geneSPLASH: An initial, ocean-wide survey of mitochondrial (mt) DNA diversity and population structure among humpback whales in the North Pacific

Final report for Contract 2006-0093-008

Principal Investigator: C. Scott Baker
Marine Mammal Institute
Oregon State University
2030 SE Marine Science Dr.
Newport, Oregon 97365
Email: scott.baker@oregonstate.edu

For the **National Fish and Wildlife Foundation**

with
Steel, Debbie[1], Calambokidis, John[5], Barlow, Jay[3], Burdin, Alexander M.[4], Clapham, Phillip J.[6], Falcone, Erin[5], Ford, John K. B.[7], Gabriele, Christine M.[8], González-Peral, Ursula[13], LeDuc, Rick[3], Mattila, David[9], Quinn, Terrance J.[10], Rojas-Bracho, Lorenzo[11], Straley, Janice M.[12], Taylor, Barbara L.[3], Urbán R., Jorge[13], Vant, Murdoch[2], Wade, Paul R.[6], Weller, David[3], Witteveen, Briana H.[14], Wynne, Kate[14], Yamaguchi, Manami[15]

(1) Marine Mammal Institute and Department of Fisheries and Wildlife, Oregon State University, 2030 SE Marine Science Drive, Newport, OR 97365 USA
(2) School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand
(3) NOAA Southwest Fisheries Science Center, 8604 La Jolla Shores Dr., La Jolla, CA 92037 USA
(4) Kamchatka Branch of Pacific Institute of Geography, Far East Branch of the Russian Academy of Sciences, Rybakov Prospect, 19-a Petropavlovsk-Kamchatsky 683024 Russia
(5) Cascadia Research Collective, 218½ West Fourth Avenue, Olympia, WA 98501 USA
(6) National Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115 USA
(7) Fisheries and Oceans Canada, Pacific Biological Station, 3190 Hammond Bay Road, Nanaimo, BC V9T 6N7 Canada
(8) Glacier Bay National Park, PO Box 140, Gustavus, AK 99826 USA
(9) Hawaiian Islands Humpback Whale National Marine Sanctuary, 726 S. Kihei Road, Kihei, HI 96753 USA
(10) Juneau Center for Fisheries and Ocean Sciences, University of Alaska, 11120 Glacier Highway, Juneau, AK 99801 USA
(11) Instituto Nacional de Ecologia, c/p CICESE, Apto. Postal 2732, Ensenada, Baja California 22860, México
(12) University of Alaska, Southeast Sitka Campus, 1332 Seward Avenue, Sitka, AK 99835 USA
(13) Programa de Mamíferos Marinos. Universidad Autónoma de Baja California Sur. La Paz, B.C.S. 23080 México
(14) *University of Alaska Fairbanks, School Fisheries & Ocean Sciences, 118 Trident Way, Kodiak, AK 99615 USA*
(15) Ogasawara Marine Center, Byobudani, Chichijima, Ogasawara-mura, Tokyo 100-21, Japan

Abstract
We report on initial results of a comprehensive, ocean-wide survey of mitochondrial (mt) DNA diversity and population structure among humpback whales in the North Pacific. Using n = 2,188 samples collected from 10 feeding and 8 breeding regions by the program *Structure of Populations, Levels of Abundance and Status of Humpbacks* (SPLASH) primarily in the winter and summer of 2004, we first used microsatellite genotyping (average, 9.5 loci) to identify replicate samples within regions and matches between regions. After review for quality control and removal of replicate samples, we identified n = 1,856 regional individuals, 33 of which demonstrated migratory movement between feeding and breeding regions based on genotype matching. Feeding regions showed no overall sex bias in sampled individuals (1.06:1, n = 1031), although a significant departure from unity was found for two regions, the Bering Sea (female bias) and Eastern Aleutians (male bias). Breeding grounds showed a strong overall male bias (2.19:1, n = 825), with the greatest bias found in Mexico-Baja (3.08:1, n = 118). From 500 bp of the mtDNA control region sequence representing each of the n = 1,856 regional individuals, we identified 28 unique haplotypes representing two divergent lineages or clades, one of which is thought to have originated from an historical connection with the Southern Hemisphere. Haplotype frequencies differed markedly among feeding regions (overall $F_{ST} = 0.179$, $p < 0.001$, n = 1031) and among breeding grounds (overall $F_{ST} = 0.106$, $p < 0.001$, n = 825), supporting previous characterization of strong maternal fidelity to migratory destinations. Among feeding regions, Russia, southeastern Alaska and California/Oregon, were notable for particularly high levels of differentiation from each other (up to $F_{ST} = 0.478$, $p < 0.001$). Among breeding regions, Okinawa and the Central American were notable for particularly high level of differentiation from each other ($F_{ST} = 0.454$, $p < 0.001$) and from most other breeding grounds. However, a comparison between feeding and breeding regions also showed a large number of significant differences, even for those regions known to be strongly connected by patterns of individual migration (e.g., by photo-identification, Calambokidis et al. 2008). Thus the influence of maternal fidelity seems to operate somewhat independently on feeding and breeding grounds over an evolutionary time scale, confounding a simple longitudinal division of the oceanic population into ‘stocks’. The potential to define multiple Genetic Management Units or Units to Conserve, on both feeding grounds and breeding grounds presents the most complex pattern of population structure yet described for large whales.

Introduction

Humpback whales in the North Pacific Ocean show a complex population structure as a result of seasonal migration and maternally directed fidelity to seasonal habitats. These whales feed during spring, summer and autumn in temperate and near-polar water along the rim of the North Pacific from central California in the east to the Kamchatka Peninsula in the west. During winter months, they migrate to near-tropical 'breeding grounds' to mate and give birth. Although hunting severely reduced the abundance of humpback whales and may have altered their historical range, winter concentrations are now found primarily in three regions (NMFS 1991):the Pacific coast and off-shore islands of Mexico; the main or leeward islands of Hawaii; and the Ogasawara (Bonin) and Ryukyu Islands (e.g., Okinawa) of Japan. More recently, smaller wintering concentrations have been found along the Central American coast in the east and the Philippines in the west.

The ocean-wide program on the *Structure of Populations, Levels of Abundance and Status of Humpbacks* (SPLASH) was initiated to estimate the current abundance of North Pacific humpback whales and to describe the seasonal structure of this population. SPLASH represents one of the largest studies of any population of living whales ever conducted, involving collaboration among 50 research groups and more than 400 researchers in 10 countries. Field efforts were conducted on all known winter breeding regions for humpback whales in the North Pacific during three seasons (2004, 2005, 2006) and all known summer feeding areas during two seasons (2004, 2005). A total of 18,469 quality fluke identification photographs were taken during over 27,000 approaches of humpback whales. After reconciling all within and cross-regional matches (from both the primary match and rechecks), a total of 7,971 unique individuals were cataloged in SPLASH. A total of 6,178 tissue samples were also collected for genetic studies of population structure.

The primary analysis of photo-identification records collected by SPLASH, including estimates of abundance and descriptions of individual migratory interchange, has now been completed and reported in Calambokidis et al (2008). Here, we report on the initial analysis of the genetic component of this program, referred to here as *gene*SPLASH. With funding from the National Fish and Wildlife Foundation and the Endowment of the Marine Mammal Institute, Oregon State University, we undertook to describe population structure of mtDNA control region sequence diversity using a representative subset of samples from 10 feeding and 8 breeding grounds derived primarily from the winter and summer of 2004. We chose mtDNA control region as the focus of our initial description of population structure because of its strictly maternal inheritance and relatively rapid rate of evolution. We predicted that the distribution of mtDNA diversity, in the form of distinct haplotypes, should reflect the influence of maternally-directed fidelity to seasonal habitats (Baker *et al.* 1990). Our goal was to include a minimum of 100 samples for each region, to provide accurate estimates of population differentiation using conventional analyses of molecular variance and permutation procedures (AMOVA, Excoffier 1995). Where regional sample sizes were less than 100 for the year 2004, we included additional samples from other SPLASH years, up to the total available for some regions (e.g., Russian and Aleutian feeding grounds; the Philippines, Okinawa and Central America). Supplemental funding from the Marine Acoustics Inc., allowed us to include all available samples from the Asian wintering grounds, including the larger samples from Ogasawara. To avoid any bias from replicate sampling, individual whales were first identified by microsatellite genotyping using up to 10

loci. Samples representing the same individual were omitted from within-region datasets but retained for between-region datasets (i.e., an individual found in two regions was included in both). Analyses to date have relied on the *a priori* grouping of samples according to the geographic divisions considered in estimates of abundance and migratory interchange for photo-identification records (Calambokidis *et al.* 2008). Results confirm the complex structure of migratory interchange and regional fidelity, providing new information for defining Units of Management or Units of Conservation within the oceanic population.

Methods

Sample collection and archiving

Skin samples (n = 6,178) were collected from humpback whales in 10 feeding grounds and 8 breeding regions of the North Pacific between 2004 and 2006 under the auspices of the program SPLASH (Calambokidis et al. 2008). Skin samples were collected primarily with a small stainless-steel biopsy dart deployed from a crossbow (Lambertsen 1987). In some regions, samples of sloughed skin were also collected (Amos *et al.* 1992). All samples were archived at the Southwest Fisheries Science Center (SWFSC), La Jolla, Ca, and assigned accession numbers ('z' codes) for cross-reference with the primary field records and photo-identification records stored with Cascadia Research.

A subset of the total SPLASH samples was chosen for initial analysis of mtDNA diversity and population differentiation (n = 2,188) with the objective of including a minimum of 100 individuals for each region (Table 1). For regions with large samples sizes, priority was given to use of 2004 samples. For other regions, samples from some or all other sampling periods were included. For Russia (RUS), Okinawa (OK), Ogasawara (OG), Philippines (PHI), Central America (CENTAM), California-Oregon (CA-OR), South British Colombia- Washington (SBC-WA), Eastern Aleutians (EAL) and Western Aleutians (WAL) all samples were used; for Bering (BER) all samples from 2004 and 2005 were used; for Western Gulf of Alaska (WGOA), Northern Gulf of Alaska (NGOA) and all 3 Mexican regions (Mainland MX-ML, Revillagigedo MX-AR and Baja California MX-BC) only 2004 samples were used; for Hawaii (HI) and Southeast Alaska (SEA) a subset of the 2004 samples was chosen at random.

DNA extraction, quantification and sex identification

Total genomic DNA was extracted at SWFSC and an aliquot was transferred to the Cetacean Conservation Genetic Laboratory (CCGL) of the Marine Mammal Institute, Oregon State University. At the CCGL, total genomic DNA was quantified using pico-green fluorescence and normalized to 5ng/ul before amplification, where possible. The sex of each sampled whale was identified by amplification of sex-specific markers following the protocol of Gilson et al. (Gilson *et al.* 1998). This involves a multiplex PCR with primers designed to amplify the male-specific *Sry* gene and, as positive controls, primers designed to amplify the *ZFY/ZFX* genes of males and females.

mtDNA sequencing and quality control

An approximately 800 base-pair (bp) portion of the mtDNA control region was amplified using the primers, light-strand tPro-whale Dlp-1.5 and heavy strand Dlp-8G (Garrigue *et al.* 2004). This region extended across the two shorter and partially overlapping fragments used in past analyses, referred to as the *North Atlantic* and the *Worldwide consensus* regions by Baker &

Medrano-González (2002). Unincorporated nucleotides and primers were removed from the amplified product using shrimp alkaline phosphatase (SAP) and exonuclease I (Ex). Purified products were sequenced with BigDye vs3.1 and run on an ABI 3730xl (Applied Biosystems).

Sequences of the mtDNA control region were aligned, manually edited and haplotypes identified using the software *Sequencher* (Gene Codes). All haplotypes were sequenced multiple times either by sequencing more than one individual or, for haplotypes represented by only one individual, by sequencing in both directions. All variable sites were visually assessed to confirm haplotype identity. Sequence quality was calculated in *Sequencher* from the raw ABI base scores using a cut off of 30 (i.e., an error rate of less than 1 in 1000; (Ewing and Green 1998)). In general, a sequence length of 500 bp that scored >30 for 90% of its length was considered good quality. All sequences with a quality score <30, were inspected visually. All sequences with a quality score of <20 were repeated. On average, more than 90% of each of the n = 1865 sequences had a quality control score of >30.

Microsatellite amplification
Samples were amplified for a total of 10 previously published microsatellite loci (Table 2). Amplifications were carried out in a final volume of 10µl at the following concentrations: 1x reaction buffer, 1.5 – 4mM $MgCl_2$ (Table 1) 0.4µM each primer, 0.2mM dNTP's, 0.5U Platinum *taq* (Invitrogen) and approximately 5ng DNA. Two temperature profiles were used for amplification, 'normal' and 'long', and cycle number varied depending on whether amplifications were conducted in 96 well or 384 well format. The 'normal' temperature profile consisted of initial denaturation at 94°C for three minutes followed by 35 (96 well) or 40 (384 well) cycles of denaturation at 94°C for 30 seconds, annealing at 50°C for 30 seconds and extension at 72°C for 30 seconds with a final extension at 72°C for ten minutes. The 'long' profile was as described above with extension at 72°C for 40 seconds each cycle and a final extension at 72°C for 30 minutes. To optimize genotyping, microsatellites were fluorescently labeled (Table 1) and mixed post-PCR into sets of non-overlapping loci before being run on an ABI 3730xl sequencer (Applied Biosystems) with formamide and 500 LIZ™ size standard (Applied Biosystems). Each well contained 2µl of PCR product mix and either 10ul of formamide/ladder mix with 0.25µl/well of ladder (96-well format) or 4µl of formamide/ladder mix with 0.1µl/well of ladder (384-well format).

Microsatellite quality control and genotype identity
Alleles were sized and binned with the software program *Genemapper* v3.7 (Applied Biosystems). This program also assesses the quality of each allele based on several criteria (e.g. peak height, peak shape, bin fit) and assigns a quality score (QS) for each sample. The peaks for all alleles with a quality score less than 0.75 were visually assessed, where the peaks were deemed to be readable and clean the alleles were accepted. For some loci under some conditions a varying 'plus A' signal caused problems for the automatic binning component of *Genemapper*, in these instances bins were assigned manually.

We considered the total number of loci amplified as an additional quality control for each sample. For regions with sample sizes lager than 100, samples with fewer than 8 micro satellite loci were excluded from further analysis. For regions where total sample size was lower than 100, a small number of samples with data for at least 6 microsatellite loci and mtDNA sequence

were considered sufficient for the purposes of excluding replicates within region (OK 4 samples, OG 12 samples , CENTAM 1 sample, WAL 1 sample, RUS 1 sample). These were not used for identifying between region re-matches but were included in the QC dataset (Table 2)

The expected probability of identity (P_{ID}) for each locus was calculated in GenAlex using the within region reconciled dataset and P_{ID} for individual matches was calculated in the program CERVUS (Kalinowski *et al.* 2007).

mtDNA diversity and differentiation
The software *Arlequin* (Schneider *et al.* 2000) was used to infer a haplotype network, to calculate haplotype diversity and to estimate differentiation between all regions using an Analysis of Molecular Variance (AMOVA) calculated for differences in haplotype frequency (F_{ST}). Significance of values was tested using 10000 random permutations.

Results
Within-region genotype matching and sex ratios
Of the n=2188 total samples chosen for the initial analysis, n=2087 met the quality control threshold (n>=8 loci for large regional samples, see Methods) or were found to be sufficient for exclusion of replicate samples by microsatellite genotyping (n>=6 for small regional populations, Table 1). The 10 micro satellite loci were all found to be moderately variable with 7 to 18 alleles per loci (Table 2). For samples with a minimum of 8 loci, the probability of identity (PI) was less than 1×10^{-9}, regardless of the specific combination of loci Based on this low value, we considered that matching genotypes represented multiple samples of the same individual and unique genotypes represented unique individuals. From these 2087 samples, genotype matching with the program CERVUS resolved n=1883 regional individuals (unique within a region). Within-region matches were considered to be 'replicates' and were removed from further genetic analysis (Table 3, shown in the diagonals). The largest number of within-region replicates was found in the sample from Ogasawara (n=47 replicates) and the next largest from Mexico, Revillagigedo Islands (n= 26 replicates), reflecting perhaps sampling protocols (e.g., collection of sloughed skin in Ogasawara) or a relatively low abundance.

After removal of within-region replicates, 6 of the feeding grounds and 4 of the breeding grounds approached or exceeded the sampling objective of 100 individuals. For some regions, such as Central America and the Philippines, the small number of samples available for all years reflects relatively low density or low abundance. In other regions, such as Russia, the western Aleutians and the eastern Aleutians, it may also reflect the logistic difficulties of sampling. Although larger sample size would be preferable for the 8 regions that did not meet the sampling objective, we considered that only 2 were insufficient for further statistical analyses: the western Aleutians (n = 9) and the Philippines (n = 13). However, for others, such as the eastern Aleutians and SBC-Washington, the results of statistical tests should be considered with some caution.

A significant male bias was observed in the sex ratio of samples from the breeding grounds (Table 1). The overall male bias of 2.19:1 (n=825) was consistent with previous reports from genetic sampling of breeding grounds (Baker *et al.* 1998) or migratory corridors (Brown *et al.* 1995). The sex ratio of samples from the feeding grounds was close to unity overall (1.06:1,

n=1031) but showed a significant male bias in the Eastern Aleutian Islands and a significant female bias in the Bering Sea.

Between-region genotype matches and migratory interchange
From the total of n = 1883 regional individual identified by unique genotypes, n = 49 were found in more than one region. Between region matches were considered 'rematches' representing migratory movement of individual whales and were retained for analysis of regional population differentiation (Table 3, shown below the diagonals). These migratory connections were generally consistent with those published previously or reported in the SPLASH photo-identification analysis (Calambokidis et al. 2008). A small number of rematches (n = 16) showed within-season movement between adjacent feeding grounds (e.g., n=1 between NBC and SEA) and between adjacent breeding grounds, particularly between the Mexican Mainland and Baja California (n=9, Table 3). A moderate number of rematches (n=33) showed movement between breeding and feeding grounds. These included the movement of 8 individuals from Asian breeding grounds to Russian feeding grounds and 1 individual to the Northern Gulf of Alaska feeding grounds. Northern Gulf of Alaska and the Bering Sea also showed connections to Hawaii and all three Mexican breeding grounds. SEA showed a connection only to Hawaii. The CA-OR feeding grounds showed connections to both Mexican Mainland and Central America. Further comparison of genotypes and photo-identification records is planned to improve the description of individual migratory interchange, particularly for the Asian stock where samples sizes were relatively low and migratory interchange was relatively strong.

mtDNA diversity and haplotype relationships
Of the n=1883 samples of regional individuals, n=1856 yielded high quality sequence of the mtDNA control region. In the 500 bp consensus region chosen to overlap with control region sequences described previously in this and other populations (e.g., Baker and Medrano-Gonzalez 2002), we identified 31 variable sites resolving 28 unique haplotypes (Table 4). A parsimony network reconstructed the two primary clades described previously from a worldwide survey of mtDNA diversity (Baker *et al.* 1993): the *AE* clade, largely restricted to the North Pacific, and the *CDF* clade thought to have originated from an historical association with populations in the Southern Hemisphere (Figure 1). Most haplotypes within each clade differed from each other by only a single nucleotide substitution, while haplotypes in different clades differed by more than 10 substitutions. Four haplotypes, *A+, A-, E2* and *F2*, accounted from more than 70% of all samples on both feeding grounds and breeding grounds (Figure 1). Haplotype diversity was similar in feeding and breeding grounds overall, but showed substantial differences among regions. Diversity was notably low in southeastern Alaska (h=0.48) and high in the Mexican Mainland and Baja California (h=0.89) (Table 1). There were 2 haplotypes (represented by 7 individuals each) from Asian breeding grounds and 1 haplotype (represented by 1 individual) from Hawaii that were not found on any feeding grounds (Figure 1).

Distribution and differentiation of mtDNA
The distribution of mtDNA haplotypes shows striking differences in regional frequencies (Figure 2). Quantification of differentiation by an Analysis of Molecular Variance (AMOVA), showed highly significant differences among feeding regions (overall $F_{ST} = 0.179$, $p < 0.001$) and among breeding grounds (overall $F_{ST} = 0.106$, $p < 0.001$). The large majority of pair-wise comparisons between regions were also significant (Table 5). Among feeding regions, Russia, southeastern

Alaska and California/Oregon, were notable for particularly high levels of differentiation from each other (up to $F_{ST} = 0.478$, $p < 0.001$). Russia was dominated by the *E2* haplotype while SEA was dominated by the *A+* and *A-* haplotype. California/ Oregon showed a high proportion of *F2* haplotypes, which were almost absent from the other two regions (Figure 2). Among breeding regions, Okinawa and Central America were notable for a particularly high level of differentiation from each other ($F_{ST} = 0.454$, $p < 0.001$) and from most other breeding grounds. Okinawa was dominated by *E* haplotypes and Central America showed a high proportion of *F* haplotypes.

A comparison between feeding and breeding regions also showed a large number of significant differences (51 of 63 comparisons were significant at $p<0.05$, uncorrected for multiple comparisons), even for those regions known to be strongly connected by patterns of individual migration (e.g., by photo-identification, Calambokidis et al. 2008). For example, Okinawa showed significant differences with all feeding regions, including Russia, and southeastern Alaska showed significant differences with all breeding grounds, including Hawaii. In other cases, similarities of haplotypes frequencies, as indicated by small values of differentiation (F_{ST}) were consistent with these known connections. For example, Western Gulf of Alaska showed no significant differentiation from Mexican-AR and Mexican-BC, although it did show a small but significant difference with Mexican-ML. Central America showed no differentiation from California/Oregon, although California/Oregon differed significantly from Mexican Mainland to which it is also strongly connected by migration (Calambokidis *et al.* 2001).

Discussion and Conclusions

The first ocean-wide survey of mtDNA diversity and differentiation in humpback whales from the North Pacific has confirmed the strong influence of maternal fidelity to migratory destinations for both feeding regions and breeding grounds. The results extend greatly the previous surveys, which were limited by geographic sampling and resolution of mtDNA diversity (Baker *et al.* 1998; Baker *et al.* 1994; Medrano-Gonzalez *et al.* 1995; Witteveen *et al.* 2004). In particular, it is now clear that differences in haplotype frequencies between most breeding grounds are highly significant, particularly between the Asian, Central and American components of the population. Further substructure is also apparent within the Asian and American components, although these remain less well resolved given current analyses and sampling (see below). These genetic differences parallel, to some extent, the complex structure of this population as described previously from observation of migratory movement by naturally marked individuals (Acebes *et al.* 2007; Baker *et al.* 1986; Calambokidis *et al.* 2001; Calambokidis *et al.* 1996; Darling *et al.* 1996; Darling and Cerchio 1993; Darling and Jurasz 1983; Darling and McSweeney 1985; Urbán *et al.* 2000), and most conclusively by the final SPLASH analysis (Calambokidis et al. 2008).

Defining genetic management units

The finding of significant differences in mtDNA haplotypes among the many feeding regions and breeding grounds conforms to Moritz's (1994) criterion for recognition of 'genetic management units'. However, the potential to define numerous distinct management units on both feeding and breeding grounds differs from the tradition view of whale 'stocks' as discrete population units connected by migratory interchange and assumed to be reproductively isolated (Donovan 1991). To date, we have not yet attempted to measure the extent of reproductive

isolation between regions through tests for differentiation in microsatellite allele frequencies. Although evidence of reproductive isolation or interchange will be of interest in understanding the full complexity of humpback population structure, the observed segregation of mtDNA haplotypes is sufficient evidence for recognizing demographic units that retain unique maternal traditions of migration and habitat use (Taylor 2005). The preservation of these maternal traditions seems to be an important component in the recovery of whale population following exploitation (Clapham *et al.* 2007).

Mixed stocks and pure stocks

It is clear from the pair-wise comparisons of haplotype frequencies that there is not a simple one-to-one relationship of feeding regions to breeding grounds. Some feeding regions and some breeding grounds seems to represent relatively 'pure stocks', often with relatively reduced mtDNA diversity. Other feeding regions and breeding grounds seem to represent 'mixed-stocks', with connections to multiple migratory destinations. For example, southeastern Alaska differs significantly from all other regions except Northern BC, but mix with other Alaskan feeding regions primarily on the Hawaiian breeding grounds. Conversely, the Central American breeding grounds mix with a subset of whales from the Mexican breeding grounds, on the California/Oregon feeding grounds. This pattern suggests that the influence of maternal fidelity operates somewhat independently on feeding and breeding grounds, over an evolutionary time scale. This is perhaps most apparent in the central Alaskan feeding grounds, where individuals that share a fidelity to these feeding grounds diverge in their fidelity to breeding grounds. Further analysis of mtDNA haplotype frequencies using mixed stock analysis (Benjamin M. Bolker 2007) are planned to help quantify this on a population level. Understanding these patterns on an individual level will require integration with photo-identification records from the larger SPLASH database (Calambokidis et al. 2008) and from satellite tagging (e.g., Mate *et al.* 1998).

Missing components and stock uncertainty

Our results are consistent with the analysis of photo-identification records, in suggesting that some components of the North Pacific population are missing or underrepresented in the SPLASH sampling (Calambokidis et al. 2008). This is most apparent in the western North Pacific where we found significant differences in haplotype frequencies between Okinawa and Ogasawara and between Okinawa and Russia, despite the evidence of individual migratory interchange between these regions. This suggests that a component of the Okinawa breeding grounds is not represented in the sampling of feeding regions. The finding of 2 haplotypes (represented by 7 individuals each) only on the Asian breeding grounds, is also consistent with a missing component of the feeding range.

The significant difference between Okinawa and Ogasawara also cautions against pooling these regions for the purposes of a single abundance estimate for Asia (Calambokidis et al. 2008). Like Central America and Mexico, Asia seems to have a complex migratory structure with some degree of mixing and some degree of stratification or isolation in different island regions. Further analyses are required to develop a more detailed understanding of this regional substructure.

Acknowledgements

We thank Kelly Robertson and Carrie LeDuc of the SWFSC for their diligence in archiving of samples and extraction of DNA, and their prompt communication on all matters involving the SPLASH program. This research was supported by Contract 2006-0093-008 from the National Fisheries and Wildlife Foundation with a match from the Endowed Marine Mammal Institute, Oregon State University. Additional funding for genetic analysis of Asian breeding ground samples was provided by the US Navy through Marine Acoustics Inc.

Literature cited

Acebes, J.M.V., D.J. Darling and M. Yamaguchi. 2007. Status and distribution of humpback whales (*Megeptera novaeangliae*) in northern Luzon, Philippines. J. Cetacean Res. Manage 9:37-43.

Amos, W., H. Whitehead, M.J. Ferrari, D.A. Glockner-Ferrari, R. Payne and J. Gordon. 1992. Restrictable DNA from sloughed cetacean skin: its potential for use in population analysis. Marine Mammal Science 8:275-283.

Baker, C.S., L.M. Herman, A. Perry, W.S. Lawton, J.M. Straley, A.A. Wolman, G.D. Kaufman, H.E. Winn, J.D. Hall, J.M. Reinke and J. Ostman. 1986. Migratory movement and population structure of humpback whales (*Megaptera novaeangliae*) in the central and eastern North Pacific. Mar. Ecol. - Prog. Series 31:105-119.

Baker, C.S. and L. Medrano-Gonzalez. 2002. World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. Pages 84-99 *in* C.J. Pfeiffer, ed. Cell and molecular biology of marine mammals. Krieger Publishing Co., Inc, Melbourne, FL.

Baker, C.S., L. Medrano-Gonzalez, J. Calambokidis, A. Perry, F.B. Pichler, H. Rosenbaum, J.M. Straley, J. Urban-Ramirez, M. Yamaguchi and O.V. Ziegesar. 1998. Population structure of nuclear and mitochondrial DNA variation among humpback whales in the North Pacific. Molecular Ecology 7:695-708.

Baker, C.S., S.R. Palumbi, R.H. Lambertsen, M.T. Weinrich, J. Calambokidis and S.J. O'Brien. 1990. Influence of seasonal migration on the distribution of mitochondrial DNA haplotypes in humpback whales. Nature 344:238-240.

Baker, C.S., A. Perry, J.L. Bannister, M.T. Weinrich, R.B. Abernethy, J. Calambokidis, J. Lien, R.H. Lambertsen, J. Urban-Ramirez, O. Vasquez, P.J. Clapham, A. Alling, S.J. O'Brien and S.R. Palumbi. 1993. Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. Proceedings of the National Academy of Science, USA 90:8239-8243.

Baker, C.S., R.W. Slade, J.L. Bannister, R.B. Abernethy, M.T. Weinrich, J. Lien, J. Urban-R., P. Corkeron, J. Calambokidis, O. Vasquez and S.R. Palumbi. 1994. Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. Molecular Ecology 3:313-327.

Bolker, B.M., T. Okuyama, K.A. Bjorndal and A.B. Bolten. 2007. Incorporating multiple mixed stocks in mixed stock analysis: 'many-to-many' analyses. Molecular Ecology 16:685-695.

Berube, M., H. Jorgensen, R. McEwing and P.J. Palsboll. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, Megaptera novaeangliae. Molecular Ecology 9:2181-2183.

Brown, M.R., P.J. Corkeron, P.T. Hale, K.W. Schultz and M.M. Bryden. 1995. Evidence for a sex segregated migration in the humpback whale (*Megaptera novaeangliae*). Proceeding of the Royal Society London B 259:229-234.

Calambokidis, J., E.A. Falcone, T.J. Quinn, A.M. Burdin, P.J. Clapham, J.K.B. Ford, C.M. Gabriele, R. LeDuc, D. Mattila, L. Rojas-Bracho, J.M. Straley, B.L. Taylor, J.U. R., D. Weller, B.H. Witteveen, M. Yamaguchi, A. Bendlin, D. Camacho, K. Flynn, A. Havron, J. Huggins and N. Maloney. 2008. SPLASH: Structure of Populations, Levels of Abundance and Status of Humpback Whales in the North Pacific. U.S. Dept of Commerce, Western Administrative Center, Seattle, Washington.

Calambokidis, J., G.H. Steiger, J.M. Straley, L.M. Herman, S. Cerchio, D.R Salden, R.J. Urbán, J.K. Jacobsen, O. von Ziegesar, K.C. Balcomb, C.M. Gabriele, M.E. Dahlheim, S. Uchida, G. Ellis, Y. Miyamura, P. Ladrón de Guevara P., M. Yamaguchi, F. Sato, S.A. Mizroch, L. Schlender, K. Rasmussen and J. Barlow. 2001. Movements and population structure of humpback whales in the North Pacific Basin. Marine Mammal Science 17:769-794.

Calambokidis, J., G.H. Steiger, J.R. Evenson, K.R. Flynn, K.C. Balcomb, D. Claridge, P. Bloedel, J.M. Straley, C.S. Baker, O.v. Ziegesar, M.E. Dalheim, J.M. Waite, J.D. Darling, G. Ellis and G.A. Green. 1996. Interchange and isolation of humpback whales off California and other feeding grounds in the North Pacific. Mar. Mam. Sci. 12:215-226.

Clapham, P.J., A. Aguilar and L.T. Hatch. 2007. Determining spatial and temporal scales for management: lessons from whaling. Marine Mammal Science 24:183–201.

Darling, J.D., J. Calambokidis, K.C. Balcomb, P. Bloedel, K.R. Flynn, A. Mochizuki, K. Mori, F. Sato, H. Suganuma and M. Yamaguchi. 1996. Movement of a humpback whale (*Megaptera novaeangliae*) from Japan to British Columbia and return. Mar. Mam. Sci. 12:281-287.

Darling, J.D. and S. Cerchio. 1993. Movement of a humpback whale (*Megaptera novaeangliae*) between Japan and Hawaii. Mar. Mamm. Sci. 9:84-88.

Darling, J.D. and C.M. Jurasz. 1983. Migratory destinations of North Pacific humpback whales (*Megaptera novaeangliae*). Pages 359-368 *in* R. Payne, ed. Communication and behavior of whales. Westview Press, Boulder, Co.

Darling, J.D. and D.J. McSweeney. 1985. Observations on the migrations of North Pacific humpback whales (*Megaptera novaeangliae*). Can. J. Zool. 63:308-314.

Donovan, G.P. 1991. A review of IWC stock boundaries. Rep. int. Whal. Commn (Special Issue 13):39-68.

Ewing, B. and P. Green. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. Genome Research 8:186–194.

Excoffier, L. 1995. AMOVA: Analysis of Molecular Variance. University of Geneva, Geneva.

Garrigue, C., R. Dodemont, D. Steel and C.S. Baker. 2004. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia. Marine Ecology - Progress Series 274:251-262.

Gilson, A., M. Syvanen, K. Levine and J. Banks. 1998. Deer gender determination by polymerase chain reaction: validation study and application to tissues, bloodstains, and hair forensic samples from California. California Fish and Game 84:159-169.

Kalinowski, ST, Taper, ML & Marshall, TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**: 1099-1006.

Lambertsen, R.H. 1987. A biopsy system for large whales and its use for cytogenetics. J. Mamm. 68:443-445.

Mate, B.R., R. Gisiner and J. Mobley. 1998. Local and Migratory Movements of Hawaiian Humpback Whales Tracked by Satellite Telemetry. Canadian Journal of Zoology 76:863–868.

Medrano-Gonzalez, L., A. Aguayo-Lobo, J. Urban-Ramirez and C.S. Baker. 1995. Diversity and distribution of mitochondrial DNA lineages among humpback whales, *Megaptera novaeangliae*, in the Mexican Pacific. Can. J. Zool. 73:1735-1743.

NMFS. 1991. Recovery plan for the humpback whale (Megaptera novaeangliae). National Marine Fisheries Service, Silver Spring.

Olavarria, C., S.C. Baker, C. Garrigue, M. Poole, N. Hauser, S. Caballero, L. Florez-Gonzalez, M. Brasseur, J.L. Bannister, J. Capella, P.J. Clapham, R. Dodemont, M. Donoghue, C. Jenner, M.-N. Jenner, D. Moro, M. Oremus, D. Paton, H. Rosenbaum and K. Russell. 2007. Population structure of South Pacific humpback whales and the origin of the eastern polyneisan breeding grounds. Marine Ecology Progress Series 330:257-268.

Palsboll, P.J., M. Berube, A.H. Larsen and H. Jorgensen. 1997. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. Molecular Ecology 6:893-895.

Schneider, S., D. Roessli and L. Excoffier. 2000. Arlequin. A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Taylor, B.L. 2005. Identifying units to conserve *in* John E. Reynolds III, William F. Perrin, Randall R. Reeves, Suzanne Montgomery and T.J. Ragen, eds. *Marine Mammal Research: Conservation Beyond Crisi*s. JHU Press.

Urbán, J., A. Jaramillo, A. Aguayo, P.L.D. Guevara, M. Salinas, C. Alvarez, L. Medrano, J.K. Jacobsen, K.C. Balcomb, D.E. Claridge, J. Calambokidis, G.H. Steiger, J.M. Straley, O.V. Ziegesar, J.M. Waite, S. Mizroch, M.E. Dahlheim, J.D. Darling and C.S. Baker. 2000. Migratory destinations of humpback whales wintering in the Mexican Pacific. Journal of Cetacean Research and Management 2:101-110.

Valsecchi, E. and W. Amos. 1996. Microsatellite markers for the study of cetacean populations. Molecular Ecology 5:151-156

Waldick, R.C., M.W. Brown and B.N. White. 1999. Characterization and isolation of microsatellite loci from the endangered North Atlantic right whale. Molecular Ecology 8:1763-1765.

Witteveen, B.H., J.M. Straley, O. von Ziegesar, D. Steel and C.S. Baker. 2004. Abundance and mtDNA differentiation of humpback whales (*Megaptera novaeangliae*) in the Shumagin Islands, Alaska. Canadian Journal of Zoology 82:1352–1359.

Table 1: Summary of samples available by region for genetic analysis of humpback whales in the North Pacific (*gene*SPLASH). N_{QC} refers to the number of samples after quality control (see text); N_I is the number of individuals in each region; N_{seq} is the number of individuals with control region sequence; *h* is the haplotype diversity; M:F is the ratio of males to females and *p* is the significance of the Chi-squared test of males to females.

Region	All	N_{QC}	N_I	N_{seq}	*h*	M:F	*p*
Russia	82	80	73	72	0.837	0.73:1	0.185
Western Aleutians	9	9	9	9	0.972	1.33:1	-
Bering	137	129	119	117	0.800	0.68:1	**0.039**
Eastern Aleutians	37	37	36	36	0.848	2.2:1	**0.034**
Western Gulf of Alaska	117	109	98	98	0.845	0.69:1	0.075
Northern Gulf of Alaska	246	242	233	231	0.779	1.22:1	0.136
South East Alaska	214	197	187	185	0.480	1.12:1	0.459
Northern British Colombia	124	123	110	109	0.547	1.24:1	0.297
SBC-Washington	57	57	51	51	0.831	1.53:1	0.149
California-Oregon	136	133	124	123	0.827	1.26:1	0.205
Total Feeding regions	1159	1116	1040	1031	0.811	1.06:1	0.353
Philippines	13	13	13	13	0.628	1.17:1	-
Okinawa	96	93	78	78	0.648	2.17:1	**0.002**
Ogasawara	241	227	180	173	0.864	1.84:1	**0.000**
Hawaii	278	244	231	229	0.714	2.73:1	**0.000**
Mexico-Revillagigedo	146	142	116	115	0.857	1.52:1	**0.033**
Mexico-Baja California	137	137	122	118	0.890	3.08:1	**0.000**
Mexico-Mainland	75	74	63	63	0.893	2.21:1	**0.003**
Central America	43	41	40	37	0.755	2.25:1	**0.016**
Total Breeding regions	1029	971	843	825	0.867	2.19:1	**0.000**
Total by regions	2188	2087	1883	1856	0.838	1.43:1	**0.000**
Total number of individuals for all regions			1834				

Table 2: Summary of microsatellite loci used for individual identification of humpback whales in the North Pacific. The normal and long profiles refer to PCR extension times as described in the text.

locus	N*	Source	label	[Mg] mM	profile	repeat	range	n alleles	Probability of Identity
Ev14	2121	(Valsecchi and Amos 1996)	VIC	2.5	long	2	129-141	7	0.176
Ev37	2109	(Valsecchi and Amos 1996)	NED	3.5	normal	2	192-220	15	0.024
Ev96	2108	(Valsecchi and Amos 1996)	FAM	1.5	long	2	143-171	14	0.079
GATA417	2057	(Palsboll *et al.* 1997)	FAM	2.5	long	4	183-274	18	0.027
GATA28	2110	(Palsboll *et al.* 1997)	NED	2.5	long	4	143-191	9	0.349
GT211	2134	(Berube *et al.* 2000)	FAM	2.5	normal	2	100-118	10	0.076
GT23	2060	(Berube *et al.* 2000)	VIC	2.5	long	2	109-121	7	0.111
GT575	2044	(Berube *et al.* 2000)	FAM	1.5	normal	2	137-165	13	0.056
rw4-10	2120	(Waldick *et al.* 1999)	VIC	3	normal	2	194-208	8	0.130
rw48	2103	(Waldick *et al.* 1999)	NED	2.5	normal	2	110-122	7	0.118
mtDNA	2123	-	-	-	-	-	500bp	28	0.042

* including within region replicates and between region rematches

Table 3: Genotype matches for individual humpback whales in the North Pacific, showing within region replicates (on diagonal) and between regions 'rematches' (below diagonal).

	RUS	WAL	BER	EAL	WGOA	NGOA	SEA	NBC	SBC-WAS	CA-OR	PHI	OK	OG	HI	MXAR	MXBC	MXML	CENTAM
RUS (73)	7																	
WAL (9)		0																
BER (119)			10															
EAL (36)			1	1														
WGOA (98)					11													
NGOA (233)					1	9												
SEA (187)							10											
NBC (110)							1	13										
SBC-WAS (51)									6									
CA-OR (124)									1	9								
PHI (13)	1										0							
OK (78)	3											15						
OG (180)	4					1							47					
HI (231)			1			2	5							13				
MXAR (116)			2		1	2			1						26			
MXBC (122)			1			3									2	15		
MXML (63)			1			1				2					1	9	11	
CENTAM (40)										2								1

Table 4: Variable sites in the first 500bp of the humpback whale mtDNA control region. Relative base pair positions are in reference to a published humpback whale haplotype SP1 Genbank #DQ768307 (Olavarria *et al.* 2007).

	24	55	63	83	84	99	116	124	132	144	159	160	165	237	238	244	245	246	255	262	263	265	267	271	282	314	315	378	380	445	490
A+	G	G	C	T	T	C	G	C	G	T	T	T	C	G	T	T	C	C	T	T	C	A	T	A	T	T	C	C	A	T	A
A-	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A3	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A4	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
A5	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	C	.	.	.	.	.	.	.	.
E3	.	.	.	.	C	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E4	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	.	.	.	.	.
E5	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.
E7	.	.	.	C	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	C	.	.	.	.	.	.
E9	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E10	.	.	.	.	C	.	.	.	.	C	.	.	.	.	.	C	.	.	.	.	T	.	C	.	.	.	.	T	.	.	G
E11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	C	.
E12	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.
E13	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	G	C	.	.	A	.	.	.	C	.
E14	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	C	.	.	.	.	.	.	.	.
E15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	G	.	.
F1	.	.	.	.	C	A	.	T	A	C	.	.	.	.	.	C	.	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F2	.	.	.	.	C	A	.	T	A	C	.	C	.	.	.	C	.	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F3	.	.	.	.	C	A	.	T	A	C	.	C	.	.	C	.	.	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F4	.	.	.	.	.	A	.	T	A	C	.	C	.	.	.	C	.	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F5	.	.	.	.	C	A	.	T	A	C	.	.	.	.	.	C	T	.	.	.	T	.	.	G	.	C	T	T	.	.	G
F6	.	.	.	.	C	A	.	T	A	C	.	C	.	.	.	C	.	.	.	.	T	G	.	G	.	C	T	.	.	.	G
F7	.	.	.	.	C	A	.	T	A	C	.	.	.	.	.	C	T	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F8	.	.	T	.	C	A	.	T	A	C	.	C	.	.	.	C	.	.	.	.	T	.	.	G	.	C	T	.	.	.	G
F9	.	.	.	.	C	A	.	T	A	C	.	.	.	.	.	C	T	.	C	.	T	.	.	G	.	C	T	T	.	.	G

Table 5: Pair-wise F_{ST} values for frequencies of humpback whale mtDNA control region haplotypes between a) feeding grounds b) breeding and feeding grounds and c) breeding grounds of the North Pacific. **Bold*** show significance at 0.01, **bold** show significance at 0.05 but not at 0.01. Empty rows and columns indicate insufficient sample sizes to perform this test (n < 20).

a)

	RUS	WAL	BER	EAL	WGOA	NGOA	SEA	NBC	SBC/WA	CA/OR
RUS	--									
WAL										
BER	**0.094***		--							
EAL	**0.114***		-0.012	--						
WGOA	**0.039***		0.012	0.010	--					
NGOA	**0.105***		0.013	0.007	**0.014**	--				
SEA	**0.389***		**0.242***	**0.343***	**0.220***	**0.116***	--			
NBC	**0.293***		**0.174***	**0.245***	**0.148***	**0.080***	0.003	--		
SBC/WA	**0.038***		**0.088***	**0.104***	**0.035**	**0.076***	**0.314***	**0.223***	--	
CA/OR	**0.268***		**0.157***	**0.108***	**0.202***	**0.229***	**0.478***	**0.401***	**0.268***	--

b)

	RUS	WAL	BER	EAL	WGOA	NGOA	SEA	NBC	SBC/WA	CA/OR
PHI										
OK	**0.031***		**0.200***	**0.283***	**0.130***	**0.198***	**0.577***	**0.497***	**0.127***	**0.360***
OG	0.002		**0.101***	**0.118***	**0.042***	**0.111***	**0.326***	**0.253***	**0.029**	**0.297***
HI	**0.135***		**0.029***	0.025	**0.033***	0.000	**0.096***	**0.065***	**0.097***	**0.252***
MX-AR	**0.042***		0.010	0.008	-0.006	**0.021**	**0.234***	**0.162***	**0.048***	**0.206***
MX-BC	**0.042***		0.015	0.002	0.000	**0.032***	**0.246***	**0.176***	**0.045***	**0.152***
MX-ML	**0.088***		0.018	-0.002	**0.031**	**0.059***	**0.366***	**0.272***	**0.095***	**0.079***
CENTAM	**0.302***		**0.168***	**0.109**	**0.218***	**0.250***	**0.625***	**0.527***	**0.303***	-0.014

c)

	PHI	OK	OG	HI	MX-AR	MX-BC	MX-ML	CENTAM
PHI								
OK		--						
OG		**0.032***	--					
HI		**0.236***	**0.142***	--				
MX-AR		**0.128***	**0.046***	**0.043***	--			
MX-BC		**0.120***	**0.044***	**0.054***	0.003	--		
MX-ML		**0.202***	**0.093***	**0.084***	**0.032**	0.005	--	
CENTAM		**0.454***	**0.328***	**0.282***	**0.223***	**0.148***	**0.068***	--

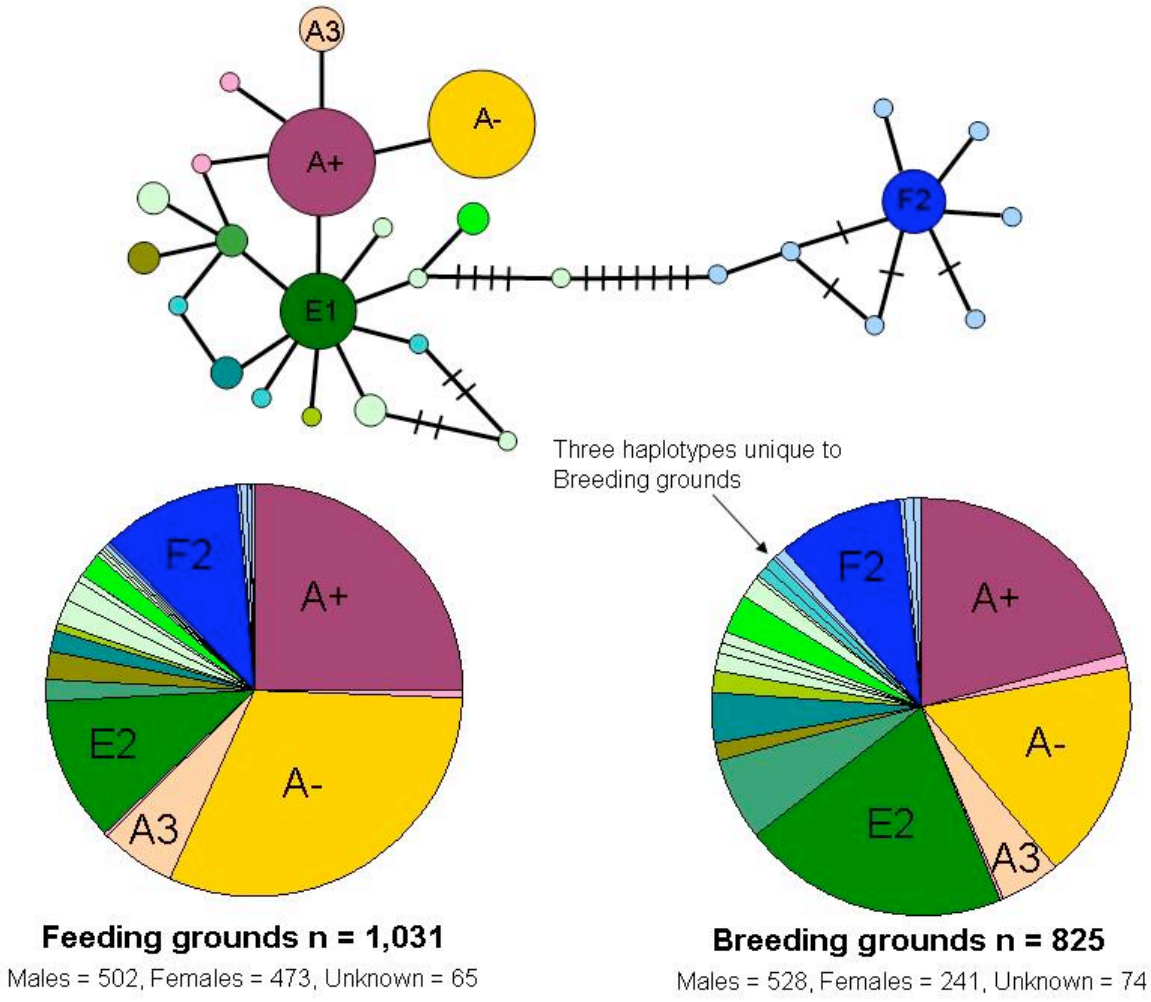

Figure 1: Parsimony network showing connections between the 28 described mtDNA haplotypes for humpback whales in the North Pacific. A-types are shaded pink and yellow, E-types are shaded green, F-types are shaded blue. The circle size is roughly representative of the proportion of the haplotype in the overall dataset. Each line indicates one base change between haplotypes. Below the network are pie charts showing proportions of haplotypes on feeding and breeding grounds.

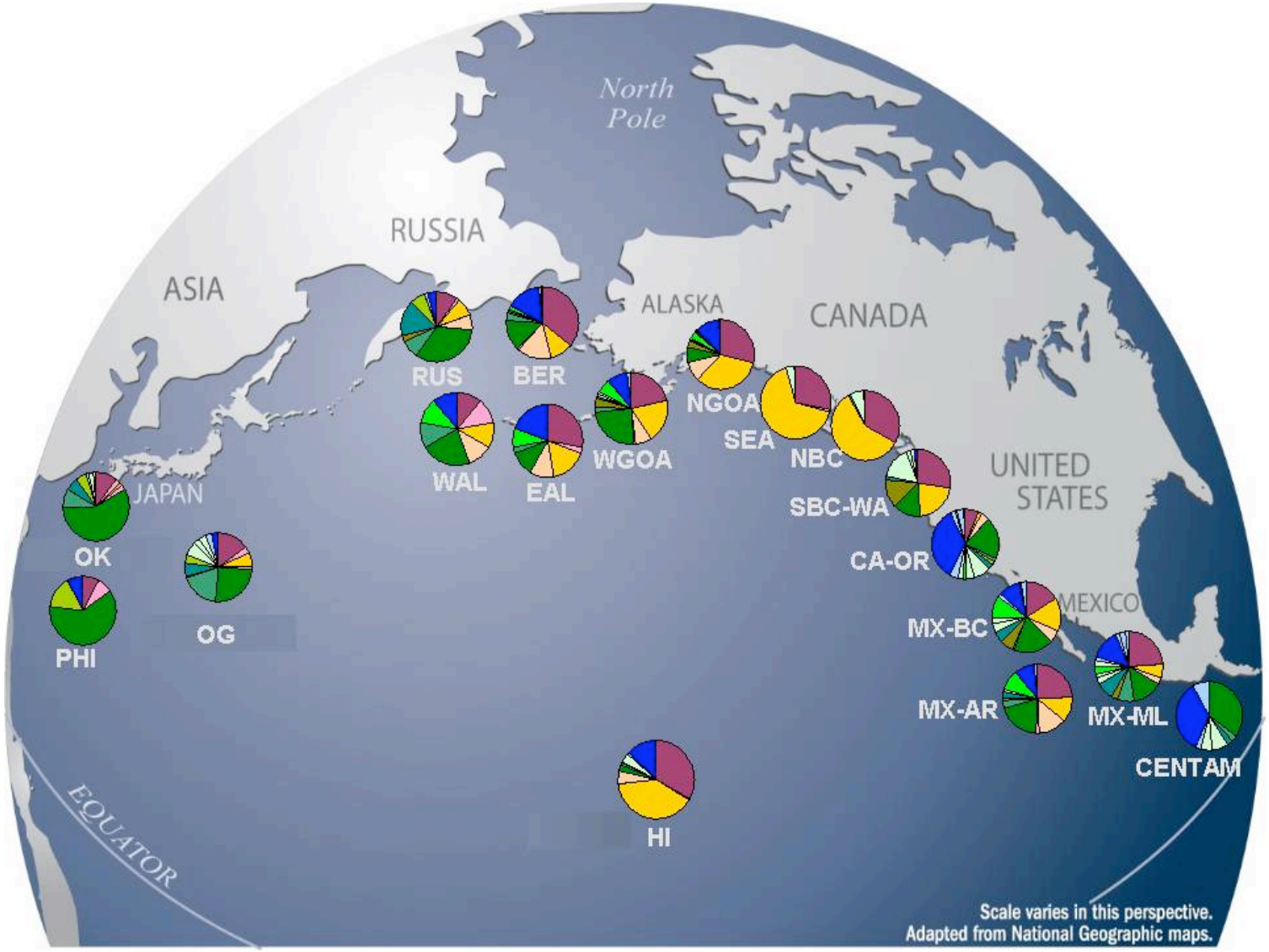

Figure 2: Regional frequencies of mtDNA haplotypes for humpback whales in the North Pacific. See Figure 1 for haplotype color codes.