

# Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm low abundance and reproductive autonomy of humpback whales on the wintering grounds of New Caledonia

Claire Garrigue<sup>1,2</sup>, Rémi Dodemont<sup>1</sup>, Debbie Steel<sup>2</sup>, C. Scott Baker<sup>2,\*</sup>

<sup>1</sup>Opération Cétacés, BP 12827, 98802 Nouméa, New Caledonia

<sup>2</sup>School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

**ABSTRACT:** The abundance and reproductive autonomy of humpback whales from the New Caledonia (South Pacific) wintering grounds were investigated using capture-recapture models and paternity inference based on nuclear microsatellite genotyping, mitochondrial DNA sequencing, molecular sex identification and photographs of natural markings (photo-identification). The analyses included records of 213 individuals (excluding 16 calves used in paternity inference) identified by genotypes (9 loci), and 210 identified by fluke photographs, collected from 1995 to 2001. By comparing records of 175 individuals identified using both genotypes and photographs, a small number of errors were detected (and corrected) in the photo-identification catalogue and in the field and laboratory notes of the genotypes. Using the weighted mean of the Petersen capture-recapture model, the estimate of abundance based on genotyping ( $N = 533$ ,  $CV = 0.15$ ) was larger than the estimate based on photo-identification ( $N = 327$ ,  $CV = 0.11$ ). Sex-specific estimates of abundance based on genotypes were similar for males and females although the variance of the female estimate was greater than the male ( $N_f = 248$ ,  $CV = 0.30$ ;  $N_m = 288$ ,  $CV = 0.18$ , respectively). The paternity of 5 calves from 16 sampled cow/calf pairs was inferred from the total sample of 133 non-calf males (one offspring each). The 16 sampled cow/calf pairs and the 5 inferred paternities were used for an alternate 'gametic recapture' estimate of male abundance. This gametic recapture estimate ( $N_m = 379$ ,  $CV = 0.30$ ) was similar to the sex-specific estimate based on the organismal recapture using a 2-sample model ( $N_m = 382$ ,  $CV = 0.22$ ). The close agreement of the organismal and gametic recapture estimates support the assumption that this humpback whale wintering ground represents an autonomous population unit that is relatively closed to demographic and reproductive interchange. The current low abundance of the New Caledonian breeding unit highlights the slow recovery of this and some other regional populations in the Southern Hemisphere following intensive 20th century commercial and illegal whaling.

**KEY WORDS:** Humpback whales · *Megaptera novaeangliae* · Abundance · Microsatellite genotyping · Reproductive autonomy · Photo-identification · Paternity · Population recovery

Resale or republication not permitted without written consent of the publisher

## INTRODUCTION

The waters of Oceania (South Pacific) include a number of known and suspected winter breeding grounds for humpback whales that feed in Areas V, VI and I of the Antarctic (i.e. Group V, VI and I stocks, see Fig. 1; Donovan 1991). Whaling during the 20th century

reduced these populations to very low numbers. In total, more than 200 000 humpback whales were killed in the Southern Hemisphere (Clapham & Baker 2002). The illegal Soviet whaling program, operating from 1947 to 1973, was particularly damaging. During this period, the USSR reported to the International Whaling Commission (IWC) on catches of only 2710 humpback

\*Corresponding author. Email: cs.baker@auckland.ac.nz

whales when, in fact, they had taken more than 48 000 (Zemsky et al. 1995). Of these, more than 13 000 humpbacks were killed in Areas V and VI south of Oceania during the single summer season of 1961/62 (Mikhalev 2000). Since official protection from commercial hunting was offered in 1964 (with violations by the Soviet whaling fleet until 1973), recovery of populations in regions around the Southern Hemisphere has been highly variable (Anonymous 2000).

Over the last several years, a number of projects have used individual identification photographs (photo-identification) collected during non-systematic vessel surveys to assess the status of humpback whales in various parts of Oceania, including New Caledonia, Fiji, Tonga, New Zealand, the Cook Islands and French Polynesia (Garrigue et al. 2000, 2002). Comparisons of photo-identification catalogues have revealed some interchange between neighboring wintering grounds (e.g. New Caledonia to Tonga), but this level of interchange is low when compared to interannual return, suggesting fidelity to individual wintering grounds. A comparison of mitochondrial (mt) DNA differentiation has confirmed a degree of genetic isolation of Oceania from wintering grounds near Colombian (presumed Group I) and Western Australia (presumed Group IV), but sample sizes were not sufficient to evaluate isolation among regions within Oceania (Baker et al. 1998). However, differences in reported rates of recovery in eastern Australia and Oceania suggest demographic independence of these regions. Preliminary estimates of abundance based on capture-recapture analyses of photo-identification records indicate that abundance remains low in New Caledonia (approximately 314 individuals in the year 2000, Garrigue et al. 2001) and Tonga (approximately 770 individuals in the year 2000, Baker et al. 2001). In other regions of the South Pacific (e.g. Cook Islands, New Zealand, French Polynesia) investigators note that the density of humpback whales seems to be low (Garrigue et al. 2002). This is in contrast with the increase in numbers exhibited along the eastern coast of Australia (Group V) and Western Australia (Group IV) (e.g. Bryden et al. 1990, Bannister 1994, Paterson et al. 1994, Anonymous 2000).

Although photo-identification has proven effective for estimating population abundance and describing migratory movement of humpback whales and other cetaceans (e.g. Hammond et al. 1990), there are known limitations of this method. These include the potential for changes in natural markings (equivalent to tag loss), the lack of distinctive markings for some individuals and sex-specific patterns of behavior that result in heterogeneity of capture probability. The development of genetic markers for sex determination, individual identification (e.g. DNA profiling or genotyping) and

parentage assignment has provided information that can extend, and in some cases replace, photo-identification or tagging (e.g. Amos et al. 1993, Richards et al. 1996, Palsboll et al. 1997a, Taberlet et al. 1997, Kohn et al. 1999, Pearse et al. 2001). For example, a 2 yr capture-recapture analysis of humpback whales from the West Indies based on microsatellite genotyping from biopsy samples provided a markedly lower estimate of abundance for females than for males (Palsboll et al. 1997a). The lower estimate of female abundance was interpreted as the result of a reduced period of residency or incomplete migration by females on this winter breeding ground in one or both of the 2 yr sampling periods.

Despite the potential for improvement of capture-recapture estimates through microsatellite genotyping and sex identification, a central question often remains unanswered: what population or population unit is being estimated? The genetic structure of humpback whale populations is complex (Baker & Palumbi 1995, Baker & Medrano 2002), reflecting long-distance migration from summer feeding grounds to winter breeding grounds (Clapham 1996). In the North Pacific and North Atlantic, humpbacks show considerable maternal fidelity to summer destinations, resulting in a segregation of mtDNA haplotypes on regional feeding grounds (Baker et al. 1990, 1993, Palsboll et al. 1995, Larsen et al. 1996). Abundance estimates from these feeding grounds are assumed to reflect relatively discrete population units defined by maternal fidelity. However, individuals from different regional feeding grounds are not necessarily isolated from each other reproductively. Instead, humpbacks from different feeding grounds congregate to breed during winter months in shallow coastal or insular waters of tropical latitudes (Mackintosh 1965). Although these geographically isolated wintering grounds are assumed to represent breeding units or 'stocks', the evidence that mating occurs in these regions is indirect, and the degree of reproductive isolation has not been established.

Here we present capture-recapture estimates of abundance for humpback whales on the wintering grounds of New Caledonia using photo-identification, microsatellite genotyping and molecular identification of sex. We also used the genotypes of individual males in the population to identify the likely paternity of calves. The frequency of these assignments formed the basis of a novel 'gametic recapture' analysis as an alternative to direct estimates of male abundance. This approach differed from methods that focus on the number of breeding males in a population (e.g. Nielsen et al. 2001, Pearse et al. 2001). Instead, our purpose was to evaluate the reproductive autonomy of this wintering ground by comparing the gametic recapture estimate to the organismal recapture estimate, i.e. are

the proportion of fathers captured by the paternity analysis consistent with the estimated size of the local male population?

## MATERIALS AND METHODS

### Field collection and photographic identification.

Humpback whales were photographed for individual identification during small-boat surveys around New Caledonia (Fig. 1) in the winter breeding and calving seasons (July to September) from 1995 to 2001 (see Garrigue et al. 2001). Individual humpback whales were identified from photographs of the unique markings on the ventral surface of their tail flukes (photo-identification, Katona et al. 1979), as well as using microsatellite genotypes (see below). For the purposes of the present study, some sighting records reported previously were deleted from the catalogue (e.g. Garrigue et al. 2001). These include photographs taken prior to initiation of biopsy samples in 1995, individuals identified only by dorsal fins or lateral markings, and calves. Fluke photographs were classified as good, fair or poor according to photographic quality (e.g. focus and exposure, Mizroch et al. 1990, Perry et al. 1990). To assure accuracy of individual identification and avoid biases in estimates of abundance (Perry et al. 1990, Friday et al. 2000), all photographs of poor quality were deleted from the capture-recapture analyses presented here. To investigate the possibility that flukes with predominantly white coloration were difficult to individually identify (Anonymous 1998), flukes were ranked from all white (Type I) to all black (Type V) following Carlson & Mayo (1990).

**Molecular analysis.** Skin tissue for genetic analysis was collected when possible from photo-identified whales (Garrigue & Greaves 1999). Most tissue samples ( $n = 285$ ) were collected using a stainless-steel biopsy dart deployed from a crossbow (Lambertsen 1987). The remaining samples ( $n = 36$ ) were collected from skin sloughed naturally from individuals during surface-active behavior (Clapham et al. 1993). The samples were preserved in 70% ethanol and stored at  $-20^{\circ}\text{C}$ . Total cellular DNA was isolated from tissue by digestion with ProK, followed by phenol/chloroform extraction and ethanol precipitation, according to Sambrook et al. (1989), as modified for small samples (Baker et al. 1994).

An approximately 800 base-pair (bp) fragment of the 5'-end of the mitochondrial (mt) DNA control region (i.e. D-loop) was amplified via the polymerase chain reaction (PCR) using the primers light-strand M13Dlp1.5 (tPro whale, 5'-TGTA AACGACAGCCA-GTTCACCCAAAGCTGRARTTCTA-3') and heavy-strand Dlp8G (5'-GGAGTACTATGTCTCTGTAACCA-3'). Amplifications were conducted in a final volume of 50  $\mu\text{l}$  at the following concentrations: 2.5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  dNTP, 0.4 mM each primer, 0.7 U *AmpliTaq* (Applied Biosystems), 1  $\times$  PCRII reaction buffer (Applied Biosystems) and 1  $\mu\text{l}$  DNA (approximately 50 ng). The thermocycle profile consisted of initial denaturation at  $94^{\circ}\text{C}$  for 2 min, followed by 30 cycles of denaturation at  $94^{\circ}\text{C}$  for 30 s, annealing at  $56^{\circ}\text{C}$  for 40 s, and extension at  $72^{\circ}\text{C}$  for 40 s. The amplified fragment was sequenced on an ABI 377 (Applied Biosystems) using BigDye™ Dye Terminator Chemistry (ver. 3 Applied Biosystems).

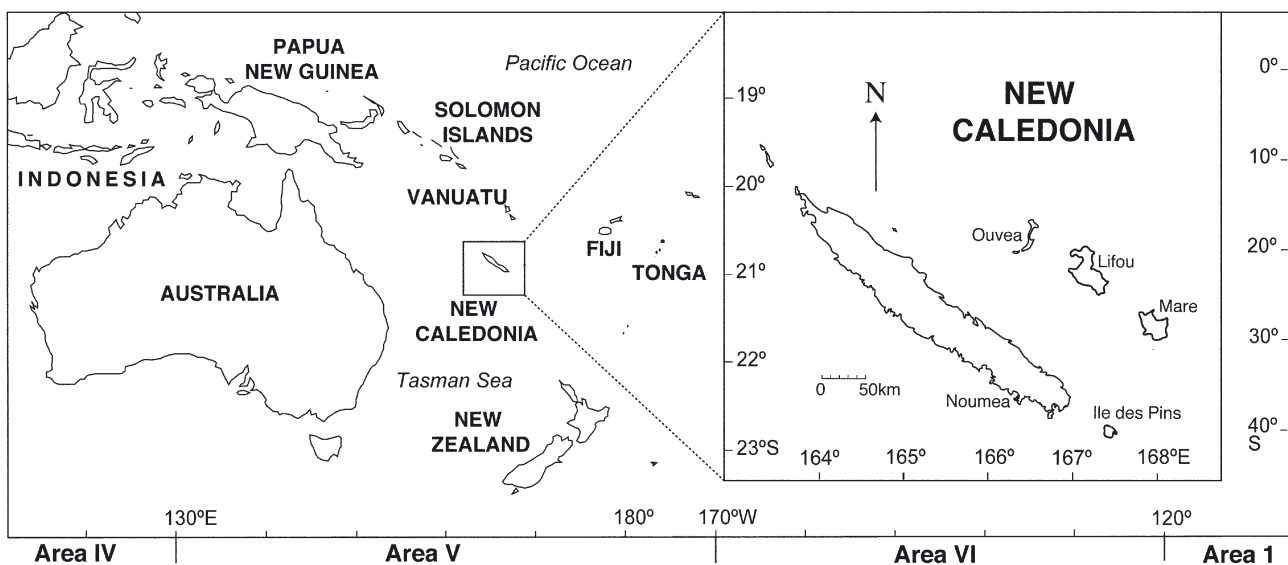


Fig. 1. Oceania and New Caledonia, with the location of primary study site and boundaries of Antarctic feeding divisions (Areas VI, V and VI) used by the International Whaling Commission for management purposes (Donovan 1991)

The sex of each sample was identified by the amplification of a male-specific SRY marker, with a ZFX positive control, following the methods of Gilson et al. (1998). Reaction volumes and reagents were as described for mtDNA. The thermocycle profile consisted of initial denaturation at 94°C for 3 min, 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 45 s, and extension at 72°C for 40 s.

Eleven published microsatellite loci were screened for genotyping and standardized with collaborators (Anderson et al. 2001, Medrano et al. 2001): 3 tetranucleotides (GATA28, GATA53, GATA417, Palsboll et al. 1997b), 1 trinucleotide (TAA 31, Palsboll et al. 1997b) and 7 dinucleotides (464/465, Schlotterer et al. 1991, and EV1, EV14, EV21, EV37, EV94, EV104, Valsecchi & Amos 1996). Amplifications were conducted in a final volume of 20  $\mu$ l at the following reaction concentrations: 2.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTP, 0.4 mM each primer, 0.5 U AmpliTaq Gold (Applied Biosystems), 1  $\times$  PCRII reaction buffer (Applied Biosystems) and 1  $\mu$ l of DNA (approximately 50 ng). All loci were amplified as individual reactions (i.e. PCR reactions were not multiplexed). The thermocycle profile consisted of initial denaturation at 94°C for 10 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 50°C for 30 s, and extension at 72°C for 30 s, with a final extension of 10 min at 72°C.

To allow for efficient genotyping, the 11 microsatellites were labeled fluorescently and organized into 2 sets to avoid overlap in size ranges prior to multiplex electrophoresis on an ABI373 automated sequencer (Applied Biosystems): Set 1, GATA28 (FAM), GATA53 (TET), EV14 (FAM), EV104 (TET), and EV94 (FAM); Set 2, EV1 (HEX), EV21 (FAM), EV37 (HEX), 464/465 (FAM), GATA417 (FAM), and TAA31 (TET). A 2  $\mu$ l aliquot of each multiplex was dried and resuspended in loading buffer (76% formamide, 14% dye, 10% TAMRA350 size standard, Applied Biosystems). Of each set, 1  $\mu$ l was loaded on a 6% bis-acrylamide (29:1) gel containing 50% urea and electrophoresed for 3.5 h, following the manufacturer's recommendations (Applied Biosystems). Data were collected and alleles were sized with the software programs GENESCAN and GENOTYPER 2.5 (Applied Biosystems). Alleles were sized by comparison with the internal size standard using the local Southern algorithm and binned to the nearest whole base in length. A standard set of 4 to 8 samples was run in all gels as internal controls and to provide an 'allelic ladder' for standard binning of alleles across gels (Ghosh et al. 1997, Dalebout et al. 2002).

**Probability of identity and paternity inference from nuclear markers.** Two microsatellite loci proved difficult to interpret consistently and were omitted from the statistical analyses. One locus (GATA53) generated occasional artefactual fragments in the size range of

known alleles, and another (EV14) proved difficult to size automatically because of low peak height or stutter. For the remaining 9 loci, the level of variation was estimated as the number of alleles per locus and the expected heterozygosity. The loci were tested for deviation from Hardy Weinberg equilibrium and linkage disequilibria using the program ARLEQUIN (Schneider et al. 2000). The probability of identity was calculated following Paetkau & Strobeck (1994). This represents the average probability that 2 unrelated animals share the same genotype by chance alone. For the purposes of individual identification, genotypes differing by a single allele were reviewed by eye and repeated if there was evidence of ambiguous sizing or allelic dropout (Waits et al. 2001).

Paternity analysis was conducted with the computer program CERVUS 2.0 (Marshall 2001). The program uses the frequency of alleles at each locus to calculate the probability of non-exclusion for each male relative to each cow/calf pair (i.e. the probability of a non-father matching by chance) and the likelihood of paternity for a particular male, relative to the likelihood of paternity for a randomly selected male in the population (Marshall et al. 1998). This likelihood is expressed as the natural log of the likelihood ratios (referred to as the LOD score) summed across all (presumably unlinked) loci. For each cow/calf pair, CERVUS ranks all candidate males according to their LOD score, allowing for correction of a specified level of genotyping error (default value of 0.01; Marshall et al. 1998, Slate et al. 2000).

To assist in judging the statistical confidence of inferred paternities, CERVUS calculates the difference between the LOD score of the most likely male and the second most likely male in the sample. The critical value of this difference, referred to as the  $\Delta$ -value, is generated by simulations of parental and offspring genotypes based on the observed allele frequencies, a specified genotype error-rate parameter, and the estimated size of the candidate male population, including unsampled males (Marshall et al. 1998, Slate et al. 2000). The 80 and 95% confidence levels for  $\Delta$ -values were derived from the overlap of distributions in LOD scores for true fathers and the most likely non-fathers generated in the simulations. These confidence values reflect different levels of tolerance for 'false-positive' paternities (type I error), relative to 'false-negative' or falsely excluded paternities (type II error) (Marshall et al. 1998). The choice of confidence limits and the interpretation of  $\Delta$ -values (especially between a non-excluded male and the next most-likely but excluded male) depend on the purpose of the analysis (see 'Discussion'; Slate et al. 2000, Jones & Ardren 2003).

**Organismal capture-recapture estimates.** Individual identification by genotyping and photographs was conducted independently and then compared to con-

sider potential errors in matching. Organismal population abundance was estimated using 2 models: the weighted mean of the Petersen (Begon 1979) based on comparisons between each year, and the sum of all previous years (i.e. 1996 to 1995; 1997 to 1995–96, 1998 to 1995–97, etc.); and Chapman's modification of the Petersen 2-sample model (Seber 1982) based on comparison of the 2001 samples to all previous years (i.e. 2001 to 1995–2000). The latter estimate was considered most comparable to the gametic recapture estimate discussed below. The coefficient of variation (CV) and the symmetrical 95% confidence intervals were calculated following Begon (1979). Two data matrices were used for the organismal estimates. In the photo-identification matrix, the first identification of an individual by photograph constitutes the capture, and the following photographic identification of this animal corresponds to the recaptures. In the genotype and sex matrix, each genotype was treated as a mark, and a recapture is recorded whenever an identical genotype was found in subsequent skin samples. The genotype matrix was partitioned based on molecular sexing, allowing sex-specific estimates of abundance (e.g. Palsboll et al. 1997a). Individuals first identified as calves were excluded from the capture-recapture analyses.

We also considered the potential to merge information from the photo-identification and the genotyping sighting-resighting matrices. For example, records of a whale identified in Year 1 by both a photograph and a genotype, and subsequently identified in Year 3 only by a photograph, could be merged to extend the resight matrix. However, estimates from such mergers using standard methods were biased by the requirement to be 'sighted' at least twice—once in the photo-identification records and once in the genotype records. We did not explore this further, but suggest that new analytic methods are required to take advantage of genotyping for correcting heterogeneity or sex biases in capture probabilities, while utilizing long-term photo-identification records available in some whale populations.

**Gametic capture-recapture estimates.** An alternative estimate of male abundance was derived by 'gametic' capture-recapture analysis based on paternity assignment (e.g. Pearse et al. 2001). The total number of males genotyped across the 7 yr study formed the first capture sample, and the total number of calves genotyped formed the second capture. The number of assigned paternities formed the gametic recapture of the males. The gametic capture-recapture estimate was based on Chapman's modification to the Petersen 2-sample model:

$$N_m = [(n_1 + 1)(n_2 + 1) / (m + 1)] - 1$$

where  $N_m$  is the estimated male abundance,  $n_1$  is the total sample of non-calf males,  $n_2$  is the total number of calves sampled and  $m$  is the total number of inferred paternities. One male, first genotyped as a calf and subsequently as a non-calf, was excluded from the male capture sample, but no attempt was made to exclude other genotyped males based on judgments of size or sexual maturity.

Unlike some other recent studies (e.g. Nielsen et al. 2001, Pearse et al. 2001), the intent of our gametic recapture analysis was not to estimate the effective size of the male population (male  $N_e$ ), but rather to test the null hypothesis that this wintering ground population was closed to male reproductive immigration. We assumed that an agreement between the gametic recapture estimate and the sex-specific organismal estimate would be evidence of reproductive autonomy for the New Caledonia wintering grounds. Alternatively, a gametic recapture estimate that was larger than the organismal recapture estimate would be evidence that male reproduction reflected a larger population, perhaps including males from other wintering grounds (e.g. eastern Australia or Tonga, Garrigue et al. 2002). A smaller gametic estimate would evidence that reproductive males were over-represented in the regional sample. As the gametic capture-recapture estimate was based on inferred paternities summed across all years, it was considered most comparable to the 2-sample organismal estimate of males.

## RESULTS

### Individual identification and organismal recapture estimates

A total of 321 skin samples were analyzed for all 9 microsatellite loci (with the exception of 2 individuals not

Table 1. *Megaptera novaeangliae*. Sample size (n = individual whales), diversity and probability of identity (PI) of microsatellite loci

Loci	n	Observed heterozygosity	Expected heterozygosity	Number of alleles	PI
GATA28	229	0.533	0.549	9	0.221
GATA417	229	0.895	0.902	15	0.018
TAA31	229	0.860	0.870	14	0.031
464/465	229	0.629	0.639	6	0.186
EV1	229	0.520	0.530	4	0.284
EV21	229	0.703	0.673	6	0.146
EV37	229	0.900	0.929	20	0.010
EV94	227	0.855	0.784	10	0.061
EV104	229	0.371	0.356	5	0.407
Mean	228.8	0.696	0.768	9.9	$2.5 \times 10^{-10}$

Table 2. *Megaptera novaeangliae*. Total number of individual whales sighted by photographic identification or by genotyping and the number of individuals newly sighted by each method for each year of the survey

Year	Days at sea (mo)	Photo ID		Genotype ID	
		Total sightings (yr <sup>-1</sup> )	New sightings (yr <sup>-1</sup> )	Total sightings (yr <sup>-1</sup> )	New sightings (yr <sup>-1</sup> )
1995	32 (Aug–Sep)	28	28	5	5
1996	56 (Jul–Sep)	50	41	37	37
1997	43 (Jul–Aug)	47	30	29	25
1998	50 (Jul–Sep)	48	30	43	34
1999	46 (Jul–Aug)	18	11	22	19
2000	50 (Jul–Sep)	36	22	41	32
2001	60 (Jul–Sep)	66	48	79	61

typed for EV94). The number of alleles per locus varied from 4 (EV1) to 20 (EV37) and the observed level of heterozygosity ranged from 0.371 to 0.900 (Table 1).

Average heterozygosity was 0.768 and the average number of alleles was 9.9 over the 9 loci. No significant deviation from the Hardy-Weinberg equilibrium or evidence of linkage disequilibrium was found for the 9 loci (data not shown). The combined probability of identity for the 9 microsatellite loci was  $2.5 \times 10^{-10}$ . After adjusting for multiple comparisons of genotypes in this sample, the expected number of random matches was  $1.3 \times 10^{-5}$ . Based on this low probability, samples with identical genotypes were considered to be replicates or resights of the same individual.

By careful comparison of field records and laboratory replication, 7 field errors and 6 laboratory errors were discovered. Field errors included double sampling of individuals and attribution of samples to the wrong individual in a group. Laboratory errors included mislabeling or mixing of tubes during the extraction or micropipetting. In the photo-identification catalogue, 2 'missed' identifications and 3 cataloguing errors were discovered (i.e. photographs previously considered to represent different individuals were found to represent the same individual). The 2 missed identifications were Type I or II category flukes (mostly white, Carlson & Mayo 1990), one of which showed substantial changes in marks over the years sighted. No 'mis-identifications' were found in the photo-identification records (i.e. no photographs previously considered to represent the same individual were found to represent different individuals).

After correction for errors, the complete photo-identification catalogue for the years 1995 to 2001 contained

293 yearly sightings of 210 individual whales. The complete genotype database contained 321 skin samples derived from 213 unique individuals (excluding 16 calves), of which 175 were also photo-identified. The numbers of individual whales sighted or genotyped in each year are shown in Table 2, and the total for the entire study is shown in Table 3. There was a significant bias towards males in the number of individuals genotyped ( $n = 213$ , 62% males and 38% females;  $p < 0.001$ ) and the number of photo-identified individuals of known sex ( $n = 175$ , 63% males and 37% females;  $p < 0.001$ ). Although these sex ratios were similar to each other, comparisons suggest that not all individuals were equally available for the 2 methods. For example, 45 individuals were photo-identified but not sampled genetically, and

54 individuals were sampled genetically but not photo-identified. In this last set, 30% of the animals were cow/calf pods.

Using the weighted mean of the Petersen model, the abundance estimates of the New Caledonian population ranged from 327 (CV = 0.11) for the photo-identification matrix ID to 533 (CV = 0.15) for the genotype matrix (Table 3). The estimate and CV were larger for the genotype matrix than for the photo-identification matrix. The sex-specific estimates based on genotyping were similar ( $N = 288$  males, CV = 0.18;  $N = 248$  females, CV = 0.30) despite the biased sex ratio of the sample. The Petersen 2-sample model, using all individuals sighted from 1995 to 2000 as the first sample and individuals sighted in 2001 as the second, gave a slightly larger and less precise estimate ( $N = 574$ , CV = 0.18, for photo-identification;  $N = 643$ , CV = 0.18, for genotyping) (Table 4).

#### Maternity confirmation, paternity inference and gametic recapture

A total of 38 cow/calf pairs were sighted in the 7 yr study, including at least 3 females sighted with calves

Table 3. *Megaptera novaeangliae*. Abundance estimates of New Caledonia humpback whales using the weighted mean of the Petersen model based on photo-identification and genotyping

Matrix	Number of individual whales	Total yearly sightings	Total re-sightings	Estimated abundance (95% CI), CV
Photo-identification	210	293	83	327 (256–398), 0.11
Genotype	213	256	43	533 (371–694), 0.15
Females	80	92	12	248 (101–394), 0.30
Males	133	164	31	288 (185–392), 0.18

Table 4. *Megaptera novaeangliae*. Abundance estimate of New Caledonia humpback whales using Chapman's modification of the Petersen 2-sample model based on photo-identification and genotyping in 2001, compared to total for all previous years

Matrix	Individuals sampled prior to 2001	Individuals sampled 2001	Total re-sightings	Estimated abundance (95% CI), CV
Photo-identification	162	66	18	574 (373–774), 0.18
Genotype	152	79	18	643 (412–874), 0.18
Only female	59	27	6	239 (104–374), 0.29
Only male	93	52	12	382 (220–544), 0.22

in more than 1 yr. Genetic samples were collected from both the cow and her calf in 16 of the 38 yearly sightings. Two of these included replicate sightings of a cow (1 cow sampled 3 times with 3 different calves). The assumed maternity of the cow/calf was confirmed by microsatellite genotyping and by sequencing the mtDNA control region (Appendix 1). The 16 cow/calf pairs had matching mtDNA haplotypes ( $n = 12$  unique haplotypes) and shared 1 or more alleles at each of the 9 microsatellite loci (i.e. no identified cow was excluded as the true mother).

The father of 5 calves was inferred by comparing the genotypes of the 16 cow/calf pairs to the total sample of 133 males (Table 5). Each of the candidate males was considered the most likely father of only 1 calf. For 4 calves, the most likely candidate father matches the paternal alleles at all 9 loci. All other males in the sample were excluded by 1 or more mismatches at paternal alleles, although based on the default genotype

correction implemented by CERVUS, the second most likely candidate had a relatively high LOD score in some cases (Table 5). Each of the non-excluded paternity candidates was supported by a high LOD score (range 5.45 to 7.47) and a low probability of non-exclusion (less than  $1.69 \times 10^{-4}$ ). For the 5th calf, the genotype of the most likely candidate father was supported by a moderately high LOD score (4.2) but showed a mismatch at 1 of the 9 loci. At this locus (EV37), the candidate father was a homozygote for an allele 2 repeats shorter than the non-maternal allele of the calf, suggesting the possibility of allelic dropout in the candidate father or a mutation in the calf. The second most likely father also showed a mismatch for single paternal allele but had a substantially lower LOD value (2.62) than the primary candidate. For a 6th calf (data not shown), 2 candidate males matched paternal alleles at all 9 loci, but the relatively low LOD score of 2.88 for each was considered to be inconclusive evidence for paternity inference by CERVUS (Slate et al. 2000). None of the inferred fathers shared a mtDNA haplotype with its matching cow/calf pair, excluding the possibility that an older full sibling, maternal half-sibling or other maternal relative was mistakenly inferred to be the father of a calf.

Critical values of  $\Delta$  were generated by simulations with CERVUS considering the total number of candidate males to be 382, as estimated from the Petersen 2-sample model and, thus, the proportion of sampled males to be 0.35 (i.e.  $133/382$ ). These  $\Delta$  values were 2.66 for the strict confidence level of 95% and 1.48 for the relaxed confidence level of 80%. Based on the relaxed 80% confidence limits considered appropriate for population level analysis (see 'Discussion'; Slate et al. 2000), the 5 primary candidates with LOD scores greater the 3.0 were significantly more likely than the secondary candidates to be the true father. Two of these were significantly more likely to be true fathers based on the 95% confidence level.

For the purposes of the gametic recapture analysis, we considered that our sample of 133 males included 5 fathers of the 16 sampled calves (see 'Discussion' for criteria). Using Chapman's modification to the Petersen model with the non-calf sample of 133

Table 5. *Megaptera novaeangliae*. Inferred paternities for 5 of the 16 cow/calf pairs sampled in New Caledonia. The probability of non-exclusion, the LOD score (natural log of the likelihood ratios) and  $\Delta$ -values (difference between most likely father and, in italics, second most likely father) were calculated with the default error correction implemented in CERVUS. Significance of  $\Delta$ -values generated by simulations using CERVUS (\*exceeding 80% or relaxed confidence level, \*\*exceeding 95% or strict confidence level)

Calf/cow/most likely father <i>2nd most likely father</i>	Mismatch between calf, cow and male	Probability of non-exclusion	LOD	$\Delta$ values
NI9908/HNC180/HNC040	0	$1.83 \times 10^{-5}$	7.47	1.66*
<i>NI0001</i>	1	$1.83 \times 10^{-5}$	5.81	
NI0143/NI0142/HNC187	0	$1.74 \times 10^{-4}$	5.87	2.56*
<i>HNC017</i>	3	$1.74 \times 10^{-4}$	3.31	
NI0144/HNC247/NI0131	0	$4.85 \times 10^{-5}$	7.02	4.06**
<i>HNC214</i>	3	$4.85 \times 10^{-5}$	2.96	
HNC171/HNC113/HNC088	0	$1.69 \times 10^{-4}$	5.45	3.21**
<i>HNC248</i>	2	$1.69 \times 10^{-4}$	2.24	
NI0020/HNC197/HNC167	1	$2.01 \times 10^{-4}$	4.20	1.66*
<i>NI0113</i>	1	$2.01 \times 10^{-4}$	2.62	

Table 6. *Megaptera novaeangliae*. Gametic recapture estimates using Chapman's modification of the Petersen 2-sample test

Matrix	Number of males	Cow/calf pairs	Inferred paternities	Estimated abundance (95% CI), CV
Gametic	133	16	4	455 (154–755), 0.34
Gametic	133	16	5	379 (158–600), 0.30
Gametic	133	16	6	324 (156–493), 0.26

males as  $n_1$ , the 16 cow/calf pairs as  $n_2$  and the 5 inferred paternities as  $m$ , the gametic recapture estimates of male abundance was 379 (CV = 0.30; Table 6). This was essentially identical to the organismal 2-sample estimate based on male genotypes ( $N = 382$ , CV = 0.22; Table 4). Alternative estimates using the only 4 paternities supported by full non-exclusion, or including the 6th paternity, were considered inconclusive due to a low LOD value ( $m = 4$  or  $6$ , respectively); both fell within the 95% confidence limits of each other and the organismal 2-sample estimate (Table 6).

## DISCUSSION

### Abundance and demographic closure

The combined photo-identification and genotype analyses of humpback whales from New Caledonia provided improved confidence in estimation of population abundance and insight into the utility of the 2 methods for this purpose. A larger and less precise population estimate resulted from genotyping compared to the use of photo-identification for capture-recapture analysis. Similar results have been reported previously (Palsboll et al. 1997a). With humpbacks, a negative bias in abundance estimation by photo-identification has been attributed in part to heterogeneity of capture by photo-identification and female migratory patterns. Cows with a calf have a reduced probability of photo-identification, as they seldom raise their flukes before diving while on the breeding grounds. This behavior does not influence the probability of obtaining a biopsy sample, and it was often possible to collect samples from both members of a cow/calf pair. Biopsy sampling alone, however, cannot reduce heterogeneity due to non-random patterns of female migration. Considering a 2 yr sample from the Caribbean breeding grounds, Palsboll et al. (1997a) found a nearly 2-fold difference between sex-specific population estimates based on genotyping of biopsy samples. The reduced probability of capture for females on breeding grounds has been attributed to greater temporal stratification and habitat preference

by females (Mattila et al. 2001), or to the absence of females in some years of their breeding cycle (Brown et al. 1995).

Unlike the West Indies estimate, our sex-specific estimates for New Caledonia, based on the multi-year capture-recapture model, were similar (288 males; 248 females). We suggest that the long-term sampling design in New Caledonia, and the use of a multi-year capture-recapture model, reduced the effects of sex-specific heterogeneity on accuracy of abundance estimates (e.g. Baker et al. 1992), although this effect remained apparent in the larger CV of the female estimate. The relatively small population and the length of the field effort (covering most of the seasonal presence of humpbacks in New Caledonia) were also likely to have reduced capture heterogeneity for both sexes.

A small number of errors were found, and corrected, during the parallel analysis of photo-identification and genotyping records. The 2 missed identifications found in the photo-identification catalogue are examples of false negative errors (i.e. photographs considered to represent more than 1 individual were found to represent 1 individual). In retrospect, these errors can be attributed to a lack of distinctiveness in the fluke coloration of these 2 individuals, non-optimum photographic quality or fluke angle, and changes in markings over time. These sources of error have been noted before (Carlson & Mayo 1990, Perry et al. 1990, Friday et al. 2000, Stevick et al. 2001) but had not been detected in the New Caledonia catalogue prior to the genotyping, despite careful review. However, the small number of missed photo-identifications should help allay concerns (Anonymous 1998) that whales from the Group V stock are difficult to identify individually because of a high proportion of white flukes (Rosenbaum et al. 1995). Similarly, a small number of errors in laboratory and field records of genotypes were detected by comparison with photographic records. Overall, our results support the general need for scrutiny in the application of photo-identification and genotyping and the advantages of employing both methods where possible.

The combined genotyping and photo-identification records provided strong support for the assumption that the New Caledonia wintering grounds are a significant demographic unit for conservation and management. Demographic closure was suggested previously by collaborative comparisons of photo-identification catalogues from wintering grounds throughout the South Pacific (Garrigue et al. 2000, 2002). Although these comparisons demonstrated some movement of individuals from New Caledonia to eastern Australia, New Zealand and Tonga, the rate of this between-region interchange was low in comparison to the rate of within-region annual return (Garrigue et al. 2002).



### Reproductive autonomy

Our combined use of organismal and gametic recapture methods provided the first direct evidence of male reproductive autonomy for humpback whale wintering grounds in the Southern Hemisphere. The 5 inferred paternities found for the 16 cow/calf pairs (31%) were consistent with the estimated proportion of sampled males (35% based on the 2-sample model) and, consequently, the gametic recapture estimate was essentially identical to the sex-specific organismal estimate. This conformed to the hypothesis that calves were fathered primarily by males in the local New Caledonia population. An alternative hypothesis, that fathers were a random sample of the much larger population migrating past eastern Australia ( $N = 1900$  in 1992, Paterson et al. 1994), could be rejected based on the 95% confidence limits of the gametic recapture estimate for New Caledonia. However, the power of our estimate was not sufficient to reject the possibility of some male reproductive interchange with smaller neighboring wintering grounds where precise estimates of abundance are not available (e.g. Fiji or Tonga; Garrigue et al. 2002). Similar evidence of paternity assignment to males from the common wintering grounds of sampled cow/calf pairs is reported for the offshore islands of Mexico in the North Pacific (Cerchio 2003).

Our gametic recapture approach differed from both 'direct genetic' estimates of total census abundance (Palsboll et al. 1997a, Taberlet et al. 1997, Kohn et al. 1999) and 'parentage' estimates of the number of breeding males in a population (e.g. Nielsen et al. 2001, Pearse et al. 2001). By combining organismal genotype captures and paternal gametic recaptures, we sought an unbiased estimate of total males for comparison to the direct sex-specific estimates (genotype and photo-identification). The organismal genotypes used in both estimates were collected from all encountered groups and thus, we assumed, represented a random sample of the population (excluding only the known calves sampled during the study). As the Petersen model requires only one randomly collected sample (Begon 1979), we considered that the gametic recapture was valid as a second, non-random, sample for the analysis.

Confidence in our gametic-recapture estimate, and in turn our conclusion of reproductive autonomy for the New Caledonia wintering ground, depended on confidence in our paternity assignment. Although there is no general agreement on criteria required for categorical paternity inference in wild populations (e.g. Jones & Ardren 2003), several characteristics of our data and analysis indicated that paternity inferences were robust for 5 of the 16 calves. First, the LOD

score was high ( $>4.2$ ) and the probability of non-exclusion was low ( $<2 \times 10^{-4}$ ) for each of the 5 assignments. Based on a large-scale retrospective assessment of paternity inference in red deer *Cervus elaphus* on the Isle of Rum, Slate et al. (2000) considered an LOD of  $>3.0$  sufficient to confirm paternity. Second, the number of loci was relatively large and the heterozygosity of most loci was high. In a simulation of paternity inference in a population of 500 individuals, Nielsen et al. (2001) found that the probability of a correct assignment was high using as few as 5 loci with 10 alleles of equal frequency or 10 loci with 4 alleles of equal frequency. Our use of 9 loci with an average of nearly 10 alleles and average heterozygosity of 0.768 was intermediate between these 2 scenarios. Third, our experimental error for genotyping was reduced by the collection of corresponding photo-identification records and visual review of the automated size binning implemented in GENOTYPER. In the assignment involving a single missing paternal allele, we concluded that the mismatch was likely due to allelic dropout in the father (an apparent homozygote) or a mutation in the calf. Fourth, the  $\Delta$ -values for each of the 5 inferences exceeded the 80% confidence limits considered appropriate for population-level analysis (as opposed to the strict confidence limits recommended for estimating individual reproductive success; Slate et al. 2000, Jones & Ardren 2003). Furthermore, the  $\Delta$ -statistic is intended to distinguish between alternate, non-excluded males, not between non-excluded males and excluded males included as a result of the error-rate correction implemented in CERVUS (Jones & Ardren 2003). Under a strict likelihood criterion, the second most likely males would have an LOD of 0.00 in 4 of the inferred paternities in our sample.

Finally, we used mtDNA haplotypes of cow/calf pairs to exclude the possibility that an older full sib of the calf or other maternal relative was not mistakenly inferred to be the father. In none of the 5 inferred paternities did the mtDNA of the candidate male match that of the cow/calf pair. Surprisingly, the asymmetry of nuclear and mtDNA inheritance has not been exploited systematically in other statistical approaches to paternity inference in wild populations (e.g. Marshall et al. 1998, Nielsen et al. 2001).

### Reproductive autonomy and evolutionary differentiation

The confirmation of reproductive autonomy by gametic recapture analysis has several advantages over the traditional analysis of evolutionary differentiation (i.e. gene flow). First, the gametic recapture can

be applied with samples from a single wintering ground, whereas estimates of differentiation require samples from other regions for comparison. Second, the gametic recapture, like assignment methods (e.g. Davies et al. 1999), is not dependent on traditional population genetic assumptions about drift/mutation/migration equilibrium (although the paternity calculations assume Hardy-Weinberg equilibrium, Marshall et al. 1998). These advantages, however, can also be viewed as limitations. Although the gametic recapture analysis can help confirm current patterns of reproduction and migration, it provides little information on the stochasticity of these patterns. As such, it is limited in inference about the question of units to conserve on an evolutionary time scale. Further, it is essentially a demographic method extended by one generation and requires relatively intensive sampling to achieve reasonable precision. In the case of New Caledonia, skin sampling included roughly 1/3 of the estimated male population and half of the calves observed during the 7 yr study. Such an intensive effort might not be practical in larger populations.

Variance in male reproductive success will also influence the efficiency of the gametic recapture. If variance is high, the required sampling effort for this approach will be greater, and the resulting estimate subject to a form of capture heterogeneity. On an evolutionary time scale, however, high variance in male reproductive success will enhance the effect of drift, requiring less sampling effort to detect differences due to low gene flow. We suggest the greatest understanding of humpback whale populations is likely to result from combining and contrasting individual and evolutionary approaches.

#### Current abundance and recovery

Together with the evidence of demographic closure and maternal fidelity to migratory destinations (e.g. Baker & Medrano 2002), the evidence of reproductive autonomy emphasizes the potential vulnerability of subdivided populations to local extirpation (Clapham et al. 1999). Nearly 4 decades have passed since humpback whales in Oceania reached their lowest numbers, probably following the intensive unreported Soviet hunt in Antarctic waters during the summer of 1961/62 (Mikhalev 2000). Assuming maximum rates of annual increase used by the IWC in models of recovering humpback whale populations (e.g. 7 to 10%, Brandao et al. 2000), extrapolation backwards from current estimates of abundance would suggest that the New Caledonia population was perilously close to extinction at that time. Humpbacks from other wintering grounds in Oceania were also exposed to intensive hunting on the feeding

grounds but, by chance, might have differed in the number of survivors. Even small differences in such low numbers of individuals surviving could help explain the variability now observed in abundance and rates of increase among stocks in the Southern Hemisphere.

*Acknowledgements.* Surveys of humpback whales in New Caledonia were made possible by contributions of Les Editions Catherine Ledru, and the Provinces Sud, North and Isles, and Inco S.A. Partial support for the 1999, 2000 and 2001 research seasons was funded by grants from the International Fund for Animal Welfare (IFAW) to the South Pacific Whale Research Consortium. We thank M. Donoghue from the Department of Conservation of New Zealand for initiating and managing this funding and Consortium members, and M. Poole, N. Hauser, D. Patton and P. Clapham for discussion. We thank J. Greaves, D. Breitenstein, M. Chambellant and V. Ducreux and all the voluntary workers that helped in the field. Field research was carried out by 'Opération Cétacés' and laboratory analysis at the Laboratory of Molecular Ecology and Evolution, University of Auckland. Support for laboratory analysis was provided by the University of Auckland Research Council (to C.S.B. and S. Lavery), the Conservation Action Fund of the New England Aquarium (to C.S.B and M. Poole), the International Fund for Animal Welfare (to C.S.B.) and the Marsden Fund of New Zealand (to C.S.B.).

#### LITERATURE CITED

- Amos B, Schlotterer C, Tautz D (1993) Social structure of pilot whales revealed by analytical DNA profiling. *Science* 30: 670–672
- Anderson MJ, Hinton G, Paton D, Baverstock PR (2001) A model for integration of microsatellite genotyping with photographic identification of humpback whales. *Mem Queensl Mus* 47:451–458
- Anonymous (1998) Annex G Report of the sub-committee on comprehensive assessment of southern hemisphere humpback whales. *Rep Int Whal Comm* 48:170–182
- Anonymous (2000) Report of the scientific committee. *J Cetacean Res Manag (Suppl)* 2:1–65
- Baker CS, Medrano-Gonzalez L (2002) World-wide distribution and diversity of humpback whale mitochondrial DNA lineages. In: Pfeiffer CJ (ed) *Cell and molecular biology of marine mammals*. Krieger Publishing, Melbourne, FL, p 84–99
- Baker CS, Palumbi SR (1995) Population structure, molecular systematics and forensic identification of whales and dolphins. In: Avise J, Hamrick JL (eds) *Conservation genetics: case histories from nature*. Chapman & Hall, New York, p 10–49
- Baker CS, Palumbi SR, Lambertsen RH, Weinrich MT, Calambokidis J, O'Brien SJ (1990) Influence of seasonal migration on the distribution of mitochondrial DNA haplotypes in humpback whales. *Nature* 344:238–240
- Baker CS, Straley JM, Perry A (1992) Population characteristics of individually identified humpback whales in southeastern Alaska: summer and fall 1986. *Fish Bull* 90:429–437
- Baker CS, Perry A, Bannister JL, Weinrich MT and 10 others (1993) Abundant mitochondrial DNA variation and world-wide population structure in humpback whales. *Proc Natl Acad Sci USA* 90:8239–8243
- Baker CS, Slade RW, Bannister JL, Abernethy RB and 7 others

- (1994) Hierarchical structure of mitochondrial DNA gene flow among humpback whales *Megaptera novaeangliae*, world-wide. *Mol Ecol* 3:313–327
- Baker CS, Flórez-González L, Abernethy B, Rosenbaum HC, Slade RW, Capella J, Bannister JL (1998) Mitochondrial DNA variation and maternal gene flow among humpback whales of the Southern Hemisphere. *Mar Mamm Sci* 14: 721–737
- Baker CS, Donoghue M, Constantine R (2001) Population status and genetic identity of humpback whales on the wintering grounds of Tonga—a slow recovery or colonization? XIVth Biennial Conf Biol Mar Mamm, November 28 to December 3, 2001, Vancouver
- Bannister JL (1994) Continued increase in Group IV humpbacks off Western Australia. *Rep Int Whal Comm* 44:309–310
- Begon M (1979) Investigating animal abundance: capture-recapture for biologist. University Park Press, Baltimore
- Brandao A, Butterworth DS, Brown MR (2000) Maximum possible humpback whale increase rates as a function of biological parameter values. *J Cetacean Res Manag (Suppl)* 2:192–193
- Brown MR, Corkeron PJ, Hale PT, Schultz KW, Bryden MM (1995) Evidence for a sex segregated migration in the humpback whale (*Megaptera novaeangliae*). *Proc R Soc Lond B* 259:229–234
- Bryden MM, Kirkwood GP, Slade RW (1990) Humpback whales, Area V. An increase in numbers off Australia's east coast. In: Kerry KH, Hempel G (eds) Antarctic ecosystems. Ecological change and conservation. Springer-Verlag, Berlin, p 271–277
- Carlson CA, Mayo CA (1990) Changes in the ventral fluke pattern of the humpback whale (*Megaptera novaeangliae*), and its effect on matching; evaluation of its significance to photo-identification research. *Rep Int Whal Comm* 12(Spec Issue):105–111
- Cerchio S (2003) Paternity, polygyny and alternative mating tactics in humpback whales (*Megaptera novaeangliae*). PhD, University of Michigan, Ann Arbor, MI
- Clapham PJ (1996) The social and reproductive biology of humpback whales: an ecological perspective. *Mamm Rev* 26:27–49
- Clapham PJ, Baker CS (2002) Modern whaling. In: Perrin WF, Würsig B, Thewissen JGM (eds) Encyclopedia of marine mammals. Academic Press, New York, p 1328–1332
- Clapham PJ, Palsboll PJ, Mattila DK (1993) High-energy behaviors in humpback whales as a source of sloughed skin for molecular analysis. *Mar Mamm Sci* 9(2):213–220
- Clapham PJ, Young SB, Brownell RL Jr (1999) Baleen whales: conservation issues and status of the most endangered populations. *Mamm Rev* 29:35–60
- Dalebout ML, Lento GM, Cipriano F, Funahashi N, Baker CS (2002) How many protected minke whales are sold in Japan and Korea? A census by DNA profiling. *Anim Conserv* 5:143–152
- Davies N, Villablanca FX, Roderick GK (1999) Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. *Trends Ecol Evol* 14:17–21
- Donovan GP (1991) A review of IWC stock boundaries. *Rep Int Whal Comm* 13(Spec Issue):39–68
- Friday NA, Smith TD, Stevick PT, Allen J (2000) Measurement of photographic quality and individual distinctiveness for the photographic identification of humpback whales, *Megaptera novaeangliae*. *Mar Mamm Sci* 16(2):355–374
- Garrigue C, Greaves J (1999) New Caledonia, a rendezvous for the humpback whale. C. Ledru, Nouméa
- Garrigue C, Forestell P, Greaves J, Gill P, Naessig P, Baker CS, Patenaude N (2000) Migratory movement of humpback whales (*Megaptera novaeangliae*) between New Caledonia, East Australia and New Zealand. *J Cetacean Res Manag* 2:111–115
- Garrigue C, Greaves J, Chambellant M (2001) Characteristics of the humpback whale population at New Caledonia (South West Pacific). *Mem Queensl Mus* 47:539–546
- Garrigue C, Aguayo A, Amante-Helweg VLU, Baker CS and 12 others (2002) Movements of humpback whales in Oceania, South Pacific. *J Cetacean Res Manag* 4(3):255–260
- Ghosh S, Karanjawala ZE, Hauser ER, Ally D and 20 others (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labelled dinucleotide markers. *Genome Res* 7:165–178
- Gilson A, Syvanen M, Levine K, Banks J (1998) Deer gender determination by polymerase chain reaction: validation study and application to tissues, bloodstains and hair forensic samples from California. *Calif Fish Game* 84(4): 159–169
- Hammond PS, Mizroch SA, Donovan GP (eds) (1990) Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. *Rep Int Whal Com (Spec Iss)* 12:1–440
- Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Mol Ecol* 12:2511–2523
- Katona S, Baxter B, Brazier O, Kraus S, Perkins J Whitehead H (1979) Identification of humpback whales by fluke photographs. In: Winn HE, Olla BL (eds) Behavior of marine animals, Vol 3. Plenum Press, New York, p 33–44
- Kohn MH, York EC, Kamradt DA, Haught G, Sauvajot RM, Wayne RK (1999) Estimating population size by genotyping faeces. *Proc R Soc Lond B* 266:657–663
- Lambertsen RH (1987) A biopsy system for large whales and its use for cytogenetics. *J Mamm* 68:443–445
- Larsen AH, Sigurjonsson J, Oien N, Vikingsson G, Palsboll P (1996) Population genetic analysis of nuclear and mitochondrial loci in skin biopsies collected from central and northeastern North Atlantic humpback whales (*Megaptera novaeangliae*): population identity and migratory destinations. *Proc R Soc Lond B* 263:1611–1618
- Mackintosh NA (1965) The stocks of whales. Fishing News, London
- Marshall T (2001) CERVUS 2.0 software. University of Edinburgh
- Marshall TC, Slate S, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* 7:639–655
- Mattila DK, Allen J, Clapham PJ, Friday N and 11 others (2001) Recent findings concerning the migration and breeding ground composition of North Atlantic humpback whales (Abstract). *Mem Queensl Mus* 47:554
- Medrano-Gonzalez L, Baker CS, Robles-Saavedra MR, Murrell J and 11 others (2001) Transoceanic population genetic structure of humpback whales in the North and South Pacific. *Mem Queensl Mus* 47:554
- Mikhalev YA (2000) Biological characteristics of humpbacks taken in Antarctic Area V by the whaling fleets *Slava* and *Sovietskaya Ukraina*. Unpubl. report (SC/52/IA12) to the Scientific Committee of the International Whaling Commission. The Red House, 135 Station Road, Impington, Cambridge CB4 9NP
- Mizroch SA, Beard JA, Macgill L (1990) Computer assisted photo-identification of humpback whales. *Rep Int Whal Comm (Spec Issue)* 12:63–70
- Nielsen R, D Mattila, PJ Clapham, Palsboll P (2001) Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157:1673–1682

- Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. *Mol Ecol* 3:465–480
- Palsboll PJ, Clapham PJ, Mattila DK, Larsen F and 5 others (1995) Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Mar Ecol Prog Ser* 116:1–10
- Palsboll PJ, Allen J, Berubé M, Clapham PJ and 15 others (1997a) Genetic tagging of humpback whales. *Nature* 388:767–769
- Palsboll PJ, Bérubé M, Larsen AH, Jorgensen H (1997b) Primers for the amplification of tri- and tetra microsatellite loci in baleen whales. *Mol Ecol* 6:893–895
- Paterson R, Paterson P, Cato DH (1994) The status of humpback whales (*Megaptera novaeangliae*) in East Australia thirty years after whaling. *Biol Conserv* 70:135–142
- Pearse DE, Eckerman CM, Janzen FJ, Avise C (2001) A genetic analogue of 'mark-recapture' methods for estimating population size: an approach based on molecular parentage assessments. *Mol Ecol* 10:2711–2718
- Perry A, Baker CS, Herman LM (1990) Population characteristics of individually identified humpback whales in the central and eastern North Pacific: a summary and critique. *Rep Int Whal Comm (Spec Issue 12)*:307–318
- Richards KR, Whitehead H, Wright JM (1996) Polymorphic microsatellites from sperm whales and their use in the genetic identification of individual from naturally sloughed skin. *Mol Ecol* 5:313–315
- Rosenbaum HC, Clapham PJ, Allen J, Nicole-Jenner M and 6 others (1995) Geographic variation in ventral fluke pigmentation of humpback whale (*Megaptera novaeangliae*) populations worldwide. *Mar Ecol Prog Ser* 124:1–7
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor Laboratory Press, New York
- Schlotterer C, Amos B, Tautz D (1991) Conservation of polymorphic simple sequence loci in cetacean species. *Nature* 354:63–65
- Schneider S, Roesslo D, Excoffier L (2000) Arlequin: a software for population genetics analysis. Genetics and Biometry Lab, University of Geneva
- Seber GAF (1982) *The estimation of animal abundance and related parameters*, 2nd edn. Charles Griffin, London
- Slate J, Marshall T, Pemberton J (2000) A retrospective assessment of the accuracy of the paternity inference program CERVUS. *Mol Ecol* 9:801–808
- Stevick PT, Palsboll PJ, Smith TD, Bravington MV, Hammond PS (2001) Errors in identification using natural markings: rates, sources, and effects on capture-recapture estimates of abundance. *Can J Fish Aquat Sci* 58:1861–1870
- Taberlet P, Waits LP, Luikart G (1997) Non-invasive genetic tracking of the endangered Pyrenean brown bear population. *Mol Ecol* 6:869–876
- Valsecchi E, Amos W (1996) Microsatellite markers for the study of cetacean populations. *Mol Ecol* 5:151–156
- Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions, guidelines. *Mol Ecol* 10:249–256
- Zemsky VA, Berzin AA, Mikhaliyev YA, Tormosov DD (1995) Soviet Antarctic pelagic whaling after WWII: Review of actual catch data. *Rep Int Whal Comm* 45:131–135

**Appendix 1.** *Megaptera novaeangliae*. Genetic profiles of the 5 cow/calf and candidate father trios, including 2nd most likely father based on CERVUS analysis (F, female; M, male). In the first 4 cases, the 2nd most likely male is excluded by 1, 2 or 3 mismatches (see also Table 5). na: not available

ID	Status	Sex	mtDNA	464/465	GATA417	TAA31	EV1A	EV37	EV21	EV94	EV104	GATA28
(1)												
NI9908	Calf	F	NC9709	149/149	187/214	106/112	123/127	206/214	113/117	208/216	149/149	147/147
HNC180	Cow	F	NC9709	143/149	214/222	106/112	123/127	212/214	111/117	208/210	149/149	147/147
HNC040	Most likely father	M	NC9601	143/149	187/214	106/109	123/127	202/206	113/117	212/216	149/149	147/147
NI0001	2nd most likely father	M	NC9635	137/149	187/230	97/106	123/123	202/202	111/113	214/216	149/149	147/147
(2)												
NI0143	Calf	F	NC9505	139/149	195/199	88/103	123/127	194/204	109/111	212/216	147/149	147/187
NI0142	Cow	F	NC9505	143/149	199/226	88/103	123/127	192/204	109/111	214/216	147/153	147/183
HNC187	Most likely father	M	NC0034	137/139	195/199	88/103	123/127	194/194	109/111	212/220	149/149	147/187
HNC017	2nd most likely father	M	NC9603	139/139	199/222	100/103	127/127	214/214	111/113	202/218	147/149	147/187
(3)												
NI0144	Calf	F	NC9608	139/141	222/226	97/106	123/123	192/194	109/111	212/218	149/149	147/179
HNC247	Cow	F	NC9608	139/139	222/222	106/115	123/123	192/198	111/111	212/218	149/149	179/183
NI0131	Most likely father	M	NC9504	141/143	203/226	97/103	123/123	194/214	109/115	212/216	149/149	147/147
HNC214	2nd most likely father	M	NC0123	141/143	187/195	106/112	123/123	198/204	109/111	212/214	149/149	147/147
(4)												
HNC171	Calf	M	NC9713	137/139	199/207	97/100	123/127	198/202	109/111	212/214	149/149	175/187
HNC113	Cow	F	NC9713	137/139	199/199	100/106	123/127	198/212	111/111	212/214	149/149	147/175
HNC088	Most likely father	M	NC9907	131/139	187/207	97/100	123/123	202/208	109/115	214/216	149/149	147/187
HNC248	2nd most likely father	M	na	139/139	199/207	100/100	123/127	202/208	109/115	210/214	149/149	147/147
(5)												
NI0020	Calf	M	NC9805	139/139	199/195	97/103	123/127	198/218	109/111	206/208	149/149	147/155
HNC197	Cow	F	NC9805	139/139	199/195	97/103	123/127	198/214	109/117	208/208	149/151	147/147
HNC167	Most likely father	M	NC9841	139/139	199/207	97/115	123/127	210/210	111/115	206/218	149/149	147/155
NI0113	2nd most likely father	M	NC9608	139/139	199/211	97/97	123/127	204/218	111/111	206/212	149/151	147/147