95% CI**	***PD***
Control region sequences						
STRUCTURE-derived nuclear group samples only						
Naisa (69) vs. Shiho (22)	-	0.32	**<0.001**	0.004	0.003-0.005	98.75%
Atlantic (14) vs. Shiho (22)	-	0.10	**0.04**	0.003	0.002-0.003	96.6%
Atlantic (14) vs. Naisa (69)	-	-0.01	0.47	0.002	0.001-0.003	73.17%
Proposed type						
Naisa (485) vs. Shiho (190)	-	0.46	**<0.001**	0.006	0.005 – 0.006	97.9%
Atlantic (50) vs. Shiho (190)	-	0.31	**<0.001**	0.004	0.003-0.004	99.2%
Atlantic (50) vs. Naisa (485)	-	0.15	**<0.001**	0.005	0.004-0.005	82.3%
SNPs						
STRUCTURE-derived nuclear group samples only						
Naisa (69) vs. Shiho (14)	0.3	-	-	-	-	-
Atlantic (22) vs. Shiho (14)	0.4	-	-	-	-	-
Atlantic (22) vs. Naisa (69)	0.1	-	-	-	-	-

Table 3. Estimates of pairwise genetic differentiation among mitogenomic clades within geographic regions with multiple clades, and vice versa. The magnitude of SNP differentiation was estimated using F_{ST}. Mitochondrial control region differentiation was estimated using Φ_{ST}, Nei's d_A, and PD. IO = Indian Ocean. Sample sizes for each stratum are shown in parentheses. Significant values are shown in bold. SNP comparisons with Clade 3 in the eastern Pacific could not be conducted due to small sample size within that region (n=2).

Strata	F_{ST}	Φ_{ST}	Φ_{ST} ***P*-value**	d_A	d_A **95% CI**	***PD***
Control region sequences						
Clades w/in Geographic Regions						
W. Pacific/IO: Naisa (398) vs. Clade 3 (62)	-	0.03	**0.020**	0.010	0.009-0.011	99.3%
E. Pacific: Shiho (172) vs. Clade 3 (17)	-	0.84	**<0.001**	0.006	0.006-0.006	99.5%
Regions w/in Clades						
Clade 3: E. Pacific (17) vs. W. Pacific/Indian Ocean (62)	-	0.73	**<0.001**	0.002	0.002-0.003	96.0%
SNPs						
Clades w/in Geographic Regions		-	-	-	-	-
W. Pacific/Indian Ocean: Naisa (52) vs. Clade 3 (19)	0.008	-	-	-	-	-

964 Figure 1. Global sampling locations of samples used to generate mitogenome and SNP sequences
965 used in this study. The shaded region indicates the general global distribution of short-finned pilot
966 whales. A detailed map of sample distribution can be found in Supplemental Figure S6.
967



968

969

Figure 2. BEAST phylogenetic tree of mitogenome data, rooted with *G. melas*, showing four distinct clades. The x-axis is Kya. The posterior probability of each branch is shown above the branch, on a scale from 0 to 1. Each branch represents a mitogenome haplotype, which may be shared by multiple individuals. Mitogenome haplotype labels can be seen in Supplemental Figure S2, and mitogenome haplotype frequencies can be found in Supplemental Table S4. The vertical bar on the far right shows the ocean basin where each haplotype was found.

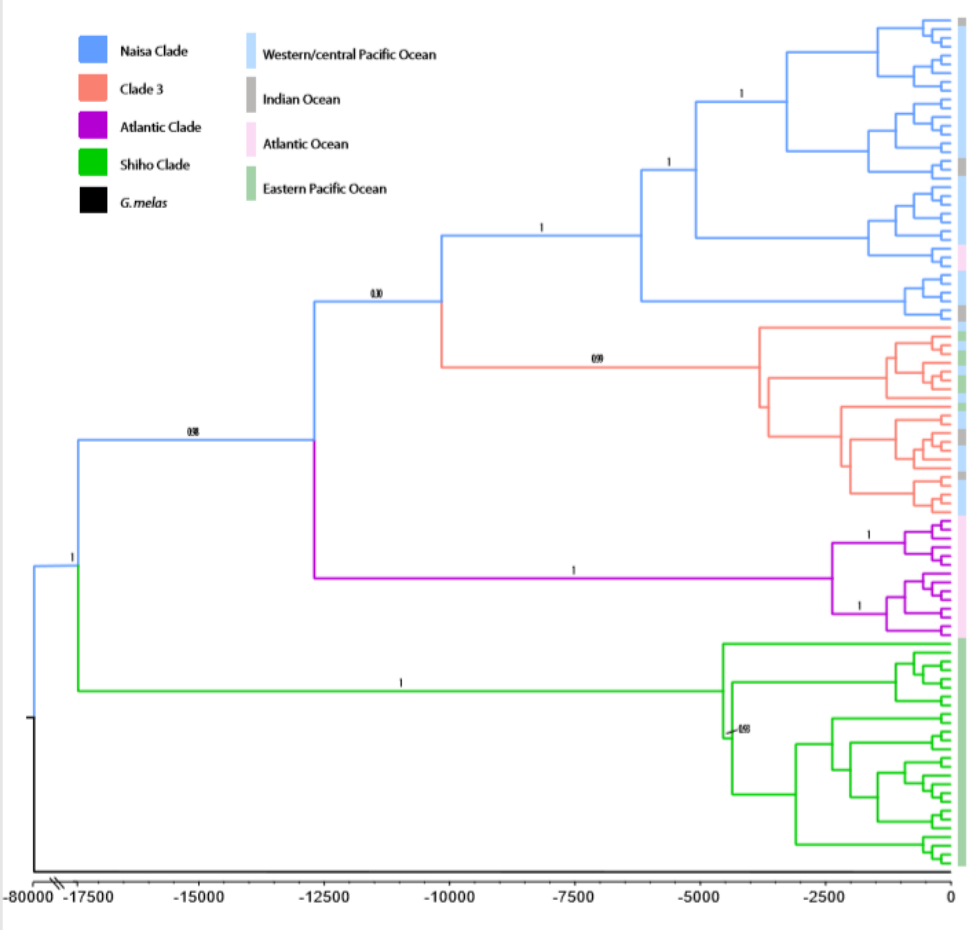


Figure 3. Median joining network (MJN) displaying the relationships among whole mitogenome haplotypes by ocean basin and mitogenomic clades. Circles are proportional in size to the number of samples with each haplotype. Cross hatches on lines indicate the number of differences between haplotypes. Missing haplotypes are indicated by a black node.

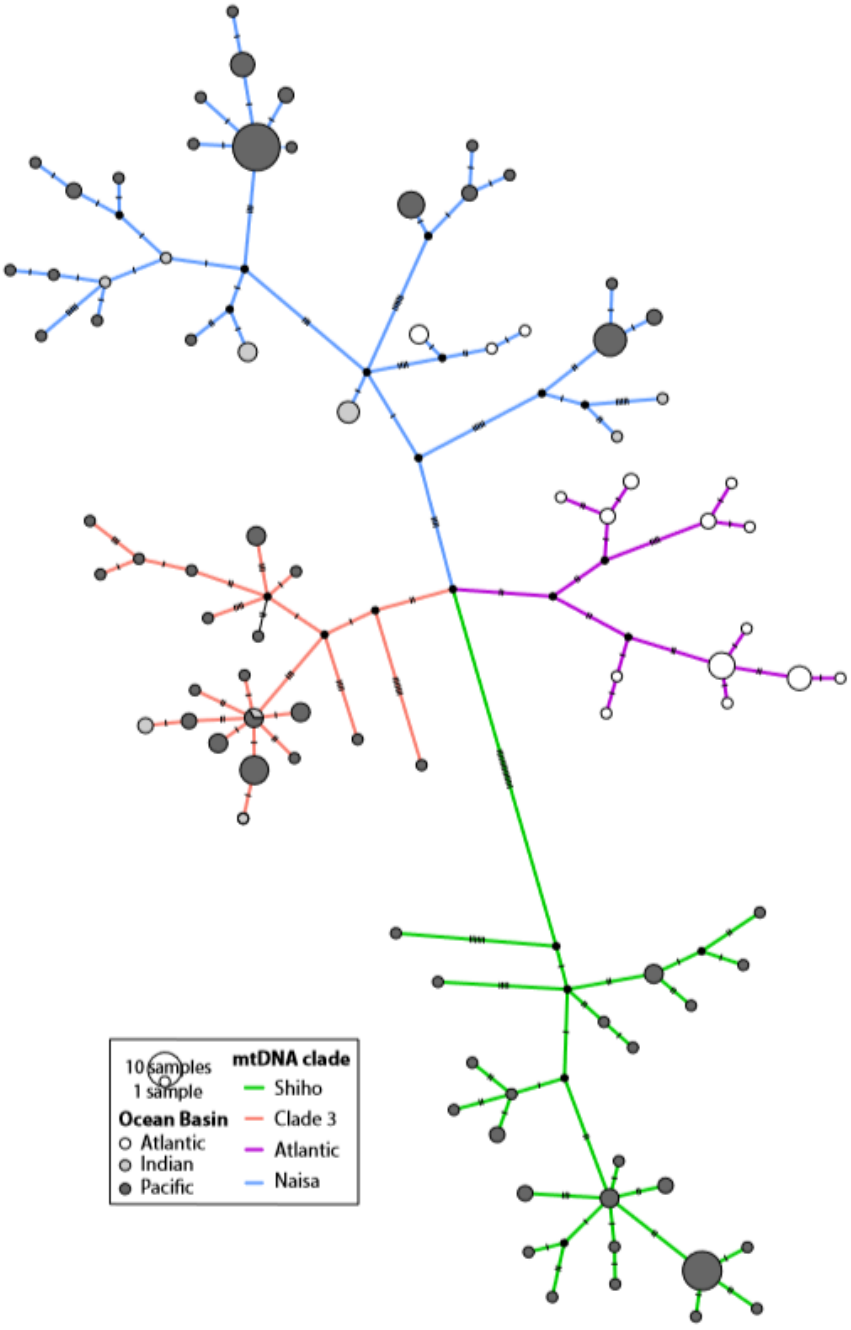
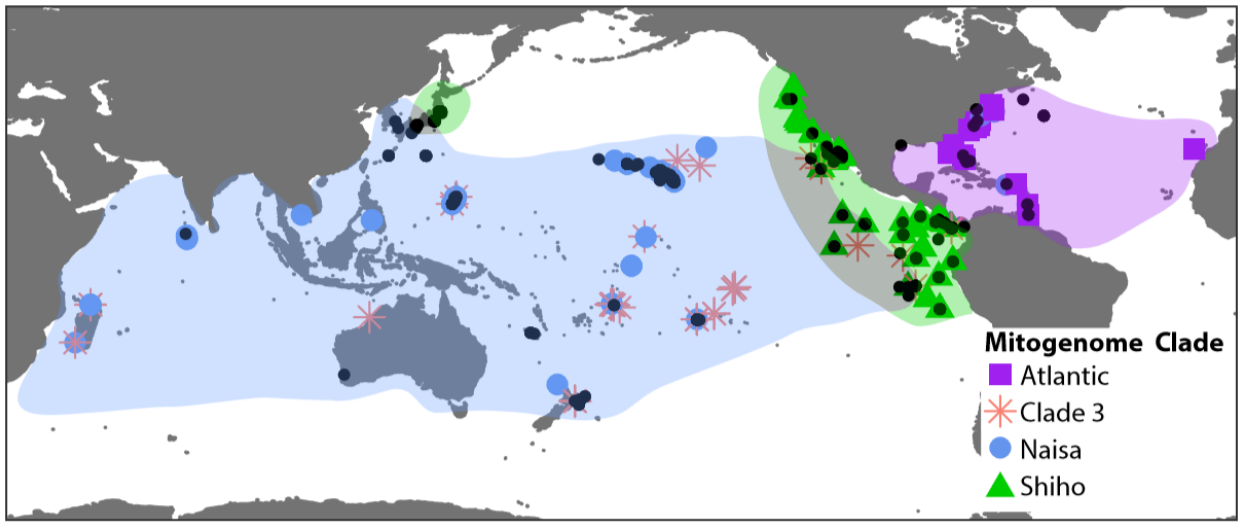


Figure 4. Distributions of A) four mitogenomic clades (black points represent samples for which we have control region samples but no mitogenome or SNP data), and B) three nuclear groups (open shapes represent samples for which we have mitogenome but no nuclear data). Two samples in panel B, outlined in red, are the only eastern Pacific Clade 3 samples with nuclear DNA; one grouped with Naisa, and the other grouped with Shiho. Colored areas in each panel encompass regions of general geographic concordance between the mitogenome and nuclear datasets, which were used to stratify samples into three regions for the estimation of divergence and differentiation among the three hypothesized types. Grey areas show regions of possible sympatry, or regions of recent or historic introgression between types.



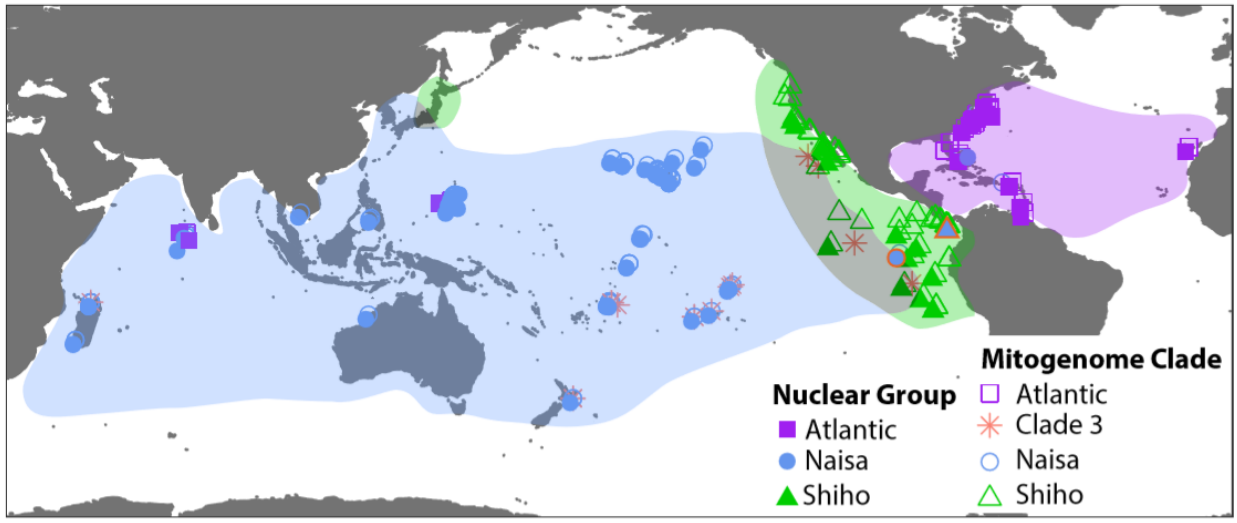


Figure 5. A) Assignment plot based on STRUCTURE analyses, with k = 3. Mitogenome clade stratification is on the x axis, and probability of assignment is on the y axis. B) Supervised Discriminant Analysis of Principal Components, colored according to the nuclear groups defined in the STRUCTURE analysis in Figure 5A.