Patterns of genetic variation in Southern Hemisphere blue whales and the use of assignment test to detect mixing on the feeding grounds

R.G. LEDUC*, A.E. DIZON*, M. GOTO+, L.A. PASTENE+, H. KATO#, S. NISHIWAKI+, C.A. LEDUC* AND R.L. BROWNELL*

ABSTRACT

A total of 111 samples from Southern Hemisphere blue whales were sequenced for 420 base pairs of the mitochondrial control region and all but one of those were genotyped over seven microsatellite loci. Comparisons were made between samples from three broad geographic regions: the southeast Pacific Ocean; Indian Ocean; and around the Antarctic continent. Each of these strata was found to be highly differentiated from the others, in both mitochondrial and nuclear data. The genetic differentiation between the geographic ranges of the nominal subspecies (i.e. true blue whales in Antarctica *vs*. pygmy blues in Pacific and Indian Oceans) was not markedly greater than between the populations of pygmy blue whales. Assignment tests using the microsatellite data provide some insight into detection of feeding-season mixing, although existing methods have some limitations.

KEYWORDS: BLUE WHALE; GENETICS; FEEDING GROUNDS; SOUTHERN HEMISPHERE; PACIFIC OCEAN; INDIAN OCEAN

INTRODUCTION

Blue whales (*Balaenoptera musculus*) in the Southern Ocean are currently divided into two subspecies, the pygmy blue whale (*B.m. brevicauda*) and the true blue whale (*B.m. intermedia*). The subspecies differ in body size and proportions and ostensibly segregate latitudinally during the austral summer, with the true blue whales occurring primarily south of the Antarctic convergence and the pygmy blue whales to the north in the southern Pacific, Atlantic, and Indian Oceans. Presumably the subspecies maintain separate breeding grounds, although little is known about their movements during the winter. Of special concern to the management of southern blue whales is the potential for mixing of the two subspecies on the feeding grounds. The catch data on body lengths indicates that although the summer segregation is nearly complete, the occurrence of some pygmy blue whales at high latitudes in the Antarctic has been inferred, with the estimates of their proportion of all Antarctic blue whales ranging from 0.28% (Kato *et al*., 2000) to about 2% (Donovan, 2000). These values were based on data collected decades ago, and subsequent changes in the relative abundances of the two subspecies raise the possibility that the proportion may be much higher today. In other words, the historical abundance of true blue whales was much greater than that of pygmy blue whales, and the proportion of pygmy blue whales in high latitudes may be much greater now that pygmy whales are more abundant than true blue whales. Based on a Bayesian analysis of morphological and behavioural characters from photographs and video, Kato *et al*. (2002) concluded that between 2.3 and 6.9% of blue whales sighted in Antarctic waters might have been of the pygmy form. In addition, it is plausible that some Antarctic blue whales could occur at low latitudes. Aguayo (1974) reported both forms as occurring in catches in Chilean waters. Stafford *et al*. (1999; 2004) reported some blue whale calls of the type typical of Antarctica were recorded in fall and winter from low latitudes south of the equator in the Indian and Pacific Oceans. It is important to try to determine the extent of mixing during the austral summer feeding season, because this is the season when the vast majority of the southern blue whale research, including surveys to estimate abundance, is conducted.

In this paper, measurement of the genetic differentiation between blue whale subspecies and among ocean basins is attempted, using biopsies taken during various research efforts. However, the aforementioned seasonal bias in research efforts presents problems for this type of analysis. Breeding ranges are an important part of understanding the differentiation of subspecies. Since these ranges are largely unknown for southern blue whales and all the available samples are from outside the breeding season, there is no way to confidently determine the subspecies membership of any given sample based solely on geographic location. Although there are measurable differences between the subspecies in body size, proportions, and other morphological characters, these are difficult to ascertain for the free-swimming whales that are the source of the present tissue samples, so subspecies identity based on appearance is also questionable. Due to these limitations, any *a priori* stratification of the data into subspecies is at best equivocal. Nevertheless, the feeding ground segregation was used as a proxy for subspecies identity; whales sampled in Antarctica are provisionally identified as true blue whales and those sampled in lower latitudes as pygmy blue whales. This is congruent with the results of McDonald *et al*. (2006), which found strong differentiation in blue whale calls among feeding areas in different ocean basins. In other words, here the samples from the different feeding grounds are treated as representative of different breeding populations, with gene flow assumed to be negligible and mixing assumed to be minimal. From this starting point, traditional and Bayesian analyses of genetic data are used to infer the extent of genetic differentiation of the two forms and to try to detect evidence of mixing.

* *Southwest Fisheries Science Center, 8604 La Jolla Shores, La Jolla CA 92037, USA.*
+ *The Institute of Cetacean Research, Toyomi-cho 4-5, Chuo-ku, Tokyo 104-0055, Japan.*
# *Laboratory of Marine Biology, Department of Ocean Sciences, Tokyo University of Marine Science and Technology 4-5-7 Konan, Minato-Ku, Tokyo 108-8477, Japan.*

MATERIALS AND METHODS

Biopsy samples from Antarctica, Chile and the southwestern Indian Ocean were obtained from International Whaling Commission/Southern Ocean Whale and Ecosystem Research (IWC/SOWER) cruises and Japanese Whale Research Program under Special Permit in the Antarctic (JARPA) research cruises. In addition, samples from Ecuador and Peru were obtained from research cruises conducted by the Southwestern Fisheries Science Center (SWFSC), and samples from the Maldives were obtained on an Indian Ocean charter survey cruise aboard the *M/V Girity*. Tissue samples from Australian waters were collected by researchers from the Western Australian Museum and the South Australian Museum. In total, there were 47 samples from around the Antarctic continent south of the convergence, 28 from the southern and western coasts of Australia, two from around the Maldives, six from off the east coast of South Africa, 16 from the coast of Chile and 12 from Peruvian and Ecuadorian waters.

An approximately 400 base pair region of the mitochondrial control region was amplified from extracted genomic DNA (Qiagen *DNeasy* No. 69506) using primers H16498 (5'-*cctgaagtaagaaccagatg*-3') (Rosel *et al*., 1994) and L15812 (5'-*cctccctaagactcaaggaag*-3'; developed at SWFSC). Using the same primers, both strands of the amplified DNA product were sequenced independently using standard protocols on the Applied Biosystems Inc. (ABI) model *3100 Sequencer*. All sequences were aligned using *Sequencher* software, version 4.1 (Genecodes).

Each sample was typed for seven polymorphic microsatellite loci: ACCC392, GATA028, GATA098 and GATA417 from Palsbøll *et al*. (1997), GT023 from Bérubé *et al*. (2000), EV37 from Valsecchi and Amos (1996) and DlrFCB17 from Buchanan *et al*. (1996). Polymerase Chain Reaction (PCR) amplifications were carried out in 25ml reaction volumes using the following conditions:

MQ Water		18.35
10 X Buffer	20mM MgCl2	2.5
dNTP's	10mM	1.5
FWD	10uM	0.75
REV	10uM	0.75
Taq DNA Pol	5u/ul	0.15
DNA	50-100ng/ul	1

The following PCR cycling conditions were used for all primers, but note varying annealing temperatures between 52°C and 55°C.

	1st Denature	Denature	Anneal	Extend	Final Extend	No. Cycles
Temp	92	94	52-55	72	72	35
Time	2 min	10-45 sec	30 sec-1 min	1 min-1 min 30 sec	3 min	

Samples were run on an ABI *3100* as per the ABI protocols.

Random samples from each locus were replicated from the DNA with 100% matching.

In addition, each animal was sexed genetically. The samples from the JARPA cruises were sexed according to the method by Abe *et al*. (2001). The rest were sexed according to Fain and LeMay (1995).

The data were stratified into three groups: Antarctica; Indian Ocean; and Pacific Ocean. The analysis focused on examining variation on large geographic scales. However, it is possible that there is population structure within each stratum. The samples of blue whales from both the Indian and Pacific Ocean basins were collected from areas within these regions that were geographically disjunct. For example, the Indian Ocean samples were collected from the southern and western coasts of Australia, the southwest Indian Ocean near Madagascar and from near the Maldives. Similarly, the samples from the Pacific came from the waters off Chile as well as off Peru and Ecuador. However, it is probably the Antarctic sample set that has the greatest chance of including divergent breeding populations. Assuming these whales move to lower latitudes to breed, it is plausible that these whales segregate into different ocean basins at that time and thus may be isolated during the breeding season. The sample sizes herein are too small to adequately test levels of difference within ocean basins. In the following analyses and discussion, therefore, the focus is on differences at the subspecies and ocean basin levels, but should not be used to imply any uniformity or panmixia within these larger strata. The working assumption is that any population structure that may exist within ocean basins, which may be detectable with larger sample sizes, are developed to a much lesser degree than differences between ocean basins or between pygmy and true blue whales. For example, in the case of the whales in the northern Indian Ocean (i.e. the two samples from the Maldives), this means that they are assumed to be less differentiated from the whales in the other parts of the Indian Ocean than they are from those in the Pacific or the Antarctic. This assumption is consistent with the degree of similarity in blue whale calls recorded in different parts of the Pacific and Indian Ocean basins (McDonald *et al*., 2006; Stafford *et al*., 1999).

There are some discrepancies between the number of samples sequenced and the total number of alleles generated for the microsatellite markers, resulting from a few instances of failed amplification of the microsatellite markers. These missing data are proportionally few in number and likely do not affect the overall results. There was one sample for which the sequence could be obtained but no microsatellite data could be derived. This sample was therefore excluded from the analyses of nuclear data. The phylogenetic tree was generated using *PAUP** v. 4.0b10 (Swofford, 2001), using the neighbour-joining algorithm and simple percent genetic distances between mitochondrial haplotypes. Pairwise population comparisons using F_{st} for the mitochondrial haplotype data was performed using *Arlequin* v. 3.0 (Excoffier *et al*., 2005). For the genotype data, *GenePop* v. 3.1c (*http://wbiomed.curtin.edu.au/genepop/index.html*) was used for all population analyses, including examining genic differentiation between population pairs for each locus separately as well as across all loci using Fisher's exact test (Raymond and Rousset, 1995).

Analysis of individual assignment probabilities was performed using two methods. Manel *et al*. (2005) reviewed the use of different assignment test methods for various biological questions. Using their criteria, the present question (determining population membership of individuals and not estimating migration rates, population sizes or divergence times) suggests the use of a Bayesian approach and of an exclusionary assignment test (Cornuet *et al*., 1999). *Structure* v. 2 (Pritchard *et al*., 2000) employs Bayesian analysis of diploid genetic data such as microsatellite data to determine the most probable number

of populations (*K*) given the data and infers probabilities of membership of each sample in each of *K* genetic clusters. For the exclusionary assignment test, the web-based program *Doh* (*http://www2.biology.ualberta.ca/jbrzusto/Doh.php*) (Brzustowski, 2002) was employed, using the options for collecting extremal statistics and drawing new individuals from each population gene pool. In this approach, the allele frequencies of each sample stratum were used to generate 10,000 random genotypes, whose assignment likelihoods to their own population were calculated using the method of Paetkau *et al*. (1995). For each real sample, a *p*-value was then generated which represents the proportion of the random genotypes that had assignment scores less than that value. Samples with scores less than a threshold value (e.g. 0.05 or 0.01) were determined to be unlikely to have come from the population in which they were sampled. By changing the population membership of these samples, they were then tested to see if they scored above the threshold value for one of the other possible populations. If this occurred, the individual could provisionally be considered to have moved from one population to another. The assignment of individuals from both methods can be evaluated in light of independent data, such as geographic location or mitochondrial sequence data. For example, if some of the animals that display haplotypes shared between regions are indicated to be members of a cluster or population characteristic of a different geographic region, this strengthens the inference that they moved from the other region.

For the determinations of population membership probabilities, the most stringent criteria for detecting dispersers were used. For example, for the *Structure* analysis, we assumed that there is no admixture (i.e. gene flow) between strata. The highly significant differences seen in the other analyses (see below) support this assumption. This is also reasonable given our current knowledge; mixing on the feeding grounds does not necessarily indicate interbreeding. In addition, prior population information, in this case geographic origin, was used to help define the genetic clusters, although it does not constrain the assignment probability for any given sample. The distribution of alleles was assumed to be independent. There were 100,000 MCMC (Markov Chain Monte Carlo) repetitions after a burn-in of 50,000 years. Default values were used for all other settings. The complete parameter settings for *Structure* are given below:

Values of parameters used in *Structure*:
NUMINDS=110, NUMLOCI=7, MISSING=-9,
LABEL=1, POPDATA=1, POPFLAG=0, PHENOTYPE=0,
EXTRACOLS=0, MAXPOPS=3, BURNIN=50000,
NUMREPS=100000, USEPOPINFO=1, INFERALPHA=1,
INFERLAMBDA=1, POPSPECIFICLAMBDA=0,
POPALPHAS=0, COMPUTEPROB=0, NOADMIX=1,
ADMBURNIN=25000, UPDATEFREQ=1,
PRINTLIKES=0, INTERMEDSAVE=0, PRINTKLD=0,
PRINTLAMBDA=0, ANCESTDIST=0,
NUMBOXES=1000, ANCESTPINT=0.90000,
GENSBACK=0, MIGRPRIOR=0.05000, PRINTQHAT=0,
PRINTQSUM=0, ALPHA=1.0000, FREQSCORR=0,
FPRIORMEAN=0.1000, FPRIORSD=0.1000, ONEFST=0,
LAMBDA=1.0000, UNIFPRIORALPHA=1,
ALPHAMAX=10.0000, ALPHAPRIORA=1.0000,
ALPHAPRIORB=2.0000, ALPHAPROPSD=0.0500,
STARTATPOPINFO=0, RANDOMISE=1, LINKAGE=0,
METROFREQ=10, REPORTHITRATE=0,
MARKOVPHASE=-1, PHASED=0, PLOIDY=2,
PHASEINFO=0.
[STRAT parameters]: NUMSIMSTATS=1000,
PHENOTYPECOL=-9, POOLFREQ=10,
LOCUSxONLY=0, EMERROR=0.00100,
MISSINGPHENO=-9,

Sequence haplotypes have been deposited in GenBank (Accession Numbers EU093919 – EU093962).

Complete sample and genetic data are available at *http://www.iwcoffice.org/publications/additions.htm*.

RESULTS AND DISCUSSION

Population differentiation

Mitochondrial DNA

Control region sequencing yielded 44 unique haplotypes for the entire sample set. Fig. 1 and Table 1 provide information on frequencies and summary statistics for each geographic region. There were relatively few haplotypes shared between the three primary geographic regions: only one (*d*) between the Indian Ocean and the Antarctic, one (*k*) between the Pacific Ocean and the Antarctic; and one (*r*) among all three strata. As Fig. 1 shows, there is no distinct phylogeographic pattern to the haplotype relationships, although the Indian Ocean haplotypes are somewhat less disparate on the tree than those from the other regions.

The Antarctic sample set had the highest haplotypic diversity. This is a reflection of not only the greater number of haplotypes present in that sample set, but also their more

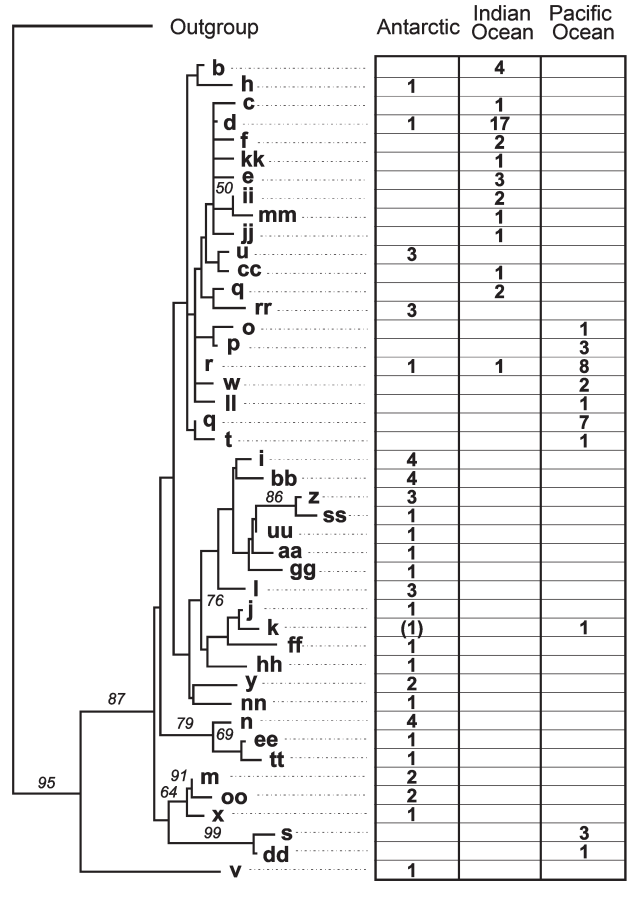

Fig. 1. Neighbour-joining tree of southern blue whale haplotypes, with frequencies of haplotypes shown from each region. Numbers of nodes indicate bootstrap values >50%, unnumbered nodes supported in <50% bootstrap replications. Tree was generated using uncorrected percent differences between sequences. The number in parentheses (1) for haplotype *k* in Antarctica indicates the sample that was sequenced but not genotyped.

Table 1

Summary statistics for mtDNA control region sequence data for each strata: Antarctica (true subspecies), Indian (pygmy subspecies), Pacific (pygmy subspecies). For F_{ST} comparisons, *p*-values are in parentheses.

Population	Antarctica	Indian Ocean	Pacific Ocean
No. haplotypes/no. samples	26/47	12/36	10/28
Haplotypic diversity	0.9685 +/- 0.0099	0.7651 +/- 0.0698	0.8519 +/- 0.0422
F_{ST} (p) *vs.* Antarctica	-	0.122 (<0.0001)	0.081 (<0.0001)
F_{ST} (p) *vs.* Indian Ocean	-	-	0.186 (<0.0001)

equitable frequency distribution; the three most common Antarctic haplotypes each occurred at a frequency of 8.5%, *vs*. 47.2% and 28.6% for the single most common haplotypes for the Indian and Pacific Ocean samples, respectively. These relative levels of diversity are consistent with the much greater historical abundance calculated for true blue whales relative to pygmy blue whales (Gambell, 1976). Table 1 also shows that the differences between regions are highly significant. This is not surprising, given the large differences in haplotype frequencies and the low number of shared haplotypes. The most divergent stratum, based on F_{ST} values, appears to be the Indian Ocean samples. One should be cautious about comparing or ranking F_{ST} values, especially when all are highly significant. However, this seems to indicate that the differences between the two populations of pygmy blue whales (Indian and Pacific Oceans) is at least as much as either shows to the putative population of true blue whales (Antarctic).

Microsatellites

Table 2 shows the allele frequencies for each of the seven microsatellite loci. The results parallel those of the sequence data. Allele frequency differences across all strata were highly significant ($p<0.001$) for each locus. For pairwise comparisons between strata, only three were non-significant; these were Antarctic-Pacific Ocean comparisons using DlrFCB17, GATA028 and GT023. All of the rest, including Antarctic-Pacific Ocean comparisons using the other four loci, were significant. Parallel to the sequence data results, the Indian Ocean samples also showed the lowest level of diversity, having the lowest heterozygosity for five of the seven markers.

It is clear from the results of both the sequence and microsatellite analyses that the three strata herein described are highly differentiated, showing not only significant differences in haplotype and allele frequencies, but differences in levels of diversity, with the Indian Ocean having the lowest diversity for both types of data.

Both the sequence and microsatellite data indicate that the three regions are quite divergent, with the differences between ocean basins being at least as large as the differences between subspecies. This pattern is consistent with a recent study on the geographic pattern of variation in blue whale calls (McDonald *et al*., 2006). However, the pattern and degree of genetic differentiation do not necessarily reflect recentness of contact between these regions. As the driving mechanism of differentiation for neutral alleles, genetic drift is strongly affected by (and is inversely proportional to) population size. In this respect, the divergence of the Indian Ocean blue whales may reflect in part their smaller historical population size. In other words, they may have differentiated faster, rather than over a longer period of time. So while the patterns of genetic divergence do not necessarily negate the status of the Indian and Pacific Ocean samples as members of the same subspecies, the results do indicate that a reconsideration of blue whale taxonomy is in order. However, no matter the historical rate of differentiation or taxonomic arrangement, contemporary gene flow between regions is likely to be negligible.

Testing for vagrants

In this section the terms 'vagrant' or 'candidate for mixing' are used, rather than the usual 'disperser' or 'migrant'. This is to emphasise that these individuals are sampled on the feeding grounds and are not necessarily (and are unlikely to be) vectors for gene flow between populations. In the analysis using *Structure* (Pritchard *et al*., 2000), the first step was that of determining the most likely number of populations given the data. In this analysis, by far the most likely number of populations was three ($p>0.999$). Again, this does not mean that there is not some population structure within these groups, just that three is the most likely number of genetic clusters at this level of resolution.

The resulting probability of each individual's membership in each of the genetic clusters is shown graphically in Fig. 2. Here, the three genetic clusters strongly correspond to the three geographic regions (Antarctic, Indian and Pacific Oceans, respectively), and are referred to as such for ease of discussion. Most individuals had high probabilities of assignment to their own geographic cluster. Only two individuals (no. 36 and no. 103 in Fig. 2) had high probabilities of belonging to a cluster comprised mainly of animals from a different geographic region; i.e. potential vagrants. Each of these involved the Antarctic and Pacific Ocean; no movements involving the Indian Ocean were indicated. Number 36 (sample 7342 in Table 3) was a whale sampled in Antarctica at over 70°S in Area II. The analysis gave it a 0.941 probability of belonging to the Pacific Ocean cluster and only a 0.059 probability of belonging to the Antarctic cluster. Number 103 (sample 11172 in Table 3) was an animal sampled at about 40°S off the coast of Chile and it had a 0.924 probability of belonging to the Antarctic cluster, and 0.076 to the Pacific. There were other individuals with more equivocal assignments (e.g. no. 93 in Fig. 2), but these were not considered to be strong candidates for mixing.

The *Structure* analysis therefore proposes two potential vagrants based on the microsatellite data, in a sense indicating the occurrence of a true blue whale off Chile and a pygmy blue whale off Antarctica. Given the low numbers of shared haplotypes between regions, these results can be compared to the sequence data to identify parallels. The haplotype of the Chilean sample no. 103 (*k*) was also recorded in Antarctica (Fig. 1), lending some support to the designation of this animal as a vagrant from the south. Unfortunately, the Antarctic sample that had haplotype *k*

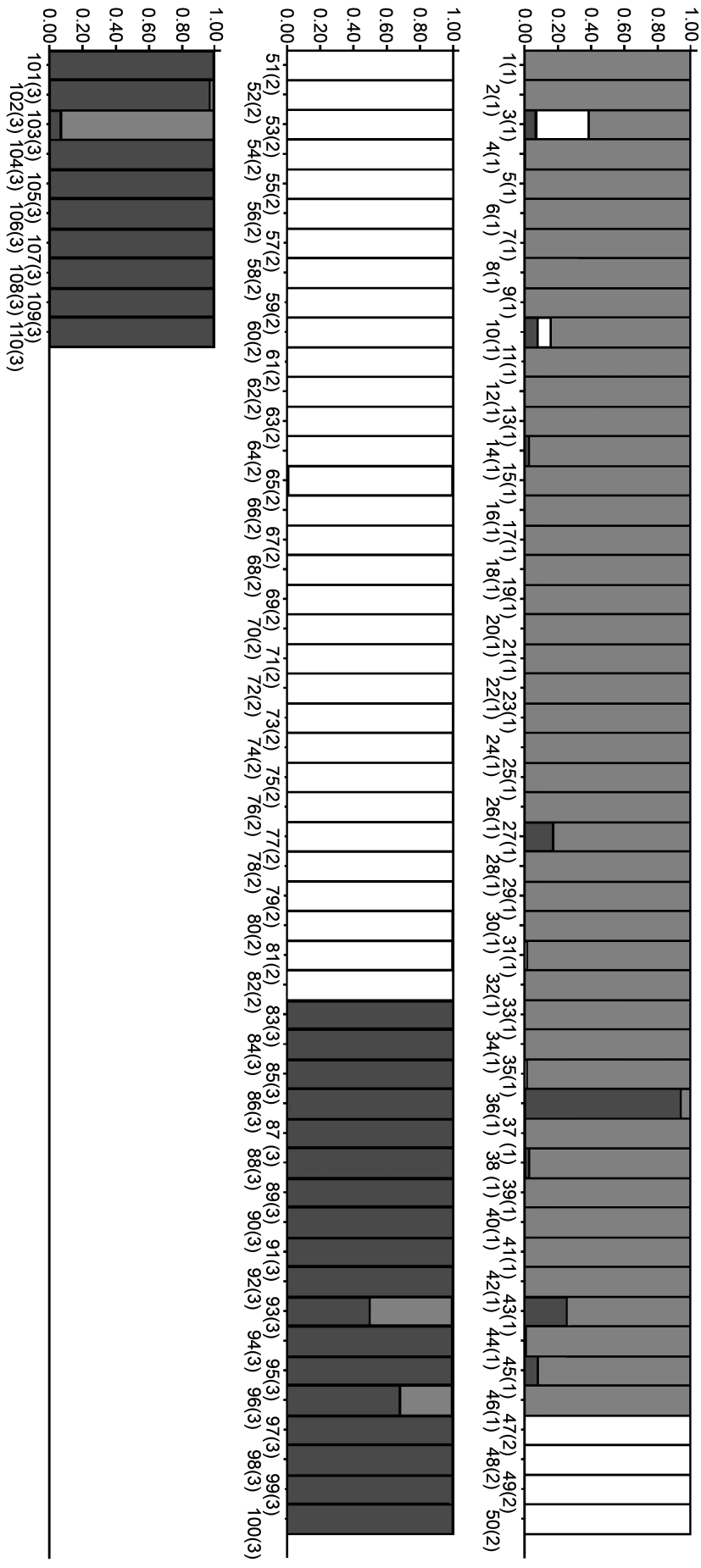

Fig. 2. Assignment probabilities for each of the samples run for the microsatellite loci. Numbers in parentheses indicate geographic origin of each sample; samples 1-46 are from the Antarctic, samples 47-81 are from the Indian Ocean and samples 82-110 are from the Pacific Ocean. Light grey indicates the assignment probability to the first cluster, white to the second and dark grey to the third.

Table 2

Allele counts and level of heterozygosity for each geographic region at seven microsatellite loci. The order of alleles from each locus is from smallest (left) to largest (right).

Population																		
Ev37 - Heterozygosity																		
Ant.	0.663	4	0	24	2	45	17	0										
Ind.	0.653	26	2	32	0	12	0	0										
Pac.	0.752	8	0	5	0	22	14	7										
DlrFCB17 - Heterozygosity																		
Ant.	0.461	9	1	67	3	3	1	2	1	1	2	2						
Ind.	0.503	33	0	39	0	0	0	0	0	0	0	0						
Pac.	0.416	8	0	42	0	5	0	1	0	0	0	0						
GATA098 - Heterozygosity																		
Ant.	0.788	1	2	1	1	1	8	18	16	34	7	3						
Ind.	0.636	0	7	0	0	0	0	0	19	37	7	0						
Pac.	0.626	0	0	0	0	0	0	5	16	30	5	0						
GATA028 - Heterozygosity																		
Ant.	0.861	10	5	0	1	6	19	16	18	9	4	2						
Ind.	0.786	1	0	1	0	7	7	24	9	20	3	0						
Pac.	0.806	2	3	0	0	0	14	18	7	8	2	2						
GT023 - Heterozygosity																		
Ant.	0.842	3	2	19	23	8	15	13	7	1	1							
Ind.	0.664	0	2	1	13	0	33	23	0	0	0							
Pac.	0.774	1	1	20	14	3	1	11	5	0	0							
ACCC392 - Heterozygosity																		
Ant.	0.789	2	13	29	5	28	2	4	0	0	1	1	1	1	0	4	0	1
Ind.	0.678	11	1	36	0	5	0	0	2	0	0	0	5	0	8	0	0	0
Pac.	0.836	0	14	14	0	3	0	9	0	2	1	0	5	0	0	1	7	0
GATA417 - Heterozygosity																		
Ant.	0.858	5	0	3	4	6	0	1	8	20	21	15	1	1	2	0	3	
Ind.	0.843	1	10	0	1	1	0	2	14	20	8	5	1	1	0	6	0	
Pac.	0.853	0	0	1	16	1	3	4	10	9	5	3	2	2	0	0	0	

failed to amplify for any of the microsatellite markers and was therefore not genotyped. In spite of this gap in the data, the assignment results based on microsatellite data are congruent with the sequence results for no. 103. In other words, that haplotype was recorded once in the Pacific Ocean and once in the Antarctic, and the microsatellite results indicate that that Pacific Ocean sample was actually a vagrant true blue whale. If this is indeed the case, then that haplotype may not actually occur in the breeding population of pygmy whales in the SE Pacific Ocean.

The results for sample no. 36, sampled in the Antarctic, are more uncertain. The haplotype (*j*) for that sample was unique to that individual, so the sequence data neither agree with nor contradict the genotype assignment. However, one could make the prediction that if no. 36 is actually a vagrant whale that was born in the SE Pacific Ocean population, then the *j* haplotype should eventually be recorded there.

For the exclusionary assignment test, Table 3 shows all 11 of the individual samples whose *p*-value was less than 0.05 for the population in which they were sampled. For six of these samples, the *p*-values for each of the other populations were also less than 0.05. By this criterion, these six would not be considered vagrants. Their low probability of membership to each of the populations is likely due to their genotypes containing alleles that are relatively rare in each of the populations. Five of the samples (those with asterisks) would be considered vagrants, in that their *p*-values for one of the other populations was greater than 0.05. However, only one of these (11172) was indicated by *Structure* to be a potential vagrant; all the rest were strongly assigned by *Structure* to the population in which they were sampled. In this respect, the results from *Structure* and the exclusionary assignment test are difficult to reconcile.

The results of the two analyses can be reconciled if we increase the stringency of the exclusionary test. If the critical *p*-value is set to 0.01, then only the first two samples (7342 and 11172) pass the exclusion test for vagrancy status. Each is excluded from its sample population (Antarctic and Pacific, respectively) and is included in only one other population (Pacific and Antarctic, respectively). This result coincides exactly with the results from *Structure*.

CONCLUSIONS

The results of the present analyses indicate that the three regions in which blue whales were sampled are highly differentiated in both mitochondrial and nuclear genes and that the blue whales in the Indian Ocean appear to be the most highly divergent of the three. The most conservative assignment analysis indicated only a single potential pygmy blue whale in the Antarctic sample set of 46 samples, a result not out of line with other estimates of mixing; i.e. it is very similar to the 2% estimated by Donovan (2000). It also indicated a single potential true blue whale in the southeast Pacific sample set of 28. This is an area where whaling catches included blue whales of both forms (Aguayo L, 1974). This approach for detecting potential vagrants holds promise, although the current results (and the actual identities of those potential vagrants) should be considered provisional. Each of the methods employed here incorporate the Hardy-Weinberg equilibrium within populations, either as an assumption (standard assignment test) or as a parameter to be maximised (*Structure*). Population structure within the broad geographic regions could compromise the ability to detect vagrants. However, this would more likely be manifested as an increase of equivocal assignments

Table 3

Samples excluded from their source population at the p<0.05 level. Those with asterisks would be assigned to another population using the same criterion. Samples in bold are those assigned to a different population using the p<0.01 levels in the exclusion test, and by Structure.

	Exclusion				*Structure*		
ID no.	*P* (SOURCE)	*P* (ANT)	*P* (IND)	*P* (PAC)	*P* (ANT)	*P* (IND)	*P* (PAC)
7342	**0.001**	-	<0.0001	**0.012**	**0.059**	<0.001	**0.941**
11172*	**0.007**	**0.593**	<0.001	**-**	**0.924**	<0.001	**0.076**
11167	0.013	-	<0.0001	0.001	0.999	<0.001	0.001
9865*	0.013	0.072	**-**	0.008	0.001	0.999	<0.001
7328*	0.016	0.1925	**-**	<0.001	0.003	0.997	<0.001
11170*	0.018	0.203	<0.001	**-**	0.029	<0.001	0.971
13187	0.028	-	<0.0001	<0.001	1.0	<0.001	<0.001
7622	0.029	-	<0.0001	<0.001	1.0	<0.001	<0.001
11895	0.029	0.040	<0.0001	-	0.001	<0.001	0.999
23981*	0.047	0.136	**-**	0.002	0.001	0.999	<0.001
26584	0.048	-	<0.001	0.037	0.742	<0.001	0.258

(individuals with uncertain status), rather than erroneous designations of individuals as vagrants. Although it is an imperfect test for population structure, tests for Hardy-Weinberg equilibrium on the three strata showed one marker in the Pacific Ocean and two in the Indian Ocean to be significantly out of equilibrium at the 0.05 level. None of the markers were out of equilibrium for the Antarctic.

The analyses to detect vagrants could be improved by adding more markers to increase the statistical power and by adding samples to better characterise allele frequencies within populations. The additional samples would also serve to better test (and perhaps incorporate) population structure within the ocean basins. In addition, the complete lack of samples from the South Atlantic Ocean and the western South Pacific Ocean means that the overall picture is far from complete, and the possible contribution of those 'ghost' populations to the present dataset cannot be estimated at this time. Future efforts will focus on increasing the sample size and refining methods of population assignments. Ideally (but perhaps unrealistically), this would include substantial sampling of the breeding grounds, which would enable better ground-truth characterisation of the separate gene pools. Lastly, the comparison of genotype-based assignment tests to patterns of mitochondrial differentiation should be done with caution. Here, the sharing of haplotypes was used as a line of evidence for potential mixing. However, this is certainly an imperfect line of evidence. Even with complete separation of the populations, haplotypes could still be shared due to incomplete lineage sorting.

The current results also highlight the need for a re-examination of blue whale taxonomy. Most notably, the two populations of pygmy blue whales were just as differentiated from each other as either was to the true blue whales. This is in contrast to the patterns of morphological variation. In addition, the non-migratory blue whales from the northern Indian Ocean (e.g. the Maldives samples) are likely reproductively isolated from the other samples from the Indian Ocean, and indeed have been considered a separate subspecies, *B. m. indica*, (Brownell and Donaghue, 1994). However, this isolation is incongruent with their genetic similarity to the Australian and African samples. The present data set is too limited to address these issues, but hopefully will provide some impetus for additional genetic and non-genetic research.

ACKNOWLEDGEMENTS

We would like to thank all the institutions and people involved in making the samples available, specifically R. Pitman and P. Ensor, K. van Waerebeek and other researchers who collected samples on the IWC/SOWER and JARPA cruises, C. Kemper and S. Donnellan (South Australia Museum), and J. Bannister (Western Australia Museum) who provided samples collected by C. and M.-N. Jenner. Thanks go to N. Kellar, J. Hyde and K. Coultrup for generating the data, to H. Yoshida for arranging to send samples from Japan to the US and to K. Robertson for arranging all the permits for importing samples into the US. Thanks also to B. Taylor and P. Morin and two anonymous reviewers for helpful comments on the manuscript and to E. Archer and K. Martien for valuable discussions regarding the analyses.

REFERENCES

Abe, H., Goto, M., Pastene, L.A., Dewa, K. and Naito, E. 2001. Practical use of multiplex flourescent PCR for cetacean sex identification. *Mar. Mammal Sci*. 17(3): 657-64.

Aguayo L, A. 1974. Baleen whales off continental Chile. pp.209-17. *In*: Schevill, W.E. (eds). *The Whale Problem: A Status Report*. Harvard University Press, Cambridge, Mass. x+419pp.

Bérubé, M., Jorgensen, H., McEwing, R. and Palsbøll, P.J. 2000. Polymorphic di-nucleotide microsatellite loci isolated from the humpback whale, *Megaptera novaeangliae*. *Mol. Ecol*. 9: 2,181-2,83.

Brownell, R.L. and Donaghue, M.A. 1994. Southern Hemisphere pelagic whaling for pygmy blue whales: review of catch statistics. Paper SC/46/SH6 presented to the IWC Scientific Committee, May 1994 (unpublished). 9pp. [Paper available from the Office of this Journal].

Brzustowski, J. 2002. Doh assignment test calculator, *http://www2.biology.ualberta.ca/jbrzusto/Doh.php*.

Buchanan, F.C., Friesen, M.K., Littlejohn, R.P. and Clayton, J.W. 1996. Microsatellites from the beluga whale *Delphinapterus leucas*. Mol. Ecol. 5: 571-75.

Cornuet, J.M., Piry, S., Luikart, G., Estoup, A. and Solignac, M. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153: 1989-2000.

Donovan, G.P. 2000. Possible occurrence of pygmy blue whales (*Balaenoptera musculus brevicauda*) south of 60°S. Paper SC/52/OS15 presented to the IWC Scientific Committee, June 2000, Adelaide, Australia (unpublished). 15pp. [Paper available from the Office of this Journal].

Excoffier, L., Laval, G. and Schneider, S. 2005. Arlequin ver. 3.0: an integrated software for popultion genetics data analysis. Submitted.

Fain, S.R. and LeMay, J.P. 1995. Gender identification of humans and mammalian wildlife species from PCR amplified sex-linked genes. Proceedings of the Annual Meeting of the American Academy of Forensic Science. Seattle, Washington February 1995.

Gambell, R. 1976. World whale stocks. *Mammal Rev*. 6(1): 41-53.

Kato, H., Honno, Y., Yoshida, H., Kojima, E., Nomura, A. and Okamura, H. 2002. Further developments on morphological and behavioural key for sub-species discrimination in southern blue whales, analyses from data through 1995/96 to 2001/02 SOWER cruises. Paper SC/54/IA8 presented to the IWC Scientific Committee, April 2002, Shimonoseki, Japan (unpublished). 16pp. [Paper available from the Office of this Journal].

Kato, H., Okamura, H., Nomura, A. and Kojima, E. 2000. Body length distribution and sexual maturity of southern blue whales, with special reference to sub-species separation. Paper SC/52/OS4 presented to the IWC Scientific Committee, June 2000, Adelaide, Australia (unpublished). 10pp. [Paper available from the Office of this Journal].

Manel, S., Gaggiotti, O.E. and Waples, R.S. 2005. Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol*. 20(3): 136-42.

McDonald, M.A., Mesnick, S.L. and Hildebrand, J.A. 2006. Biogeographic characterisation of blue whale song worldwide: using song to identify populations. *J. Cetacean Res. Manage*. 8(1): 55-65.

Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C. 1995. Microsatellite analysis of population structure in Canadian polar bears. *Mol. Ecol*. 4: 347-54.

Palsbøll, P.J., Bérubé, M., Larsen, A.H. and Jorgensen, H. 1997. Primers for the amplification of tri- and tetramer microsatellite loci in baleen whales. *Mol. Ecol*. 6: 893-95.

Pritchard, J.K., Stephens, M. and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-59.

Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal Of Heredity* 86: 248-49.

Rosel, P.E., Dizon, A.E. and Heyning, J.E. 1994. Genetic analysis of sympatric morphotypes of common dolphins (genus: Delphinus). *Mar. Biol*. 119(2): 159-67.

Stafford, K.M., Bohnenstiehl, D.R., Tolstoy, M., Chapp, E., Mellinger, D.K. and Moore, S.E. 2004. Antarctic-type blue whale calls recorded at low latitudes in the Indian and eastern Pacific Oceans. *Deep-Sea Research I*. 51:1,337-1,346.

Stafford, K.M., Nieukirk, S.L. and Fox, C.G. 1999. Low-frequency whale sounds recorded on hydrophones moored in the eastern tropical Pacific. *J. Acoust. Soc. Am*. 106: 3687-98.

Swofford, D.L. 2001. *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0b10*. Sinauer, Sunderland, Massachusetts.

Valsecchi, E. and Amos, W. 1996. Microsatellite markers for the study of cetacean populations. *Mol. Ecol*. 5: 151-56.

Date received: September 2006.
Date accepted: December 2006.