Genetic diversity of humpback whales in the Antarctic feeding ground examined by mitochondrial DNA and microsatellite

Luis A. Pastene[1], Mutsuo Goto[1], Hideaki Abe[1], Shigetoshi Nishiwaki[1] and Per Palsboll[2]

[1]*The Institute of Cetacean Research, Toyomi-cho 4-18, Chuo-ku, Tokyo 104-0055, Japan*

[2]*School of Biological Sciences, University of Wales-Bangor, Deiniol Road, Bangor, Gwynedd LL57 2UW Wales, United Kingdom*

ABSTRACT

Biopsy samples from 144 humpback whales obtained during surveys of the Japanese Whale Research Program under Special Permit in the Antarctic (JARPA), representing four feeding aggregations, were analyzed for mtDNA and nuclear DNA variation. Samples from Areas III (15), IV (73), V (40) and VI (16) were examined for i) sex determination, ii) the sequence variation of the first 333 nucleotides of the mtDNA control region and iii) the genotype at seven microsatellite loci. We also analyzed an additional set of eleven samples from Area I. The level of genetic diversity in the Antarctic was high for both genomes, the nucleotide diversity at the mtDNA was estimated at 0.026 and the mean expected heterozygosity at the nuclear loci at 0.81 from the total sample. In general the mtDNA was more sensitive to separate Antarctic Areas than microsatellite. The majority of pair-wise comparisons among Areas detected significant degrees of heterogeneity at the mtDNA. The only exceptions were between Areas I and IIIE, and VIW, respective as well as between Areas V and VIW. Areas V and VIW are contiguous adjacent Areas, whereas the sample of Area I is relatively small reducing statistical power. Significant deviation from the expected Hardy-Weinberg equilibrium genotype frequencies were detected in the total sample when all loci were combined indicating the existence of nonrandom mating, suggesting that humpback whales from different breeding populations mix on the feeding grounds. A significant level of heterogeneity at nuclear loci was only detected in two cases between Area I and V and Areas IV and V, respectively. However, several other pair-wise comparisons yielded near-to-significant P values. A single instance case of 'mark-recapture' from the 'genetic tag' was observed. A female humpback whale first sampled in the western part of Area V in January 1995 was subsequently 're-sampled' in the eastern part of Area IV in January 2000. Both samples presented the same genotype, same sex determination and identical mtDNA sequence.

KEYWORDS: HUMPBACK WHALE, ANTARCTIC FEEDING GROUND, GENETICS, STOCK IDENTITY

INTRODUCTION

There is a considerable amount of information on the pattern of distribution and seasonal migratory movement of humpback whales in the Southern Hemisphere. Most of that information is derived from the analysis of 'Discovery'-type marks and catch distribution conducted in the past. Mackintosh (1965) showed that humpback whales tend to gather into five or six distinct feeding concentrations in the Antarctic during the austral summer season. He denominated these concentrations as Groups I-V (with a Group IIa and IIb). These concentrations correspond roughly to Management Areas I-VI.

Two of the areas of feeding concentrations in the Antarctic are Areas IV (70°-130°E) and V (130°E-170°W), where JARPA surveys are mainly conducted. The geographic boundary of these two Areas in the Antarctic were defined considering the distribution of catches and the results of mark-recapture analysis in the humpback whale (Omura, 1953; Chittleborough, 1959).

Fig.1 reproduce a figure of Dawbin (1966). This figure summarize the distribution and seasonal migratory movement of humpback whales from Groups IV (W) and V (E, NZ) as demonstrated by mark-recapture data. Whales from Group IV move mainly between Antarctic Area IV and Western Australia. Whales from Group V move between Antarctic Area V and Eastern Australia and along the coast of New Zealand and southwest Pacific islands. It should be noted the interchange of a few individuals between Groups IV and V. Also it should be noted

that the boundary of these Groups in the Antarctic do not correspond to the actual boundary of Areas IV and V. Also it is noted that some whales marked in Area VI were recovered in Eastern Australia.

Genetic analyses in the Southern Hemisphere humpback whale have been based basically in the maternal-inherited mtDNA. While genetic analyses in breeding areas and migratory corridors demonstrated significant mtDNA differences between Western Australia, Eastern Australia and Colombia (Baker *et al.*, 1998), little is known on the genetic population structure in the feeding ground. In a preliminary analysis Pastene and Baker (1997) examined 152 samples from seven Southern Hemisphere localities involving three feeding grounds in the Antarctic and four winter breeding grounds. The pattern of mtDNA heterogeneity was consistent with the hypothesis that the samples used belong to at least three genetic stocks. Samples from the Antarctic feeding grounds were less discrete genetically than the sample from the wintering grounds. Both Western Australia and Area IV were consistently different from most Group V and VI regions to the east. Area VI did not differ significantly from most of the other localities and there was some evidence that the Colombian wintering grounds were not closely linked to Areas IV, V and VI.

Making use of additional biopsy samples collected from Areas III, IV, V and VI by the JARPA, we extend our genetic analysis in the feeding ground to study the sex proportion and to examine the genetic diversity and stock structure using uni-(mtDNA control region sequencing) and bi-parental (microsatellite) inherited genetic markers. This is the first time that nuclear DNA markers are used to investigate stock identity in the Antarctic. For comparison, we included in the analysis samples from Area I (11). The mtDNA control region sequences of these samples have been already published (Palsboll *et al.*, 1995).

MATERIALS AND METHODS

Biopsy samples

Skin biopsy samples were obtained along the sighting surveys of the JARPA, on an opportunistic basis. They have been collected using an air gun described by Kasamatsu *et al.* (1991). To avoid re-sampling, ancillary information is obtained for each individual, among them the estimated body length and visual observations of external characters. All the samples obtained are checked for the possibility of re-sampling by using a set of microsatellite.

Table 1 shows the number of samples used in the analysis on stock structure and the male proportion, by Antarctic Area and year. This table exclude the re-samplings found (see Results). We examined a total of 144 samples, from Areas III (15), IV (73), V (40) and VI (16). Samples in Areas III, IV, V and VI are from three, four, three and two different years, respectively. In three cases calf accompanying the mother were observed. Only data from the mother were used in the analysis. Fig. 2 shows the geographic distribution of the biopsy samples used in the study on stock structure, by Area and sex. Samples from Area III and VI are from their eastern and western parts, respectively. After sampling, skin biopsies were stored at –20°C until use. We also used a set of samples from Area I (11). Thus the total sample size is 155.

Biochemical analysis

Extraction of DNA

Genomic DNA (nuclear+mitochondrial DNA) was extracted from approximately 0.05g of the outer epidermal layer of the skin biopsy. For extracting genomic DNA, we used established protocols (Sambrook *et al.*, 1989). The tissue was homogenized in 500ul of TES buffer. Previous addition of 25ul of Sodium Dodecyl Sulfate (20%), 25ul of Proteinase K (20mg/ml) was added and the homogenate was incubated overnight at 37°C. After incubation, the DNA solution was mixed with an equal volume of a 25:24:1 phenol/chloroform/isoamyl alcohol solution, shaken thoroughly and centrifuged to precipitate proteins. Finally DNA was precipitated by adding 1 ml of 99.5% ethanol and incubating at –70°C for 15min. The genomic DNA was then suspended in 500ul of TE buffer and stored at –20°C until use.

Sex determination

Sex of the whales sampled was determined following the method of Abe *et al.* (unpublished data). The SRY gene from the Y chromosome of humpback whale was isolated. One set of oligonucleotide primers was designed so that the presence/absence of this gene can be detected by PCR amplification. A STR locus (GATA417) was used as internal control. It should be noted that this method was tested using 60 minke whale samples for which gender information was already known. All the determinations obtained by this molecular approach were correct.

Mitochondrial DNA control region sequences

We used the polymerase chain reaction (PCR) to amplify a segment of the control region of the mitochondrial genome following instructions given by Hoelzel (1992). For amplification primers MT4-F (Arnason *et al.*, 1993)

(5'-CCTCCCTAAGACTCAAGGAAG-3') and P2-R (Hori *et al.*, 1994) (5'-GAAGAGGGATCCCTGCCAAGCGG-3') were used. F and R, respectively, denote a forward- or reverse-oriented primer, with reference to the light strand.

The DNA sequences were determined with an automatic sequencer, the Applied Biosystems 377 (ABI 377), following the protocols of the manufacturer. For each sample both strands were sequenced. Sequences were aligned using the 'Sequence Navigator', a DNA sequence comparison software developed by Applied Biosystems.

Microsatellite loci
Seven microsatellite loci (GATA417, GATA28, GATA98, TAA31, GATA53, GGAA520 and GT23) were used (Berube *et al.*, unpublished data; Palsboll, *et al.*, 1997a). They involved five tetranucleotide motifs, one trinucleotide motifs and one dinucleotide motifs. Microsatellite polymorphism was detected fluorescently using end-labeled primers. PCR amplifications were carried out in 15ul reactions containing 5pmol of each labeled and unlabeled primers, 0.625 units of Ex Taq polymerase, 2 mM of each dNTP, reaction reagent and 10-100ng of genomic DNA. Each PCR product was electrophoresed with internal size standard (N,N,N',N',-tetrametyl-6-carboxyrhodamine; TAMRA 500) through 5% polyacrylamide denaturing gel (Long Ranger) using an ABI 377 DNA Prism sequencer.

Data analysis
Analysis of mtDNA
Genetic distances among unique sequences (haplotypes) were estimated using the Kimura's two parameters method (Kimura, 1980). The degree of mtDNA diversity within each geographical locality was estimated using the nucleon (haplotypic) diversity (Nei and Tajima, 1981) and the nucleotide diversity (Nei and Li, 1979). The net genetic distance between Areas was estimated by subtracting the average level of variation within each Area, following equation 10.21 of Nei (1987).

Homogeneity tests between Areas were conducted as described by Hudson *et al.* (1992), using the chi-square statistic (which proved more powerful than either Hst or Kst*). The level of statistical significance was estimated from 10,000 Monte Carlo simulations as the proportion of simulations in which a similar or more extreme value of chi-square was observed. For the total sample in each Area we tested first for differences between male and female. If no significant differences were found we pooled male and female in the subsequent analyses.

Phylogenetic reconstruction of unique sequences (haplotypes) was made using the neighbor-joining (NJ) method (Saitou and Nei, 1987). To evaluate the confidence intervals, we used the bootstrap method (Felsenstein, 1985). The phylogenies were rooted using the homologous sequence from a North Atlantic humpback whale (Arnason *et al.* 1993).

Analysis of microsatellite loci
The level of variation at nuclear loci was estimated as the number of alleles per locus, the expected heterozygosity, and the probability of identity (I) (Paetkau and Strobeck, 1994) as implemented in GENEPOP (ver. 1.31) PC software package (Raymond and Rousset, 1995). I is the probability that two unrelated individuals have the same genotype at a single locus or at all loci.

Deviations from the expected Hardy-Weinberg (HW) genotype frequencies for all loci and Areas were examined using the chi-square test as implemented in the GENEPOP program. We employed the homogeneity test implemented in GENEPOP, which use a Markov Chain method to estimate the values of Fisher's exact test (Raymond and Rousset, 1995). For each locus, an unbiased estimate of the P value was obtained after 10,000 permutations. The P values from the seven loci were combined into a single P value as described by Sokal and Rohlf (1995, p.795).

RESULTS

Re-sampling
By using a set of seven microsatellites, we found that in 11 cases the genotype obtained from two samples was the same. In ten cases, the two samples were taken from the same school at the same time and location within a summer season. Only one of these samples was used in the analysis on stock structure.

In the other case the two samples were taken in different years and location as follow. An individual sampled on 15 January 1995 at 63°07'S, 153°38'E (western part of Area V) was re-sampled again on 16 January 2000 at 63°25'S, 108°58'E (eastern part of Area IV) (Fig. 3). In 1995 the female individual was observed in a school of

three individuals. The estimated body size was 12.8m. In 2000 the animal was observed in a school of two individuals and the estimated body size was 14.8m. These two samples were omitted from the analysis on stock structure.

The overall probability of identity (across all loci), in the Antarctic was estimated at 2.374×10^{-18} yielding an expectation of 2.833×10^{-14} samples having identical genotypes due to chance alone. Apart to present the same genotype, in all cases of re-sampling the two samples presented the same sex determination and identical mtDNA sequences.

The final number of samples used in the analysis on stock structure, included those from Area I, is 155.

Sex ratio

The male proportion in the sample is shown in Table 1, by Area and JARPA survey. The proportions for Areas IIIE, IV, V and VIW were 47%, 56%, 38% and 75%, respectively. For Area I the male proportion was 36%. A chi-square test showed a low but no significant P value (0.077) when the sex ratio was compared among Areas. Table 2 shows the distribution of the sexes by school size and Area. In general male and female were observed in school size of two individuals, followed by schools of three or more individuals. Chi-square test in Areas IV and V showed no significant differences between male and female in relation to school sizes (P=0.488, P=0.915, respectively).

Mitochondrial DNA

Haplotypes

A 333 base pairs of the mtDNA control region was analyzed in the total of 155 samples. A total of 58 polymorphic sites defined 60 haplotypes (Table 3). Apart three transversions and one insertion, all substitutions were transitions.

The frequencies of haplotypes in the five Areas are shown in Table 3. In the 11 individuals from Area I, seven haplotypes were detected, three of which were found only in single specimens (Palsboll *et al.*, 1995). In the 15 individuals from Area IIIE 14 haplotypes were detected, 13 of which were found in single specimens. In the 73 individuals from Area IV, 34 haplotypes were detected, 21 of which were found only in single specimens. In Area V, 23 haplotypes were detected in 40 individuals, 15 of which were found only in single specimens. In Area VI, 12 haplotypes were detected in 16 individuals examined, of which 9 were found in single specimens.

Most of the haplotypes were specific to one of the five localities. However, this was not the case for haplotypes found in more than a single individual. Most of the haplotypes found in Area VI were shared with other Areas. Only two haplotypes were specific to this Area.

Intrapopulational mtDNA diversity

Table 4 shows the haplotypic and nucleotide diversities for each of the five geographic localities examined. Haplotypic diversity for the total sample was estimated at 0.9745. The single estimates for each Area varied from 0.9273 (Area I) to 0.9905 (Area IIIE). Nucleotide diversity for the total sample was 0.0260 and this estimate varied from 0.0230 (Area I) and 0.0281 (Area V). By considering the standard errors of these estimates, no significant differences among Areas were found in the level of mtDNA diversity.

Net genetic distances among Areas

The net genetic distances among Areas were small ranging from –0.00033 (Areas V and VIW) and 0.00071 (Areas IV and VIW).

MtDNA genealogy

Fig. 4 shows the neighbor-joining-based tree of mtDNA haplotypes. Closed circles indicate those nodes for which bootstrap values were above 50% in 400 simulations. Of them, four clades were 'informative' regarding structure. The first, composed of haplotypes '18' and '46' were distributed only in Areas IIIE and IV; the second, composed of haplotypes '12', '13', '24' and '31' were distributed only in Areas IIIE and IV, but mainly in Area IV; the third, composed of haplotypes '50' and '51' were distributed only in Areas V and VIW. The fourth clade, composed of haplotypes '22' and '26', were distributed only in Area IV.

Homogeneity test

No significant degree of heterogeneity was detected between male and female samples in any one Area. Thus during the subsequent analyses, male and female samples for each Area were combined.

Table 5 shows the results of the homogeneity test by chi-square. The overall test showed a significant P value (0.0000) suggesting some degree of structure in the Antarctic sample. The majority of pair-wise comparisons among Areas detected significant degrees of heterogeneity (see Table 5). The only exceptions were between Areas I and IIIE, and VIW, respective as well as between Areas V and VIW.

Microsatellite loci

Levels of polymorphism

The total number of alleles per microsatellite locus ranged from nine to 20 with an average of 14.6. In the Antarctic the mean expected heterozygosity was estimated at 0.81 (range: 0.38-0.91). It varied from 0.74 in Area I (range: 0.44-0.85) to 0.81 in Areas IV (range: 0.54-0.91) and V (range: 0.65-0.90) (Table 6). The overall probability of identity (across all loci), in the Antarctic was estimated at 2.374×10^{-18} yielding an expectation of 2.833×10^{-14} samples having identical genotypes due to chance alone (Table 6).

Tests of HW genotypic proportion

The combined sample from the Antarctic deviated significantly from HW genotypic proportions. Significant deviations from HW genotypic proportions were detected in Areas IIIE, IV and V (Table 7).

Homogeneity test

The following pairwise Areas comparisons yielded significant differences at nuclear DNA: Areas I and V, Areas IV and V. The following pairwise comparisons yielded results near-to-significant: Area I and IV, Areas I and VIW and Areas IIIE and V (Table 8).

DISCUSSION

Levels of genetic diversity

The overall nucleotide diversity in the Antarctic (2.60%) is similar to that obtained for the humpback whale worldwide (2.57%) (Baker *et al.* 1993). These values ranged from 0.86% and 2.90% among feeding aggregations of the humpback whale in the North Atlantic (Palsboll *et al.* 1995). The only previous study involving nuclear loci in the southern humpback whale was that of Valsecchi *et al.* (1997). They analyzed four microsatellite loci in a worldwide study in order to examine the degree of gene flow and divergence among humpback whales from North Atlantic and Southern Hemisphere. The latter sample included samples from two localities: eastern Australia and the Antarctic Peninsula. The expected heterozygosity for these localities was estimated at 0.739 (range: 0.407-0.921) and 0.735 (range: 0.512-0.889), respectively. In our study the expected heterozygosity ranged from 0.74 in Area I (range: 0.44-0.85) to 0.81 in Areas IV (range: 0.54-0.91) and V (range: 0.65-0.90).

Stock structure in the feeding ground as suggested by maternal and bi-paternal genetic markers

Previous studies on stock identity in the feeding ground involved only mtDNA and limited sample sizes (Pastene *et al.*, 1997; Pastene and Baker, 1997). Statistical analysis of mtDNA differences showed Area IV as the only source of genetic heterogeneity. Furthermore samples from the feeding grounds were less discrete genetically that the samples from the wintering grounds.

The limited sample size in terms of individuals as well as loci and genomes of previous studies prompted this study, in order to obtain a more accurate and reliable estimate of the population structure of humpback whale in the Antarctic feeding ground. The statistic used for the analysis of mtDNA, separated almost all the Antarctic Areas and only three pairwise comparisons (out of ten) showing P values above 5% level. One of these cases involved Areas I and IIIE. However, we should note that the sample sizes in these two Areas were the smallest and the fail to detect significant differences could be due to a low statistical power. The other comparisons where no significant differences were found were Areas I and VIW and Areas V and VIW. Both cases involved Area VIW. These are contiguous adjacent Areas. It should be noted here that Pastene and Baker (1997) found that Area VIW did not differ from any of other six regions from breeding and feeding grounds, except from the Colombia sample.

The results of the nuclear DNA analysis also showed substantial level of genetic heterogeneity in the Antarctic humpback whale. Our first approach was the test of HW genotypic proportion. This test yielded a significant P value for all samples and loci combined, suggesting that whales in our sample are not mating randomly. The P values for Areas III, IV and V were significant. This result suggest that some individuals from Group V could be present in Area IV and some from Group IV in Area V, as suggested by studies of mark-recapture in the past (Chittleborough, 1959; Dawbin, 1966). This is also suggested by the result of our study of individual identification that showed a female humpback whale occupying the western part of Area V in 1995 and the eastern part of Area IV in 2000. A similar situation could occur with Groups III and IV. However, the overlap of individuals of

different stocks in the feeding ground does not prevent the detection of genetic differences among them. For instance, Both genomes clearly separated Areas IV and V.

Results of the homogeneity test showed that nuclear DNA is less powerful than mtDNA to separate Areas. Only two significant results were found for all loci combined, Areas I and V and Areas IV and V. These two pairwise comparisons were also significant in the mtDNA analysis. There were, however, some pairwise comparisons that showed near-to-significant P values: Areas I and IV, Areas I and VIW and Areas IIIE and V. All these pairwise comparisons, but Areas I and VIW, were also significant in the mtDNA analysis. In general the microsatellite analysis tended to separate the most distant Area I (see Table 8). Valsechi *et al.* (1997) found no significant differences between eastern Australia (n=100) and Antarctic Peninsula (n=9) using four microsatellite loci. They were unclear whether failure to find a difference between these two localities reflected gene flow between them or lack of statistical power arising from the small size of the Antarctic Peninsula sample.

Berube *et al.* (1998) used mtDNA and microsatellite analysis to analyze fin whales from several localities in the North Atlantic and from the Sea of Cortez. They detected higher levels and more incidences of heterogeneity in mtDNA than in microsatellite. They explained such differences by the lower divergence rate at nuclear loci relative to mt loci or because of male-mediated gene flow among populations. Following these authors, difference in divergence rates are due to differences in the effective population size for each of the two genomes, which is four times larger for nuclear loci relative to mt loci (Berube *et al.*, 1998).

With the analysis of a larger number of samples as well the use of more loci and genome, we were able to detect more incidences of genetic heterogeneity in the Antarctic feeding ground than in the previous study. However the cause seems unresolved. Indeed the deviation from the expected HW proportion shows evidence of mixing of different breeding stocks on the feeding range. However the distribution of genetic variation is not homogeneous indicating some structure. Probably the longitudinal factor is not the only one conditioning the structure of breeding stocks in the Antarctic. A given stock could occupy different longitudinal sectors in different periods within a feeding season or different stocks could occupy different longitudinal ranges in different years. These possibilities should be investigated in future with the incorporation of the temporal factor in addition to the longitudinal sub-division used in this study.

With a larger number of biopsy samples available from the Antarctic feeding ground it will be necessary to conduct a new analysis in combination with samples from low latitudes, similar to that conducted preliminary by Pastene and Baker (1997). In this context we are planing a co-operative study that would involve the combined analysis of samples from the feeding ground (used in this study) and at least one locality from lower latitudes (Eastern Australia), from where a large genetic data set is already available.

The use of microsatellite for individual identification

The analysis with a set of seven microsatellite loci proved to be a useful tool for individual identification and studies on 'mark-recapture' in the Antarctic. Previously the usefulness of microsatellite in such kind of studies was demonstrated in the North Atlantic humpback whale (Palsboll *et al.*, 1997b). In the Antarctic we 'recaptured' a female individual five years after it was marked originally. The recapture was determined by a combination of molecular techniques, matching genotypes, matching sex and identical mtDNA control region sequences. The probability of identity (the probability that two different individuals shared the same genotype) is low in the Antarctic, particularly in Areas IV and V.

We checked our photo-id catalogue for possible matching. We found some pictures but unfortunately those from the 1995 school are ventral flukes only while those from the 2000 school are lateral markings only.

ACKNOWLEDGMENTS

Skin biopsy samples of humpback whales were obtained during the JARPA surveys from 1993/94 to 1999/2000 in Areas IIIE, IV, V and VIW. We would like to acknowledge the researchers and crewmembers that participated in these surveys.

REFERENCES

Arnason, U., Gullberg, A. and Widegren, B. 1993. Cetacean mitochondrial DNA control region: sequences of all extant baleen whales and two sperm whale species. *Mol. Biol. Evol.* 10:960-70.

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

Baker, C.S., Florez-Gonzalez, L., Abernethy, B. Rosenbaum, H.C., Slade, R.W., Capella, J. and Bannister, J.L. 1998. Mitochondrial DNA variation and meternal gene flow among humpback whales of the Southern Hemisphere. *Marine Mammal Science* 14 (4): 721-737.

Berube, M., Aguilar, A., Dendanto, D., Larsen, F., Notarbartolo Di Sciara, G., Sears, R., Sigurjonsson, J., Urban, J. and Palsboll, P.J. 1998. Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus 1758): analysis of mitochondrial and nuclear loci. *Mol. Ecol.* 7:585-599.

Chittleborough, R.G. 1959. Australian marking of humpback whales. *Norsk Hvalfangsttid* 48: 47-55.

Dawbin, W.H. 1966. The seasonal migratory cycle of humpback whales. *In* Norris, K.S. (ed.). *Whales, dolphins and porpoises*. University of California Press, Berkerly. pp. 145-171.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 44: 1097-1108.

Hoelzel, A.R. (ed.). 1992. *Molecular Genetic Analysis of Populations. A practical approach*. Oxford University Press. Oxford, New York, Tokyo. 315pp.

Hori, H., Bessho, Y., Kawabata, R., Watanabe, I., Koga, A. and Pastene, L. 1994. World-wide population structure of minke whales deduced from mitochondrial DNA control region sequences. Paper SC/46/SH14 presented to the IWC Scientific Committee, May 1994 (unpublished). 11pp.

Hudson, R.R., Boos, D.D. and Kaplan, N.L. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* 9(1):138-151.

Kasamatsu, F., Iwata, S. and Nishiwaki, S. 1991. Development of biopsy skin sampling system for fast swimming whales in pelagic waters. *Rep. int. Whal. Commn* 41:555-557.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.

Mackintosh, N.A. 1965. *The stocks of whales*. Fishing News (Books) Ltd., London. 232pp.

Nei, M. 1987. *Molecular Evolutionary Genetics*. Columbia University Press, New York. x+512pp.

Nei, M and Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci.* USA 76:5269-5273.

Nei, M. and Tajima, F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics* 97:145-163.

Omura, H. 1953. Biological study on humpback whales in the Antarctic whaling areas IV and V. *Scientific Reports of the Whales Research Institute*, Tokyo 8:81-102.

Paetkau, D. and Strobeck, C. 1994. Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* 3:489-95.

Palsboll, P.J., Clapham, P.J., Mattila, D.K., Larsen, F., Sears, R., Siegismund, H.R., Sigurjonsson, J., Vasquez, O. and Arctander, P. 1995. Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behavior on population structure. *Marine Ecology Progress Series* 116:1-10.

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

Palsboll, P.J., Allen, J., Berube, M., Clapham, P.J., Feddersen, T.P., Hammond, P.S., Hudson, R.R., Jorgensen, H., Katona, S., Larsen, A.H., Larsen, F., Lien, J., Mattila, D.K., Sigurjonsson, J., Sears, R., Smith, T., Sponer, R., Stevick, P., and Oien, N. 1997b. Genetic tagging of humpback whales. *Nature* 388:767-769.

Pastene, L.A. and Baker, C.S. 1997. Diversity and distribution of mtDNA lineages from humpback whales on the feeding and wintering grounds of the Southern Hemisphere. Paper SC/49/SH26 presented to the IWC Scientific Committee, October 1997 (unpublished). 19pp.

Pastene, L.A., M. Goto, T. Kimura and S. Nishiwaki. 1997. Population structure of the humpback whale in the Antarctic feeding ground based on analysis of mitochondrial DNA control region sequences. Paper SC/49/SH21 presented to the IWC Scientific Committee, October 1997 (unpublished). 12pp.

Raymond, M. and Rousset, F. 1995. GENEPOP (version 1.2). population genetics software for exact tests and ecumenicism . *J. Heredity* 86:248-9

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic tree. *Mol. Biol. Evol.* 4: 406-425.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. *Molecular Cloning: A Laboratory Manual*. Second Edition. Cold Spring Harbor Laboratory, New York.

Sokal, R.R. and Rohlf, F.J. 1995. *Biometry: the Principles and Practice of Statistics in Biological Research*. W.H. Freeman and Company, New York. I-xix+887 pp.

Valsecchi, E., Palsboll, P., Hale, P., Glockner-Ferrari, D., Ferrari, M., Clapham, P., Larsen, F., Mattila, D., Sears, R., Sigurjonsson, J., Brown, M., Corkeron, P. and Amos, B. 1997. Microsatellite genetic distances between oceanic populations of the humpback whale (Megaptera novaeangliae). *Mol. Biol. Evol.* 14 (4): 355-362.

Table 1: Number of samples of humpback whales examined in this study, by austral summer season, Antarctic Area and sex. %M refers to the male proportion in each Area sample

YEAR	AREA IIIE				AREA IV				AREA V				AREA VIW			
	M	F	T	%M	M	F	T	%M	M	F	T	%M	M	F	T	%M
1993/94					8	10	18	0.44								
1994/95									3	6	9	0.33				
1995/96		2	2	0.00	3	5	8	0.38								
1996/97									1	4	5	0.20	11	4	15	0.73
1997/98	2	3	5	0.40	12	7	19	0.63								
1998/99									11	15	26	0.42	1		1	1.00
1999/00	5	3	8	0.63	18	10	28	0.64								
TOTAL	7	8	15	**0.47**	41	32	73	**0.56**	15	25	40	**0.38**	12	4	16	**0.75**

Table 2: Relationship between gender and school size in the Antarctic humpback whale (figures indicates number of cases)

School size	Area IIIE		Area IV		Area V		Area VIW		Total	
	M	F	M	F	M	F	M	F	M	F
1			5	3	2	4	1		8	7
2	7	4	24	23	6	11	3	3	40	41
=or>3		4	12	6	7	10	8	1	27	21

Table 3: Variable sites defining 60 mtDNA haplotypes in the Antarctic humpback whale. The column on the left are haplotype ID. The numbers above list the nucleotide position of the polymorphic sites starting from the 5' end of the mtDNA control region. Haplotypes '2' through '60' are listed with reference to haplotype '1'. A dot indicate an identical nucleotide at the position relative to haplotype '1'. A hyphen indicate a deletion. On the right side of the table are the frequencies of the 60 haplotypes in the five Antarctic Areas

	10	20	30	40	50						
		11111111	1111111111	1222222222	2222222222	22222333					
	255667788	9901111122	2334566678	9333444455	5556666666	67888011					
	7346171423	6805678914	6234905911	4678345612	5790123456	71234445	I	IIIE	IV	V	VIW
1	TGGTTCCTTT	TCAGATA-CC	TGGTTTCACT	TAATGTCCAA	TCAGGTCCGT	CATCTAAC	1	0	0	0	0
2	C.....C	.....T	.A...C....	..G.....	.....T.A.	TG....CT	1	0	0	1	1
3	....C.....	.....	.A.C..T...	..G.....	.....TTA.	T.C...CT	2	0	1	0	1
4	C	.A.....T	.AAC.....	..G..C....	.....T.A.	TG....CT	2	0	9	1	0
5	.A.....C	.A.....T	.A.C.C....	..G..C....	.....T.A.	TG....CT	2	0	0	0	0
6	C	.A.....T	.AAC.....	..G..C....	.....TTA.	TG....CT	1	0	0	0	0
7	.....	.....T	.AAC.....	..G..C....	.....TA.	TG..C.CT	2	1	0	0	1
8	.....C.	.A.....T	.A.C.C....	..G..C....	.....T.A.	TG....CT	0	1	0	0	0
9	....C.....	.....	.A.C..T...	..G.....	.....TA.	T.C...CT	0	1	0	0	0
10	....C.....	.....	.....T...	..G.....	....A..TA.	T.....CT	0	1	0	0	0
11	.....C..	.A.....T	.A.C.C....	..G..C....	.....TTA.	TG....CT	0	2	0	1	0
12	....C...C.	.....T	.....T...	..G.....	.....TTA.	T.....CT	0	1	7	0	0
13	....C...C.	.....T	.....T...	..G.....G	.....TTA.	T.....CT	0	1	1	0	0
14	C	.A.....T	.AAC.....	..G.AC....	.....T.A.	TG....CT	0	1	1	0	0
15	C	.AG.....T	.A.C.C....	..G..C....	.....T.A.	TG....CT	0	1	0	0	0
16	.....	.A.....T	.A.....	..G..C....	.....TTA.	TG....CT	0	1	0	0	0
17	....C.T...	.....	.A.C..T...	..G.....	.....TTA.	T.C...CT	0	1	0	0	0
18	....C.....	.....	.....T...	.GG.....	.....TA.	T.....CT	0	1	0	0	0
19	C	.A.....T	.AAC.....	..G.AC....	....C.T.A.	TG....CT	0	1	1	4	0
20	.....	.A.....T	.A.C.C....	..G.....	C.....T.A.	TG....CT	0	1	0	0	0
21	.....	.....T	.A...C....	..G.....	.....T.A.	TG....CT	0	0	1	0	1
22	.....	.....T	.A.....C	..G..C....	.....CT.A.	.G....CT	0	0	1	0	0
23	C	.A.A.....T	.AAC.....	..G..C....	.....TTA.	T.....TT	0	0	3	0	0
24	....C...C.	.....T	.....T...	..G...T...	.....TTA.	T.....CT	0	0	1	0	0
25	.....	.....T	.A...C....	..G.....	.....T.A.	.G....CT	0	0	3	2	0
26	.....	.....T	.A...C...C	..G..C....	.....CT.A.	.G....CT	0	0	5	0	0
27	C	.A.....T	.A.C.C....	..G.....	.....TA.	TG....TT	0	0	1	0	0
28	....C.....	C.....T..	.A.C..T...	..G..C....	.....TTA.	T.C...CT	0	0	6	1	1
29	....C.....	C.....T..	.A.C..T...	..G.....	.....T.A.	T.....CT	0	0	1	2	0
30	.....C.	.A.....T	.AAC.C....	..G..C....	.....TTA.	TG...GCT	0	0	4	4	0
31	....C.....	.....T	.....T...	..G...T...	.....TTA.	T.....CT	0	0	4	0	0
32	.....	.....T	.A...C....	..G..C....	C....C..A.	.G....CT	0	0	3	5	3
33	....C.....	C.....T..	.A.C..T...	..G..C....	.....TTA.	T.....CT	0	0	1	0	0
34	C	.....T	.A...C....	..G.....	.....T.A.	TG....CT	0	0	2	0	0
35	....C.....	.....	.A.CC.T...	..G..C.T..	..G...TTA.	T.CT..CT	0	0	2	0	0
36	.....C.	.A..TATA.T	.A.C.C....	..G..C....	.....T.A.	TG....CT	0	0	1	0	0
37	C	.A.....T	.AAC.....	..G..C....	C.....T.A.	TG....CT	0	0	1	0	0
38	.....C.	.A.....T	.A.C.C....	..G..C....	C.....T.A.	TG....CT	0	0	1	0	0
39	....C.....	C.....T..	CA.C..T...	..G.....	.....TTA.	T.C...CT	0	0	1	0	0
40	.A.....	.....T	.A.C.....	C.GC..T...	....A...A.	TG....CT	0	0	1	0	0
41	....C.....	.....T	.....T...	..G.....	.....TTA.	T.....CT	0	0	1	0	1
42	....C.....	C.....T..	.A.C..T...	..G.....	.....TTA.	T.C..GCT	0	0	2	0	0
43	...C...C..	.A.....T	.A.C.C....	..G..C....	.....TTA.	TG....CT	0	0	1	1	0
44	C	.A.....T	.A.C.C....	..G..C....	.....T.A.	TG....CT	0	0	2	0	2
45	....C.....	.....	.....T...	..G.....	....A.TTA.	T.....CT	0	0	1	0	0
46	....C.....	.....	.....T...	..G.....	.T.....TAC	T.....CT	0	0	1	0	0
47	....C...C.	C.....T..	.A.C..T...	..G..C....	.....TTA.	T.C.C.CT	0	0	1	0	0
48	.....	.....T	.AAC.....	..G..C....	.....TA.	TG..CGCT	0	0	1	1	2
49	.A..C.....	.....	.A.C..T...	..G.....	.....TTA.	T.C...CT	0	0	0	1	0
50	..A..T....	.T.....	.A.....	..G.....G.	.....TAC	T.....CT	0	0	0	1	0
51	..A..T....	.....	.A.....	..G.....G.	.....TAC	T.....CT	0	0	0	3	1
52	C	.A.....T	.AAC.C....	..G..C....	C.....T.A.	TG....CT	0	0	0	1	0
53	.....	.A.....T	.A.C.C....	..G..C....	C.....T.A.	TG....CT	0	0	0	1	0
54	C	.T.....T	.AAC.....	..G..C....	.....T.A.	TG....CT	0	0	0	3	1
55	....C.....	C.....TT.	.A.C..T...	..G...T...	.....TTA.	T.C...CT	0	0	0	2	0
56	....C...C.	.....	.....TGT.	..G.....	....A.TTA.	T.....CT	0	0	0	1	0
57	C	.A.A.....T	.AACC.....	..G..C....	.....TTA.	T.....CT	0	0	0	1	0
58	.....	.....T	.A.C.C...C	..G.....	...A..T.A.	.G....CT	0	0	0	1	1
59	....C.....	C.....T..	.....T...	..G.....	.....TTA.	T.C...CT	0	0	0	1	0
60	....C.....	C.....T..	.A.C..T...	..G.....	.....TTA.	T.C...CT	0	0	0	1	0
	2222222222	2322222222	2222222222	2222222222	2222322222	22222232	11	15	73	40	16

Table 4: Estimates of the nucleon and nucleotide diversities in the Antarctic humpback whale

Area (sample size)	Nucleon diversity	Nucleotide diversity (SE)
Area I (11)	0.9273	0.0230 (0.0039)
Area IIIE (15)	0.9905	0.0244 (0.0018)
Area IV (73)	0.9593	0.0256 (0.0008)
Area V (40)	0.9603	0.0281 (0.0014)
Area VIW (16)	0.9583	0.0243 (0.0020)
Total (155)	0.9745	0.0260 (0.0006)

Table 5: Results of the homogeneity test using randomized chi-square test in pair-wise comparisons. The overall P value for the chi-square test was 0.0000. P values below 5% level are indicated in bold.

Areas compared	Chi-square (P-value)
I and IIIE	0.1179
I and IV	**0.0032**
I and V	**0.0011**
I and VIW	0.1581
IIIE and IV	**0.0004**
IIIE and V	**0.0027**
IIIE and VIW	**0.0411**
IV and V	**0.0000**
IV and VIW	**0.0383**
V and VIW	0.3913

Table 6: Polymorphism at every locus for all Antarctic Areas

		Microsatellite loci							
		GATA 417	GATA 28	GATA 98	TAA 31	GATA 53	GGAA 520	GT23	All loci
Area I	Allele	9	5	7	8	6	8	5	
	H	0.85	0.44	0.80	0.85	0.74	0.82	0.69	0.74
	I	0.039	0.333	0.065	0.038	0.111	0.052	0.144	2.694×10^{-8}
Area IIIE	Allele	11	7	7	8	8	11	7	
	H	0.87	0.45	0.67	0.84	0.82	0.89	0.82	0.77
	I	0.028	0.318	0.138	0.047	0.054	0.023	0.058	4.134×10^{-9}
Area IV	Allele	12	11	13	19	8	19	9	
	H	0.91	0.54	0.79	0.89	0.82	0.91	0.78	0.81
	I	0.017	0.225	0.064	0.024	0.056	0.014	0.080	3.718×10^{-10}
Area V	Allele	13	10	9	13	10	13	9	
	H	0.87	0.65	0.76	0.90	0.81	0.89	0.76	0.81
	I	0.025	0.045	0.012	0.011	0.011	0.026	0.004	1.776×10^{-13}
Area VIW	Allele	11	7	6	9	8	13	7	
	H	0.87	0.38	0.73	0.78	0.82	0.88	0.79	0.75
	I	0.029	0.392	0.107	0.072	0.057	0.027	0.072	9.333×10^{-9}

Table 7: Results of the chi-square tests of HW genotypic proportions. Figures shown are probabilities. Figures in bold indicates values below 5% level

Loci	Area I	Area IIIE	Area IV	Area V	Area VIW	All Areas
GATA417	0.8416	0.7597	0.5169	**0.0001**	0.6702	0.1565
GATA28	0.7557	0.9984	0.9060	0.8337	0.8017	0.9955
GATA98	0.8425	**0.0281**	**0.0001**	0.5630	0.8568	**0.0001**
TAA31	0.8523	0.2294	0.7925	0.0344	**0.0078**	**0.0003**
GATA53	0.5255	**0.0001**	0.7574	0.6927	0.9395	0.4583
GGAA520	0.8599	0.2995	**0.0001**	**0.0001**	0.3553	**0.0001**
GT23	**0.0012**	0.2384	0.9426	0.4632	0.7469	**0.0201**
All loci	0.2777	**0.0018**	**0.0004**	**0.0001**	0.4469	**0.0001**

Table 8: Results of the homogeneity tests using microsatellite data. Figures shown are probabilities. P-values below 5% are shown in bold; those near to significant are underlined

Comparisons	GATA 417	GATA 28	GATA 98	TAA31	GATA 53	GGAA 520	GT23	All loci
I and IIIE	0.1570	0.8723	0.0790	0.3772	0.3100	0.6648	0.1727	0.2221
I and IV	**0.0046**	0.4150	0.2505	0.3352	0.4061	0.3302	0.5259	0.0643
I and V	**0.0062**	0.3818	0.4738	0.3176	0.5480	0.1329	0.0580	**0.0203**
I and VIW	**0.0251**	0.9392	0.3620	0.2352	0.2424	0.0681	0.3490	0.0648
IIIE and IV	0.6424	0.9105	0.1707	0.7832	0.2439	0.6664	0.2787	0.6633
IIIE and V	**0.0028**	0.6019	0.4057	0.1197	0.9063	0.3482	0.4623	0.0658
IIIE and VIW	0.2815	0.9596	**0.0330**	0.7047	0.3817	0.4396	0.8579	0.4486
IV and V	**0.0040**	0.3200	0.7884	0.1412	0.1592	0.9560	0.0956	**0.0244**
IV and VIW	0.3811	0.9304	0.2803	0.4544	0.1210	0.3614	0.6950	0.5123
V and VIW	0.6640	0.6330	0.2984	0.0680	0.1310	0.5526	0.8769	0.3752

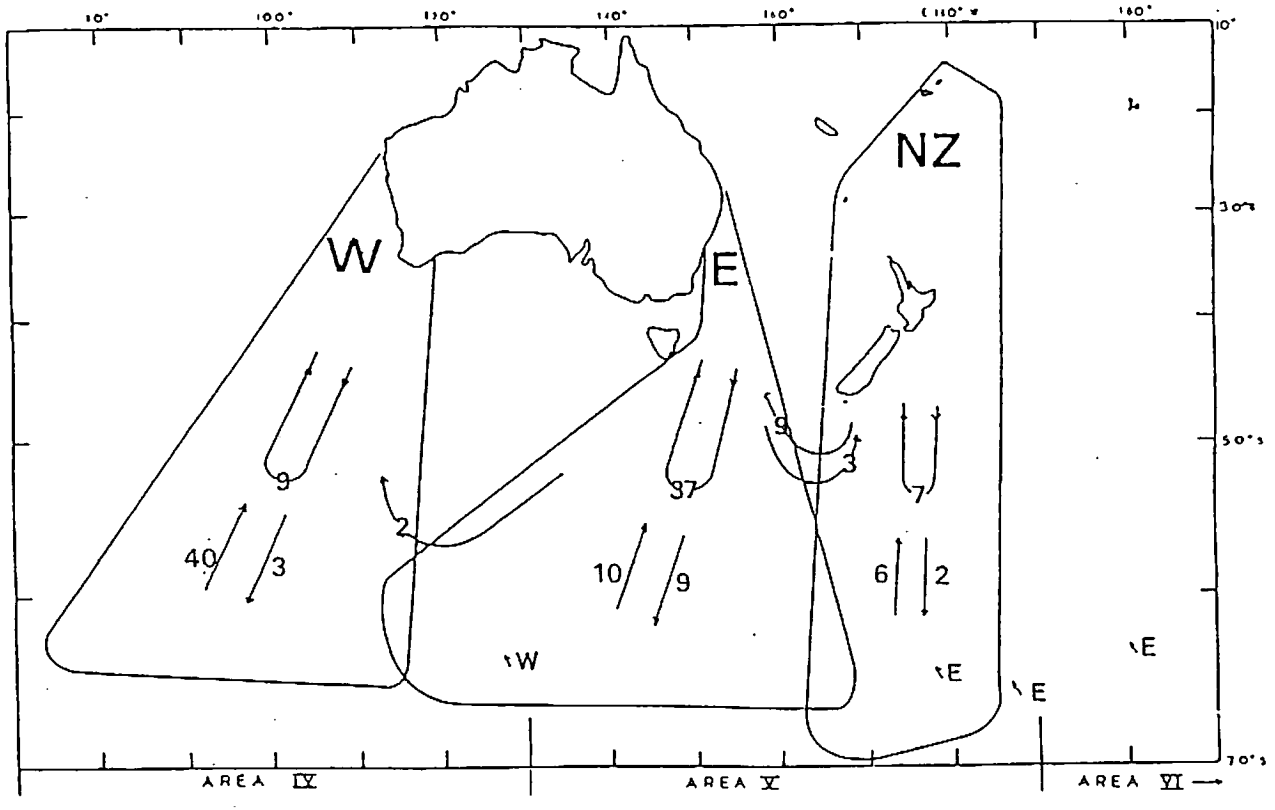

Fig. 1: Reproduction of Fig. 2 of Dawbin (1966) showing the movement of humpback whales between low-latitude localities and the Antarctic Areas IV and V, and between low-latitude localities. Information is based in the results of the analysis of mark-recapture data between 70°E and 150°W. W= Western Australian group, E= Eastern Australian group, NZ= New Zealand group. The geographical range of the three groups were defined by the positions of whales marked in the Antarctic and recovered in low-latitude localities (arrows pointing north), the position of whales marked in low-latitude localities and recovered in the Antarctic (arrows pointing south), the position of whales marked and recovered in the same low-latitude locality, in one or more season later (looped arrows within each locality). Interchange between low-latitude localities is shown by curved arrows crossing boundaries between W, E and NZ (see more details in Dawbin, 1966). Note that the boundaries in the Antarctic do not correspond with the actual boundaries of Areas IV and V and that there is an overlap at the boundaries between groups.

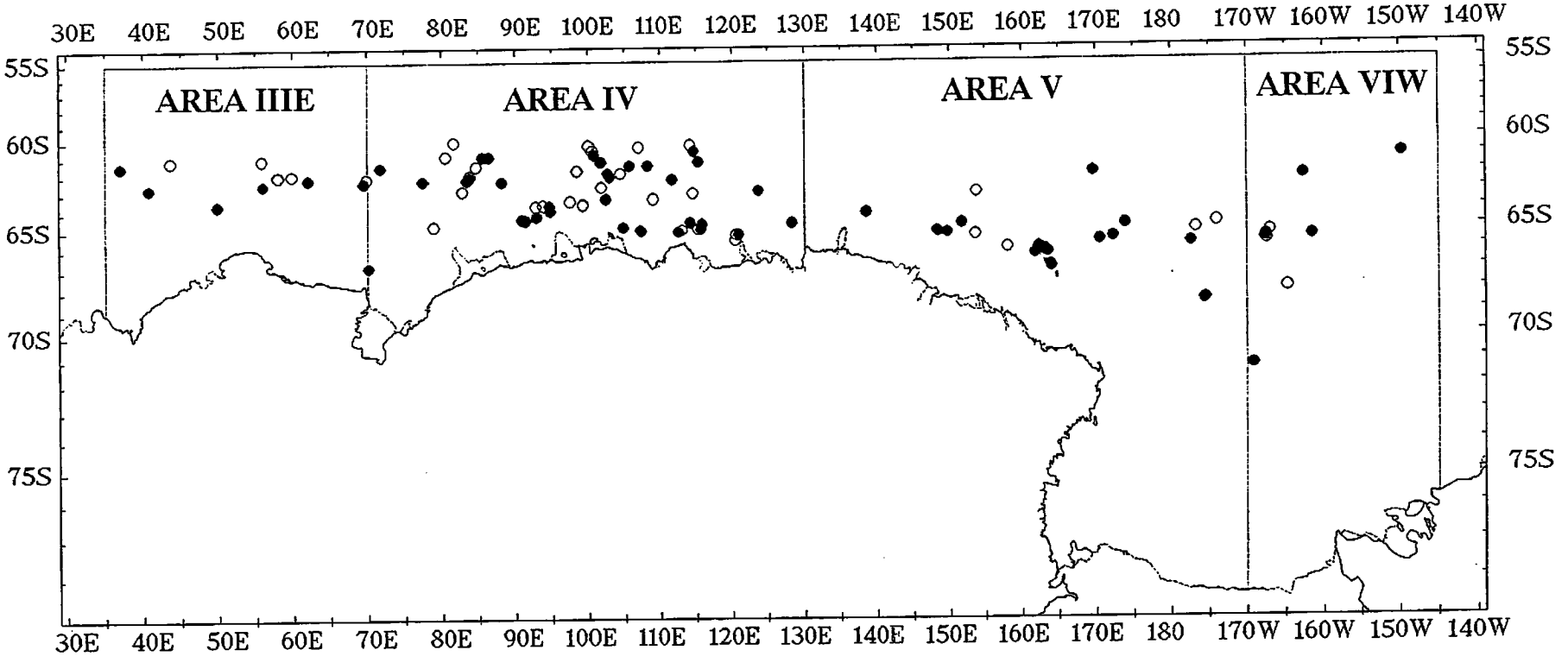

Fig. 2: Geographic distribution of the humpback whales examined in this genetic study, by Antarctic Area and sex. Closed circle = male; open circle = female

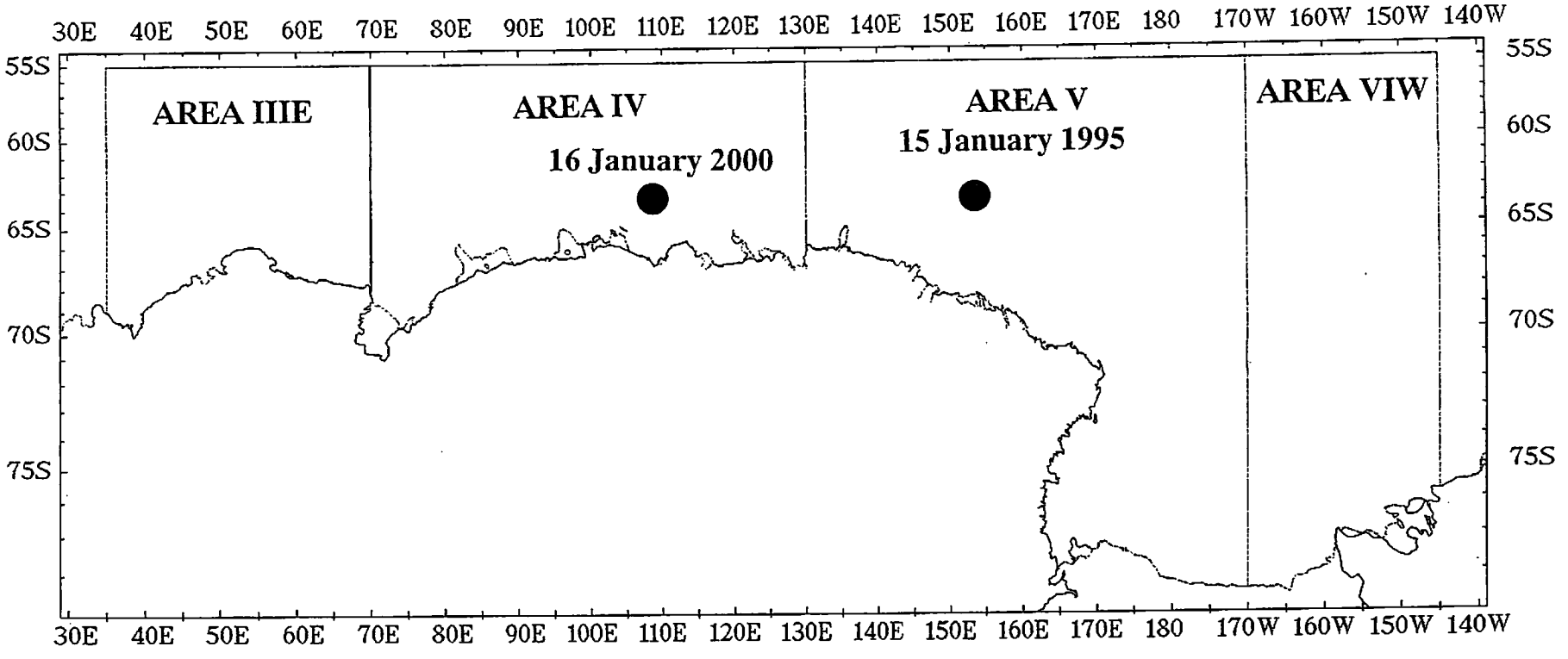

Fig. 3: Geographical position of a female humpback whale genetically marked in the western part of Area V in 1995 and 'recaptured' in the eastern part of Area IV in 2000. In 1995 it was observed in a school of three individuals and the body size was estimated at 12.8m. In 2000 it was observed in a school of two individuals and the body size was estimated at 14.8m

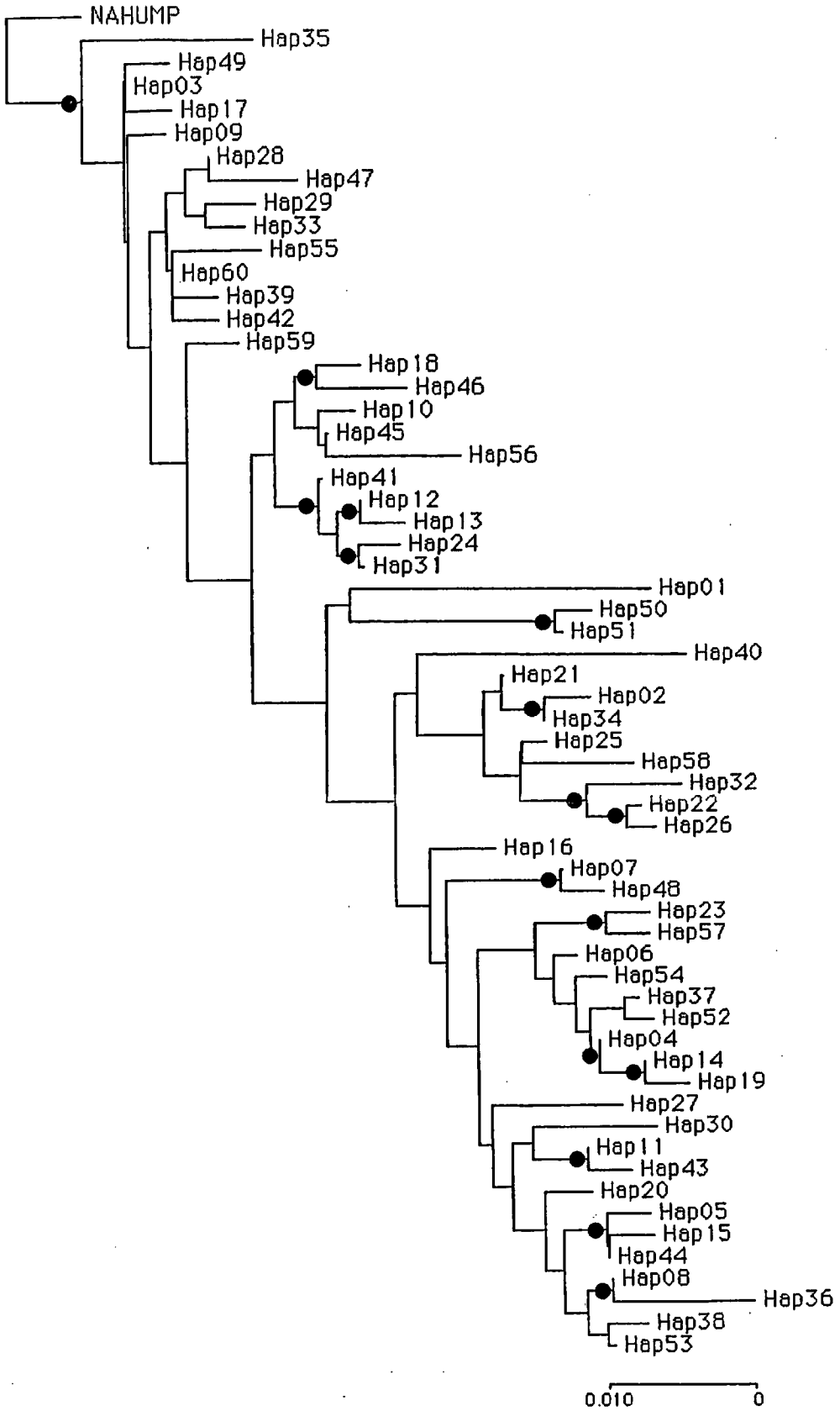

Fig. 4: Neigbor-Joining-based tree showing the phylogenetic relationships among 60 mtDNA haplotypes in the Antarctic humpback whale. A sequence of the North Atlantic humpback whale is used as out-group (Arnason *et al.*, 1993). Nodes with bootstrap values above 50% (in 400 simulations), are indicated by closed circles. See Table 3 for the frequencies of haplotypes by Area.