Genetic distinctiveness and decline of a small population of humpback whales (*Megaptera novaeangliae*) in the Arabian Sea (Region X)

CRISTINA POMILLA[1,2], TIM COLLINS [2,3], GIANNA MINTON [3], KEN P. FINDLAY [4], MATTHEW S. LESLIE[1,2], LOUISA PONNAMPALAM [5], ROBERT BALDWIN [3] AND HOWARD C. ROSENBAUM[1,2]

[1] *Sackler Institute for Comparative Genomics, American Museum of Natural History, 79th Street and Central Park West, New York, NY 10024, USA.*
[2] *Cetacean Conservation and Research Program, c/o Marine Conservation, Wildlife Conservation Society, 2300 Southern Blvd., Bronx, NY 10460-1099, USA.*
[3] *Environment Society of Oman Whale and Dolphin Research Group, P. O. Box 3955, P. C. 112, Ruwi, Sultanate of Oman.*
[4] *Department of Oceanography, University of Cape Town, Rondebosch 7701, South Africa.*
[5] *University Marine Biological Station, University of London, Millport, Isle of Cumbrae, Scotland KA28 0EG.*

ABSTRACT

Humpback whales in the Arabian Sea present a possibly unique exception to the seasonal migrations typical of this species. Early sightings were thought to be Southern Hemisphere whales extending their migration north of the equator. Several sources of evidence have questioned this hypothesis, suggesting that at least some whales may be year-round residents. Genetic analyses based on 11 microsatellite markers and mitochondrial DNA sequences (485bp) revealed significant differentiation between whales sampled off the coast of Oman (n=67), in the Arabian Sea, whales sampled in one North Pacific and four Southern Hemisphere regions (microsatellites, smallest F_{ST}=0.0387, p<0.000; mtDNA, smallest F_{ST}=0.112, p<0.000). These results combined with estimates of gene flow and divergence times, as well as clustering analyses, suggest that the Arabian Sea population originated from the Southern Indian Ocean, but is currently isolated. Genetic diversity (microsatellites, k=7, H_o=0.7011; mtDNA, h=0.6873, π=0.01734) was significantly reduced, when contrasted with values obtained for Southern Hemisphere populations. Multiple tests showed a consistent signature of recent bottleneck, current lack of recovery, and possibly ongoing decline. Recent surveys that suggested a very small population size and our genetic findings raise great concern for the survival of this potentially small and isolated population.

INTRODUCTION

The presence and distribution of humpback whales in the Arabian Sea is known from whaling records, observations from merchant vessels, and surveys off the coast of Oman (Brown, 1957; Mikhalev, 2000; Minton *et al.*, In Press; Minton *et al.*, 2008; Reeves *et al.*, 1991; Slijper *et al.*, 1964; Wray, Martin, 1983). While early sightings were thought to be Southern Hemisphere whales pushing their migration far north into this region, this collective evidence suggests the presence of a resident population with a distribution in Yemen, Southern Oman, Iran, Pakistan and India. Recent research has confirmed the continued presence of humpback whales off the Gulf of Oman and Arabian Sea coasts of Oman with the highest concentration in the Gulf of Masirah and Kuria Muria bay regions (Baldwin, 2000; Minton, 2004; Minton *et al.*, In Press). Only limited incidental observations of the species have been recorded for some areas within the remainder of the suspected range (Al-Robaae, 1974; Braulik *et al.*, In Press).

Data on reproductive parameters collected during illegal Soviet whaling operations in 1966, as well as observations of calves and recordings of humpback whale song off the coast of Oman, indicate that this population adheres to a Northern Hemisphere breeding cycle, with peak calving taking place between January and May (Mikhalev, 1997; Mikhalev, 2000;

Minton *et al*., In Press). Connections with the Northern Hemisphere are, however, highly unlikely as the Arabian Sea offers no readily accessible northward passage. In addition whale songs recorded in Oman are very different from those recorded concurrently in neighbouring stocks suggesting that whales from these populations have no recent biological connectivity (Minton *et al*., In Press; Whitehead, 1985; S. Cerchio, Pers. Comm.). Indeed whaling data indicate that feeding occurs in monsoon-driven upwelling areas along the coasts of Oman, Pakistan and India (Mikhalev, 1997), and capture-recapture data suggest the whales off the coast of Oman may move seasonally between the Dhofar region (Hallaniyat/Kuria Muria Islands) in winter and the Gulf of Masirah to the north in summer (Minton *et al*., In Press).

Due to its quirky geographic position and uncertain population origin, the Arabian Sea has been grouped by the IWC with the Southern Hemisphere distribution as Breeding Stock X (International Whaling Commission, 2004). Whether the Arabian Sea is connected to Southern Hemisphere stocks, and if so to what extent, has not yet been established with certainty, although more and more evidence seems to point to a discrete population. Humpbacks from this region carry fewer, and smaller, barnacles than other Southern Hemisphere whales, and do not exhibit the white oval scars indicative of cookie cutter shark bites, a feature very commonly seen on other Southern Hemisphere humpbacks (Mikhalev, 1997). Comparisons of the photo-identification catalogue from Oman to those from Madagascar, South Africa, Mozambique, and Zanzibar produced no matches (Minton *et al*., 2009; Minton *et al*., In Press). Recent mtDNA analyses, using a subset (n=46) of the Stock X samples used in this study, found strong genetic structure between Stock X and other Breeding Stocks in the South Atlantic Ocean (Stocks A and B) and in the southwestern Indian Ocean (Stock C) (pair-wise FST range between Oman and other Indian Ocean breeding populations 0.11-0.15, Rosenbaum *et al*., 2009). This study however did not test connections with the eastern Southern Hemisphere range, or a potential Northern Hemisphere origin of the Stock.

Mark-recapture studies using three different pairings of tail fluke photographs collected in Oman in two main research areas over a period of four and a half years yielded a population estimate of 82 individuals (95% CI=60-111) (Minton *et al*., In Press). However, sample sizes were small and certain portions of the suspected range were not surveyed, possibly introducing a negative bias (Minton *et al*., In Press). Nevertheless, the survival of this potentially small and isolated population is of great concern to the international scientific and conservation community, as confirmed by its recent listing in the IUCN Red List of Threatened Species in the Category 'Endangered' (Minton *et al*., 2008).

We combined mtDNA and nuclear microsatellites data for an expanded number of samples from the Arabian Sea to address two objectives: i) to test comprehensively hypotheses on the origin and connections, if any, of Stock X; ii) to assess the status of the population from a conservation genetics point of view. To address the first objective we also used mtDNA data, and where available microsatellite data, from four Southern Hemisphere Breeding Stocks (Stocks C to F) and one additional Northern Hemisphere area (North Pacific), as those are the most proximal regions. Our approach used a combination of traditional statistical analyses of population structuring, Bayesian clustering, coalescent-based maximum likelihood estimates of migration rates and divergence time. Genetic diversity estimates, and bottleneck and population stability tests were used to address the second aim.

MATERIAL AND METHODS

Sample collection and DNA extraction

Samples were collected from 1999 to 2004 during small-boat surveys that were conducted in two main locations: the Gulf of Masirah and Dofhar, on the Arabian Sea coast of Oman (Figure 1). Surveys occurred between October and March and were designed to target areas where published (Mikhalev, 2000) and unpublished records (held by the authors) indicated potentially higher abundance of humpback whales.

Skin tissues were mostly obtained using the biopsy dart procedure (Lambertsen, 1987) or as sloughed skin when available. Few samples were collected from carcasses of stranded animals. Samples were preserved in the field in salt saturated 20% Dimethyl Sulfoxide solution (DMSO) and later stored at –20ºC until processed. Total genomic DNA was extracted from the epidermal layer of biopsies using the DNAeasy tissue kit (Qiagen).

Mitochondrial DNA and microsatellite molecular analyses

A 520 bp fragment within the mtDNA control region (Baker *et al*., 1993; Kocher *et al*., 1989) was amplified with primers Dlp-1.5 and Dlp-5 (Baker *et al*., 1998). Polymerase Chain Reaction (PCR) products were cycle-sequenced (both forward and reverse) with dye-labeled terminators following conditions recommended by the manufacturers. Sequence reactions were analyzed using an ABI-Prism 3700 or 3730 DNA Analyzer (Applied Biosystem).

Eleven cetacean di-, tri- and tetra- nucleotide microsatellite loci were selected from literature: 199/200, 417/418 and 464/465 (Schlötterer *et al*., 1991), EV1Pm, EV37Mn, EV94Mn and EV96Mn (Valsecchi, Amos, 1996), and GATA028, TAA031, GATA053 and GATA417 (Palsbøll *et al*., 1997). One primer of each pair was labelled with a fluorescent tag (HEX, 6-FAM and TET, Qiagen-Operon; NED, Applied Biosystems) on the 5′ end. Polymerase chain reactions (PCRs) were carried out in a 20μl or 10μl volume with the following conditions: 50mM KCl, 10mM Tris-HCl pH8.8, 2.5-3.5mM $MgCl_2$, 200μM of each dNTP, 0.4μM of each primer, 0.025 U/μl *Taq* Gold polymerase (Perkin-Elmer). Amplifications were completed in either a Perkin-Elmer 9600 thermal cycler or an Eppendorf Gradient Mastercycler, after optimization of published annealing temperatures and profiles. Pooled PCR products were loaded with the addition of an internal standard ladder (Genscan-500 TAMRA or ROX, Applied Biosystems) on 3700 or 3730 DNA analyzer (Applied Biosystems). The allele size in base pairs was identified with the software GENESCAN ANALYSES and GENOTYPER 2.1 or GENEMAPPER (Applied Biosystems).

Statistical analysis

Duplicate samples were detected using genotype identity using the Excel add-in MS_TOOLKIT package (Park, 2001), and were consequently eliminated. The average probability of different random individuals sharing the same genotype by chance (Probability of Identity, PI) was estimated using the software API-CALC 1.0 (Ayres, Overall, 2004).

From the 520 bp mtDNA fragment, a 486 bp consensus region that contains the majority of variable nucleotide positions in the mtDNA control region of humpback whales was examined for all samples (Baker *et al*., 1993). Sequences were aligned and edited using SEQUENCHER v. 4.5 (Gene Codes Corp. Ann Arbor, MI). Comparisons of sequences to identify polymorphic sites and haplotypes were performed using MACCLADE v. 4.01 (Maddison, Maddison, 1992).

Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) at microsatellite loci were evaluated with a probability test (Guo and Thompson, 1992) implemented in GENEPOP v.3.4 (Raymond and Rousset, 1995).

i. Origin of the population

To test possible hypotheses on the origin of the Arabian Sea population we included in the analyses mtDNA control region sequences and 11 microsatellite markers for 740 individuals sampled at five locations within the Southwestern Indian Ocean (Breeding Stock C) (Pomilla *et al*., 2006; Rosenbaum *et al*., 2009). Additionally we obtained from GenBank sequences for 174 samples (464 bp) from the Southeastern Indian Ocean (Breeding Stock D) (Olavarria *et al*., 2007), 835 samples (464 bp) from the South Pacific Ocean (Breeding Stock E n=605, Breeding Stock F n=230) (Olavarria *et al*., 2007), and 54 samples (425 bp) from the North Pacific Ocean (Witteveen *et al*., 2004).

An unrooted phylogeny of the humpback whale mtDNA haplotypes identified from all of the above sequences was constructed using the Neighbor-Joining method (Saitou, Nei, 1987), and parsimony as implemented in PAUP* (Swofford, 2000). The sequences were adjusted for multiple substitutions using the Kimura 2-parameter model (Kimura, 1980).

The differentiation between Region X and the other areas was quantified using pairwise *F*-statistics as implemented in ARLEQUIN V..11 (Excoffier *et al*., 2005). These were calculated for mtDNA nucleotide differentiation (Φ_{ST}) and haplotype frequency differences (F_{ST}), as well as for microsatellite allele frequency differences (F_{ST}). The significance of the observed Φ_{ST} and F_{ST} values was tested using 10,000 random permutations of the data matrix. No correction for simultaneous tests was applied to significance levels of pairwise comparisons (Narum, 2006; Perneger, 1998).

Population subdivisions were also tested using microsatellites data and two Bayesian model-based clustering methods implemented in BAPS v.4.14 (Corander and Marttinen, 2006) and STRUCTURE v.2.1 (Pritchard *et al*., 2000). Both methods assume the presence of *K* populations, where *K* may be unknown. BAPS v.4.14 uses stochastic optimization to infer the posterior mode of the genetic structure, while in STRUCTURE a Markov chain Monte Carlo procedure is used for estimation. Genetic mixture analysis in BAPS was performed at the 'group level' clustering *a-priori* defined population samples. For STRUCTURE we analyzed our data without supplying prior information on the origin of the samples, using the 'correlated frequencies' model, and excluding admixture ($1.5x10^6$ dememorization iterations, 10^6 chain iterations). The true number of populations was estimated computing ΔK over 10 runs for each *K* following the method described by Evanno *et al*. (2005).

In order to estimate effective migration rates and divergence times between Region X and the other areas we used mtDNA data in the program MDIV (Nielsen, Wakeley, 2001), which uses a Markov Chain Monte Carlo (MCMC) method within a likelihood framework to estimate the posterior distributions of: the migration rate per gene per generation [$M=(2 N_e m)$], the time in generations since the two populations diverged scaled by the effective population size [$T=(t/2 N_e)$], and the parameter theta (θ), which is a product of the effective population size and the mutation rate per generation of the studied gene region [$\theta=(4 N_e \mu)$]. Default maximum values for Mmax and Tmax were used. A minimum of five runs was carried out for each comparison and results were averaged between runs (10^6 dememorization iterations, $5x10^6$ chain iterations). Terminal ends of sequences had to be truncated as they contained gaps, which MDIV is not able handle.

ii. Status of the population

The diversity of humpback whale mtDNA sequences was estimated both as haplotype diversity (*Hd*) and as mean number of pairwise nucleotide differences (*k*) (Nei, 1987; Tajima, 1993) using ARLEQUIN V.3.11 (Rozas *et al*., 2003). For nuclear microsatellites the mean number of alleles per locus (*K*), the observed heterozygosity (H_o), and the heterozygosity expected (H_e) under Hardy-Weinberg assumptions (Nei, 1987) were calculated in FSTAT v.2.9.3 (Goudet, 1995).

To test for reduction in population size we calculated from microsatellite data *M*, the ratio between number of alleles and range of the allelic array with the software M-RATIO (Garza, Williamson, 2001). *M* declines after a population is reduced in size and the magnitude of the decrease is positively correlated with the severity and duration of the reduction in size. A simulation approach was used to calculate a critical value for the ratio *M (M*c) in an equilibrium population, below which one can assume that a given data set is a sample from a population that has experienced a recent reduction in size. We ran the simulation under a conservative two-phase mutation model where the proportion of one-step mutations is p_s=90% and the average size of non one-step mutations is Δ_g=3.5, as well as a realistic model obtained from literature data (p_s=88%, Δ_g=2.8) (Garza, Williamson, 2001). The simulations were repeated for different values (from 0.004 to 5) of historical diversity θ ($4 N_e \mu$), which represents different equilibrium population sizes if we assume a constant μ (e.g. 5×10^{-4}, Garza, Williamson, 2001). The rate of recovery of *M* following a reduction in size is positively correlated with post-reduction population size, but that recovery occurs in both small and large populations. This indicates that *M* can distinguish between populations that have been recently reduced in size and those that have been small for a long time.

To further investigate population stability we applied three tests of neutral theory of molecular evolution to the mtDNA data. The raggedness statistics *rg* analyzes the distribution

of pairwise differences, or mismatch distribution (Harpending *et al*., 1993; Rogers, Harpending, 1992); Tajima's *D* (Tajima, 1989) and R_2 statistics (Ramos-Onsins, Rozas, 2002) are both based on the distribution of mutation frequencies. The significance of the tests was assessed with coalescent simulations. All tests and simulations were conducted with DNASP v.4.10.9 (Rozas *et al*., 2003).

RESULTS

Sample size and mtDNA and microsatellite variation

A total of 67 samples from the Arabian Sea were included in the study. Excluding four samples that did not yielded amplifiable DNA, for the remaining 63 the average PI was small enough to exclude re-samples with confidence and 47 individuals were included in the subsequent analyses.

A consensus region of 486 bp of the mtDNA control region was assembled in which a total of 7 maternal haplotypes was detected from 24 polymorphic sites. The other 1,798 samples included in the analysis yielded 167 haplotypes.

For microsatellites missing allelic data averaged 0.3% across all loci, the largest number of alleles (12) was found at EV37Mn, the smallest (3) was recorded at locus EV1Pm. Deviation from HWE was rejected for all single loci and across loci (p=0.79) and no significant heterozygote deficiency was found. In the same way there was no evidence of LD.

Statistical analysis

i. Origin of the population

Of the seven identified mtDNA haplotypes four were shared with Stock C, two were shared with Stock D or Stock E, one was shared with Stock F, and three were private. In the Neighbor-joining tree (Figure 2) constructed from the total 170 haplotypes, the private Stock X haplotypes were placed in a clade together with Southern Hemisphere haplotypes, the closest ones of which were haplotypes from Stocks C and D. The North Pacific haplotypes formed a separate clade with the exception of one haplotype, which was shared with Southern Hemisphere samples (SP88).

Pairwise *F*-statistics showed high levels of differentiation for Stock X and were among the highest recorded for population differentiation among any humpback whale populations worldwide. In particular for mtDNA (Figure 3) the highest divergence was found from the North Pacific, while comparisons with Southern Hemisphere Stocks showed a gradient of differentiation related to geographic distance and therefore the lowest differentiation was found with Stock C and the highest with Stock F. For microsatellites, comparisons between Stock X and five locations within Stock C showed strong structure, with the highest divergence from Sub-Stocks C2/C3 (F_{ST}= 0.046-0.048) and the lowest from Sub-Stock C1 (F_{ST}= 0.040). Incidentally, no matches were recovered between the two stocks.

For STRUCTURE computations, the highest posterior probability of the data was obtained when the individuals were partitioned in only two clusters [Pr (K=2/X)=1]. All individuals from Stock X were assigned to the same cluster for all values of K, and for K=2 with 0.85-0.99 probabilities, except two samples that had ca. 0.4 probabilities to be assigned to the other cluster. All the remaining individuals had equal probabilities to belong to any of the clusters for any value of K. Similarly BAPS resolved the presence of two clusters of individuals, and all the samples from Stock X were assigned to one of the two.

MDIV coalescent analyses showed that Stock X diverged first from the North Pacific (T>2), and at a later time from the Southern Hemisphere stocks. Figure 4 shows that the closest divergence time is between Stocks X and C (T=0.1684). This divergence time is however very similar to the divergence time between Stocks X and D, therefore it cannot be excluded that X diverged from C and D at about the same time. Using this T estimate, an average estimate for θ=17.291, a mutation rate of 5.2% per million years (Alter *et al*., 2007), a generation time of 21.5 years (Taylor *et al*., 2007), and the length of the sequence analyzed (477 bp) we obtained a divergence time of ~70,111 years ago.

The posterior distributions of migration rates showed that since the time of divergence limited gene flow has been exchanged with the Southern Hemisphere and as expected the highest exchange has been with Stock C (M=5.67, Figure 5). To render an idea of the scale it may be useful to note that M values between distinct contiguous stocks in the Southern Hemisphere are 5-6 times larger (Rosenbaum *et al*., 2009).

ii. Status of the population

Reduced genetic diversity for Stock X was confirmed by both mtDNA and microsatellite analyses (Figure 6). The mtDNA haplotype diversity (Hd) was lower than in any of the other areas analyzed. The average number of pairwise nucleotide differences (k) was larger only when compared to the North Pacific. Reduced diversity for Northern Hemisphere whales compared to Southern Hemisphere whales has been previously described (Baker, Medrano-Gonzalez, 2002). For microsatellites both observed heterozygosity (Ho) and mean number of alleles per locus (K) were smaller than in five sampling locations within Stock C.

The observed ratio between number of alleles and the range of the allelic array was M=0.7706 (Figure 7). Coalescent simulations of the data at equilibrium were carried out under a realistic mutation model for different historical diversity values (θ=0.004 to θ=5). The M-ratio produced by coalescent simulations (Mc) was higher than that produced by the current data only for θ <4.2, which corresponds to a pre-bottleneck effective population size of ~2,100 animals (μ=5 x 10^{-4}) (Figure 7). When the conservative mutation model was applied, this yielded an estimated cut-off value of θ ~1.1, which corresponds to a pre-bottleneck effective population size of ~550 animals. In other words, if the pre-bottleneck N_e was larger than those cut-off values, than there is no evidence of bottleneck. Considering that effective population sizes correspond to ~ 10% of census sizes (Frankham, 1995), these numbers are reasonably large to assume that the population has experienced a bottleneck. M starts to decline immediately, within 1-2 generation, after the bottleneck. Additionally, if after a bottleneck event a population stays small, it takes ~300 generations for M to increase to normal levels (Garza, Williamson, 2001). This means that Stock X may have experienced a bottleneck as recently as 20 ya, or as early as ~6,450 ya (generation time = 21.5 year).

In support of the above result, significant Tajima's D [D=1.8574, $P(D\geq1.8574)$=0.023], Ramos-Onsins and Rozas' R_2 (R_2=0.1763, $P(R_2\geq0.1763)$=0.017) indicated that the population is not currently at equilibrium. While these statistics are generally used to test for population expansion (significant expansion when the observed statistics is below the minimum of the 95% CI of the distribution), when the observed value falls in the upper tail of the distribution, as it is the case for both test, it may indicates population decline (J. Rozas, Pers. Comm.). Raggedness statistics (rg=0.2566 p=0.009) significantly rejected a population expansion.

DISCUSSION

Our results show that the Arabian Sea population originated most likely as a consequence of an expansion followed by a contraction of the range of Southern Indian Ocean whales, or as the result of an immigration event. Although historical gene flow seems to have occurred after the divergence which we dated back to ~58,600 ya, it is very unlikely that migrants are currently being exchanged between the Arabian Sea and the Southern Indian Ocean stocks. Shared haplotypes may in fact be simply the result of shared ancestry. Foetal evidence from whaling data indicates that northern Indian Ocean humpback whales adhere to a Northern Hemisphere humpback breeding cycle with mating and calving occurring between January and May (Mikhalev, 1997), providing a likely barrier to inter-breeding with whales in the Southern Hemiphere (Minton *et al*., 2008). Further evidence of feeding and breeding behaviours (recording of song), as well as the year-round sightings of individual whales, suggested that whales in this area may not exhibit typical migratory characteristics associated with other humpback whale populations, and represent therefore a unique and isolated population.

Population size estimates have revealed a small population size, but so far it has not been clarified whether Stock X was a small equilibrium population originated from a small founder

group, or a population that had recently declined. Our evidence of reduced genetic diversity at both mtDNA and nuclear microsatellites, as well as of a recent bottleneck, current lack of recovery, and possibly ongoing decline all favor the latter hypothesis. We estimated a time of bottleneck (20-5,400 ya) compatible with whaling (40 ya) but we cannot exclude an earlier date, or multiple bottlenecks. Whaling records indicate that 238 humpback whales were taken in the Arabian Sea in 1966-1967, and the Soviets estimated that 60% of entire population was taken (Mikhalev, 1997). In favor of a recent bottleneck hypothesis, tests of population expansion suggest that the population has not yet started recovering and may still be in decline.

This population may be unusually vulnerable to the recent threats of coastal development and pollution described above and elsewhere (Baldwin *et al*., 2010; Minton *et al*., In Press). This is of particular concern in light of the numerous lines of evidence suggesting that this population is both spatially, genetically, and demographically isolated from other humpback whale breeding populations in both hemispheres.

REFERENCES

Al-Robaae K (1974) *Tursiops aduncus* bottlenosed dolphin: a new record for Arab Gulf, with notes on Cetacea of the region. *Bulletin of the Basrah Natural History Museum* **1**, 7-15.

Alter SE, Rynes E, Palumbi SR (2007) DNA evidence for historic population size and past ecosystem impacts of gray whales. *Proceedings of the National Academy of Sciences* **104**, 15162-15167.

Ayres KL, Overall DJ (2004) api-calc 1.0: a computer program for calculating the average probability of identity allowing for substructure, inbreeding and the presence of close relatives. *Molecular Ecology Notes* **4**, 315-318.

Baker CS, Flórez-González L, Abernethy B*, et al.* (1998) Mitochondrial DNA variation and maternal gene flow among humpback whales of the Southern Hemisphere. *Marine Mammal Science* **14**, 721-737.

Baker CS, Medrano-Gonzalez L (2002) World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. In: *Cell and Molecular Biology of Marine Mammals* (ed. Pfeiffer CJ), pp. 84-99. Krieger Publishing Co.

Baker CS, Perry A, Bannister JL*, et al.* (1993) Abundant mitochondrial DNA variation and worldwide population structure in humpback whales. *Proceedings of the National Academy of Sciences USA* **90**, 8239-8243.

Baldwin R (2000) Oman's humpback whales (*Megaptera novaeangliae*). *Journal of Oman Studies* **11**, 11-18.

Baldwin R, Collins T, Minton G*, et al.* (2010) Arabian Sea humpback whales: canaries for the Northern Indian Ocean. Paper SC/62/SH20 presented to the IWC the Scientific Committee, (unpublished).

Braulik GT, Ranjbar S, Owfi F*, et al.* (In Press) Marine Mammal Records from Iran. *Journal of Cetacean Research and Management*.

Brown SG (1957) Whales observed in the Indian Ocean: notes on their distribution. *Marine Observer* **27**, 157-165.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research* **66**, 95-107.

Garza JC, Williamson EG (2001) Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**, 305-318.

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate *F*-statistics. *Journal of Heredity* **86**, 485-486.

Harpending HC, Sherry ST, Rogers AR, Stoneking M (1993) Genetic structure of ancient human populations. *Current Anthropology* **34**, 483-496.

International Whaling Commission (2004) Report of the Scientific Committee. *Report of the International Whaling Commission* **56**.

Kimura M (1980) A simple method for estimating rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.

Kocher TD, Thomas WK, Meyer A*, et al.* (1989) Dynamics of Mitochondrial DNA Evolution in Animals: Amplification and Sequencing with Conserved Primers. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 6196-6200.

Lambertsen RH (1987) A biopsy system for large whales and its use for cytogenetics. *Journal of Mammalogy* **68**, 443-445.

Maddison WP, Maddison DR (1992) MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts.

Mikhalev YA (1997) Humpback whales, *Megaptera novaeangliae,* in the Arabian Sea. *Marine Ecology Progress Series* **149**, 13-21.

Mikhalev YA (2000) Whaling in the Arabian Sea by the whaling fleets *Slava* and *Sovetskaya Ukraina*. In: *Soviet Whaling Data (1949-1979)* (eds. Yablokov AV, Zemsky VA, Tormosov DD), p. 408. Centre for Russian Environmental Policy, Moscow.

Minton G (2004) *Ecology and Conservation of Cetaceans in Oman with particular reference to humpback whales, Megaptera novaeangliae (Borowski 1781)* D.Phil., University of London.

Minton G, Cerchio S, Collins T*, et al.* (2009) A note on the comparison of humpback whale tail fluke catalogues from the Sultanate of Oman with Madagascar and the East African mainland. *Journal of Cetacean Research and Management* **11**, 67-70.

Minton G, Collins T, Findlay KP*, et al.* (In Press) Seasonal distribution, abundance, habitat use and population identity of humpback whales in Oman. *Journal of Cetacean Research and Management, Special Issue on Southern Hemisphere Humpback Whales*.

Minton G, Collins T, Pomilla C*, et al.* (2008) *Megaptera novaeangliae* (Arabian Sea sub-population). In: *2009 IUCN Red List of threatened species* (ed. IUCN).

Narum SR (2006) Beyond Bonferroni: Less conservative analyses for conservation genetics. *Conservation Genetics* **7**, 783-787.

Nei M (1987) *Molecular Evolutionary Genetics* Columbia University Press, New York.

Nielsen R, Wakeley J (2001) Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach. *Genetics* **158**, 885-896.

Olavarria C, Baker CS, Garrigue C*, et al.* (2007) Population structure of South Pacific humpback whales and the origin of the eastern Polynesian breeding grounds. *Marine Ecology Progress Series* **330**, 257-268.

Palsbøll PJ, Bérubé M, A.H. L, Jørgensen H (1997) Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Molecular Ecology* **6**, 893-895.

Park S (2001) MS_Toolkit: Excel add-in tool package for microsatellite data. Genetics Department, Trinity College, Dublin.

Perneger TV (1998) What's wrong with Bonferroni adjustments. *BMJ* **316**, 1236-1238.

Pomilla C, Best PB, Findlay KP*, et al.* (2006) Population structure and sex-biased gene flow in humpback whales from Wintering Regions A, B, C and X based on nucleare microsatellite variation, p. 22. Paper SC/A06/HW38 presented to the to the IWC, Intersessional Workshop for the Comprehensive Assessment of Southern Hemisphere humpback whales, Hobart, Tasmania

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092-2100.

Reeves RR, Leatherwood S, Papastavrou V (1991) Possible stock affinities of humpack whales in the Northern Indian Ocean. *UNEP Marine Mammal Technical Reports* **3**, 259-269.

Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.

Rosenbaum HC, Pomilla C, Mendez M*, et al.* (2009) Population structure of humpback whales from their breeding grounds in the South Atlantic and Indian Oceans. *PLoS ONE* **4**.

Rozas J, Sánchez-DelBarrio JC, Messegyer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 2496-2497.

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Schlötterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**, 63-65.

Slijper EJ, van Utrecht WL, Naaktgeboren C (1964) Remarks on the distribution and migration of whales, based on observations from Dutch ships. *Bijdragen Tot De Dierkunde* **34**, 4-86.

Swofford DL (2000) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Tajima F (1989) The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597-601.

Tajima F (1993) Measurement of DNA polymorphism. In: *Mechanisms of molecular evolution. Introduction to molecular paleopopulation biology* (eds. Takahata N, Clark AG), pp. 37-59. Japan Scientific Societies Press, Sinauer Associates, Inc, Tokyo.

Taylor BL, Chivers SJ, Larese J, Perrin WF (2007) Generation lenght and percent mature estimates for IUCN assesments of cetaceans. Administrative report LJ-07-01. National Marine Fisheries Service, Southwest Fisheries Science Center.

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

Whitehead H (1985) Humpback whale songs from the Northern Indian Ocean. *Investigations on Cetacea* **17**, 157-162.

Witteveen BH, Straley JM, Von Ziegesar O, Steel D, Baker CS (2004) Abundance and mtDNA differentiation of humpback whales (*Megaptera novaeangliae*) in the Shumagin Islands, Alaska. *Canadian Journal of Zoology* **82**, 1352-1359.

Wray P, Martin KR (1983) Historical whaling records from the Western Indian Ocean. *Reports of the International Whaling Commission (Special Issue)* **5**, 213-241.

Figure 1 Map of the coast of Oman showing two sampling locations in the Arabian Sea (Stock X).

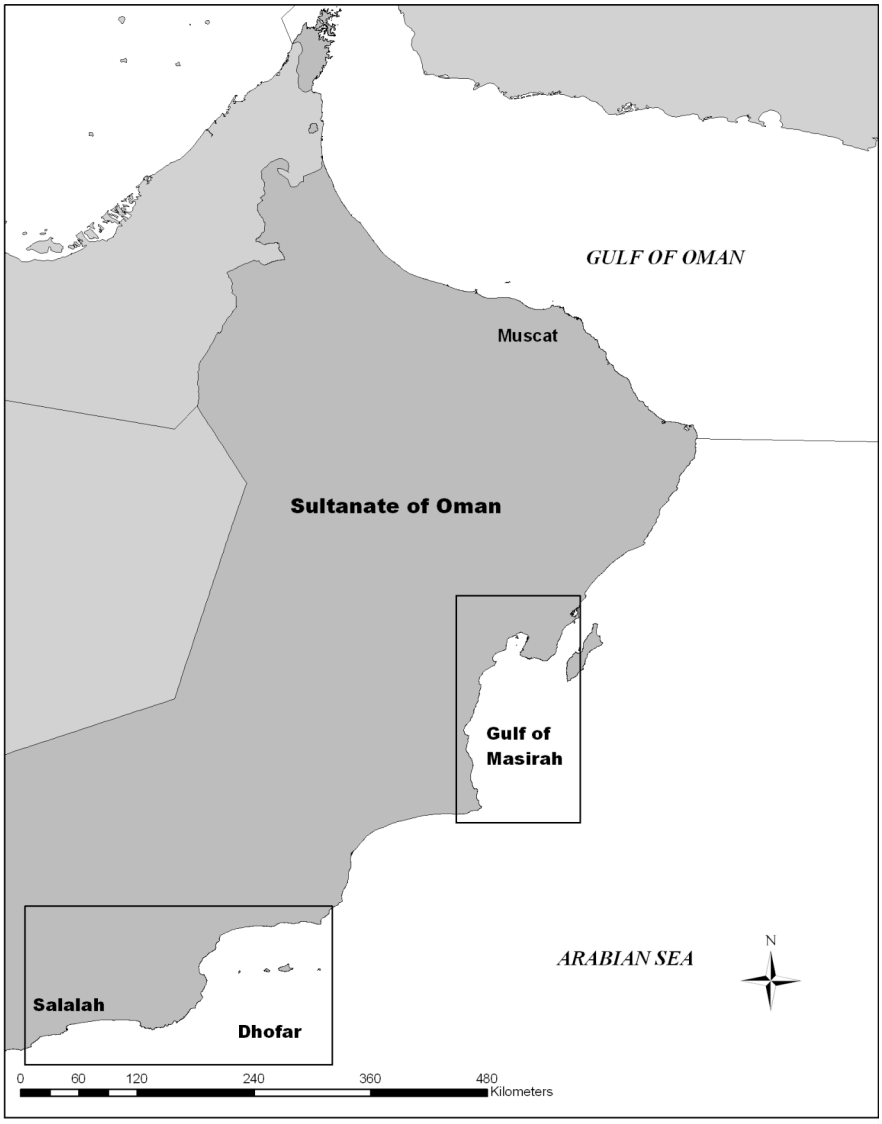


Figure 2 Unrooted Neighbor-joining tree constructed from 170 haplotypes from Stock X, C, D, E, F, and North Pacific. Stock X haplotypes are indicated by arrows.

Figure 3 Pairwise measures of mtDNA differentiation between Stock X (Arabian Sea) and Stocks C to F in the Southern Hemisphere, and two sampling location in the North Pacific (grouped as NP). Rhomboids refer to F_{ST}, while circles to Φ_{ST}.

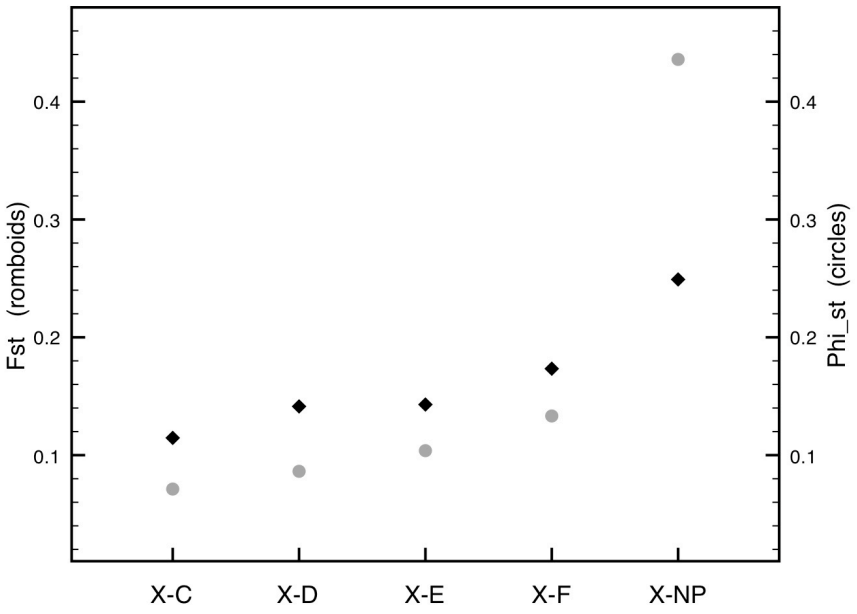


Figure 4 MDIV estimates of divergence time.

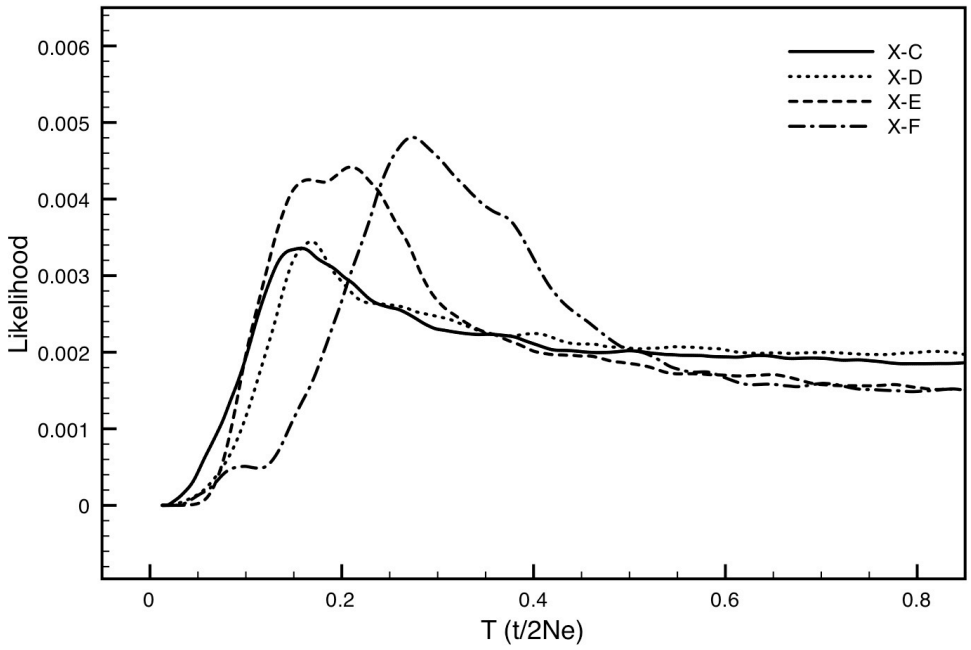


Figure 5 MDIV estimates of migration rates.

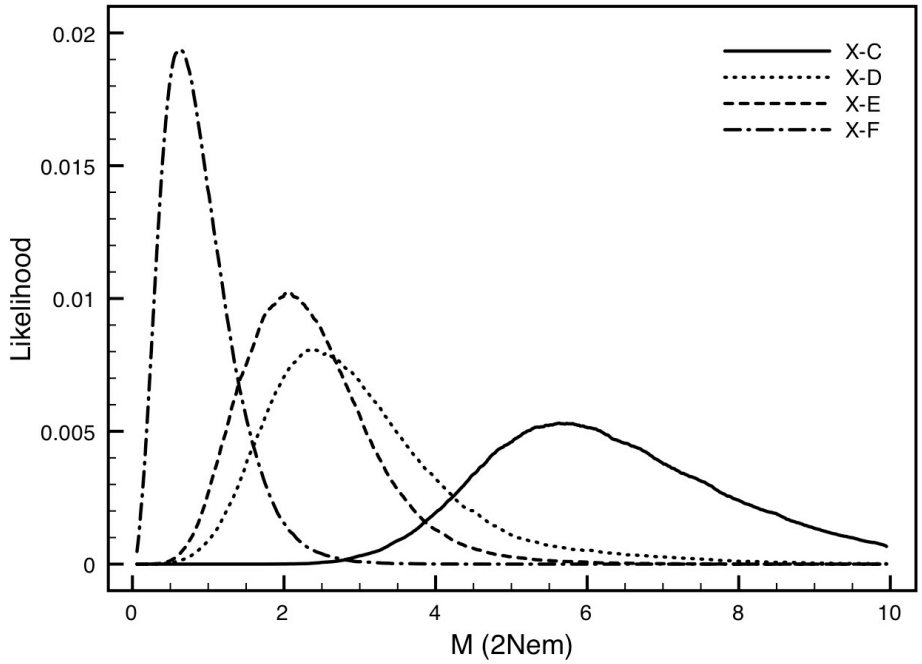
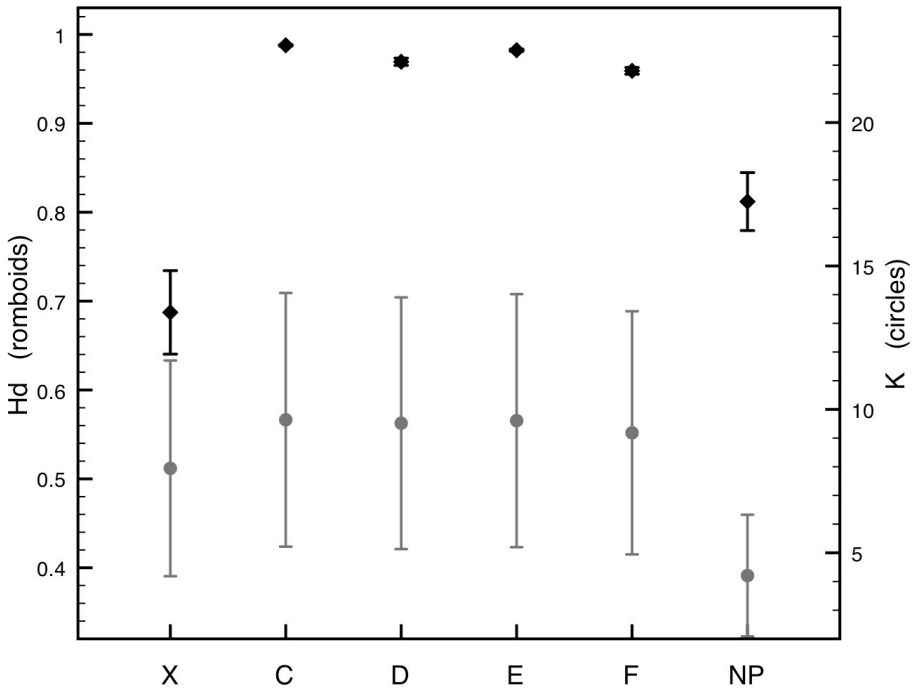


Figure 6 a) mtDNA diversity: rhomboids refer to haplotype diversity *H*d and circles to the mean number of pairwise differences k; **b)** microsatellite diversity: rhomboids refer to the observed heterozygosity *H*o, and circles to the mean number of alleles per locus K. Bars represent the standard deviation.

a)



b)

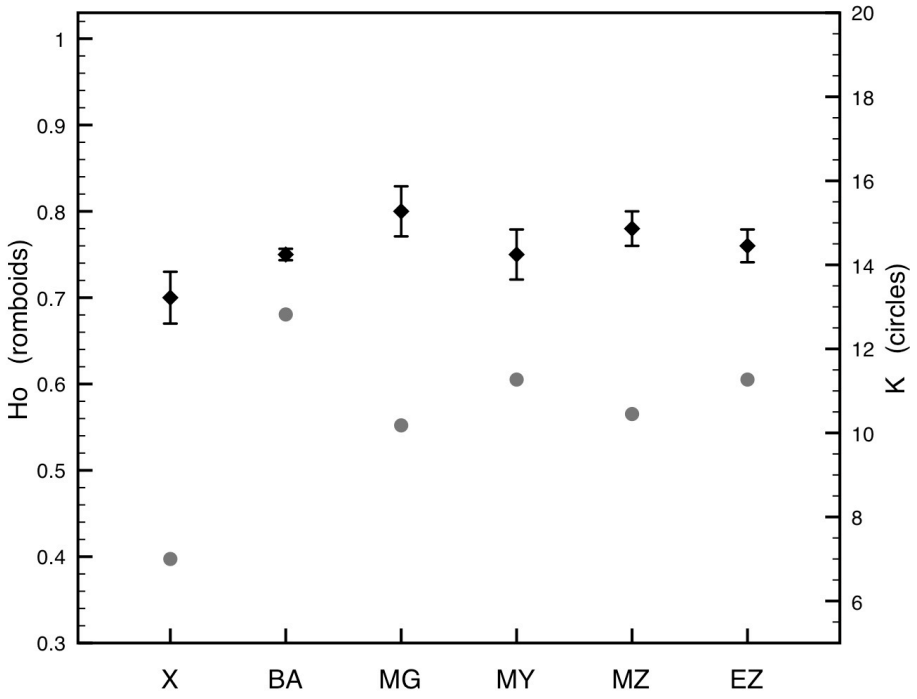


Figure 7 Bottleneck analyses (Garza, Williamson, 2001). *M* is the ratio between number of alleles and range of the allelic array. *M*c is the minimum critical value for the ratio in an equilibrium population and for different historical diversity values (theta) calculated through simulations The conservative two-phase mutation model assumes the proportion of one-step mutations to be p_s=90% and the average size of non one-step mutations Δ_g=3.5. The realistic model is based on literature data (p_s=88%, Δ_g=2.8).