Genetic tagging of male North Atlantic minke whales through comparison of maternal and foetal DNA-profiles

HANS J. SKAUG AND NILS ØIEN

Institute of Marine Research, PO Box 1870, N-5817 Bergen, Norway

Contact e-mail: Hans.Skaug@imr.no

ABSTRACT

DNA-profiles from 288 mother-foetus pairs were used to obtain partial DNA-profiles for the fathers of the foetuses. The paternal profiles were subsequently matched against those of the males on the Norwegian DNA-register for minke whales using statistical analyses. Three likely instances of paternity were identified. An estimate of the number of reproductively active males in the population was calculated and found to be consistent with previous abundance estimates. However, the associated confidence interval was very broad since it was based on few 'recaptures'. Finally the scope and potential use of such genetic and population data is discussed.

KEYWORDS: ATLANTIC OCEAN; GENETICS; MARK-RECAPTURE; MOVEMENTS; REPRODUCTION; EUROPE; DNA FINGERPRINTING; COMMON MINKE WHALE

INTRODUCTION

Parentage studies based on DNA-profiles are now commonly conducted for many species of wildlife (Marshall *et al.*, 1998; Jones and Arden, 2004), however, so far there have been few for baleen whales (Clapham and Palsbøll, 1997; Nielsen *et al.*, 2001; Garrigue *et al.*, 2003). Determination of biological paternity *per se* is not usually the primary goal of such studies. Identification of father-offspring pairs can yield information about animal abundance (Nielsen *et al.*, 2001), gene flow between subpopulations (Amos *et al.*, 1993) and reproductive success in different behavioural groups of animals (Nielsen *et al.*, 2001).

A particularly advantageous situation for paternity studies arises when DNA-profiles from mother-offspring pairs, confirmed on non-genetic grounds, are available. By comparing their DNA-profiles, one of the father's two alleles can be inferred at each locus[1], yielding a partial paternal DNA-profile. This profile can subsequently be compared against a DNA-database of potential fathers. In the studies cited above, the database covers a large proportion of the male population. In the present study, on the other hand, only a small fraction of the males are present in the database, and hence a classical mark-recapture approach is used. The partial DNA profiles serve as tags for the fathers. Unless the number of genetic loci is very large, many tags will not be unique in the population, i.e. there may be males other than the true father that match an inferred partial DNA-profile. Thus, statistical analyses are required to calculate the 'specificity' of each father profile. The specificity is a measure of the usefulness of the tag in a mark-recapture setting. A related concept is that of the 'paternity probability' (Nielsen *et al.*, 2001). When a match with the database is obtained, one can calculate the probability that the true father has been found, as opposed to an unrelated male matching by chance.

The establishment of the Norwegian DNA-register for common minke whales (Olaisen, 1997) has provided an opportunity to perform paternity studies for northeastern Atlantic common minke whales (*Balaenoptera acutorostrata*). The register currently contains DNA-profiles (10 microsatellites loci and mtDNA) for 3,301 individuals caught by Norwegian whalers in the period 1997-2002. From the year 2000, foetal tissue samples have also been collected from pregnant females.

The migration pattern of common minke whales in the North Atlantic is not known but it has been speculated that they may enter the Northeast Atlantic feeding areas through the Denmark Strait and north of the British Isles. Recent sightings surveys (Skaug *et al.*, 2004) have revealed that common minke whales summer in fairly large numbers in the Norwegian, Greenland, North and Barents Seas. Although their numbers can vary through seasons and between years, no clear migration patterns are apparent from those data. According to Jonsgård (1951; 1955) common minke whales migrate into Norwegian and Arctic waters in the spring, are most frequent there in the summer, and leave these northern waters again more or less completely in the autumn. Immigration in the spring begins apparently in the southern and western areas and continues along the coast. There is segregation both with respect to length and sex (Jonsgård, 1951; Øien, 1988). Large females dominate in Skagerrak (Fig. 1, south-eastern part of the EN area) and in the main Barents Sea and off Spitsbergen, while large males dominate in the rest of the EN area (Fig. 1). During 1974-78, 333 minke whales were marked with Discovery tags in the Barents Sea. In addition, 18 individuals had been tagged prior to 1974 and 15 individuals have been tagged after 1978. Of the total 366 tags applied, 33 have been recovered in the commercial minke whale catches (Christensen and Rorvik, 1978; Beddington *et al.*, 1984). Locations for tagging and recaptures of the Discovery tags are shown in Fig. 1.

A key question is whether the 288 inferred partial DNA-profiles obtained in the present study are sufficiently specific to provide useful information about paternity. The number of fathers that can be identified also depends crucially on the proportion of the male population covered by the DNA-register. Simple calculations, based on the number of males present in the DNA-register and the best estimate of total population size (Skaug *et al.*, 2004), show that the expected number of recaptured fathers is rather low (approximately five). Thus, our ability to gain new

[1] Not uniquely though, in the situation where a heterozygotous offspring shares both alleles with the mother.

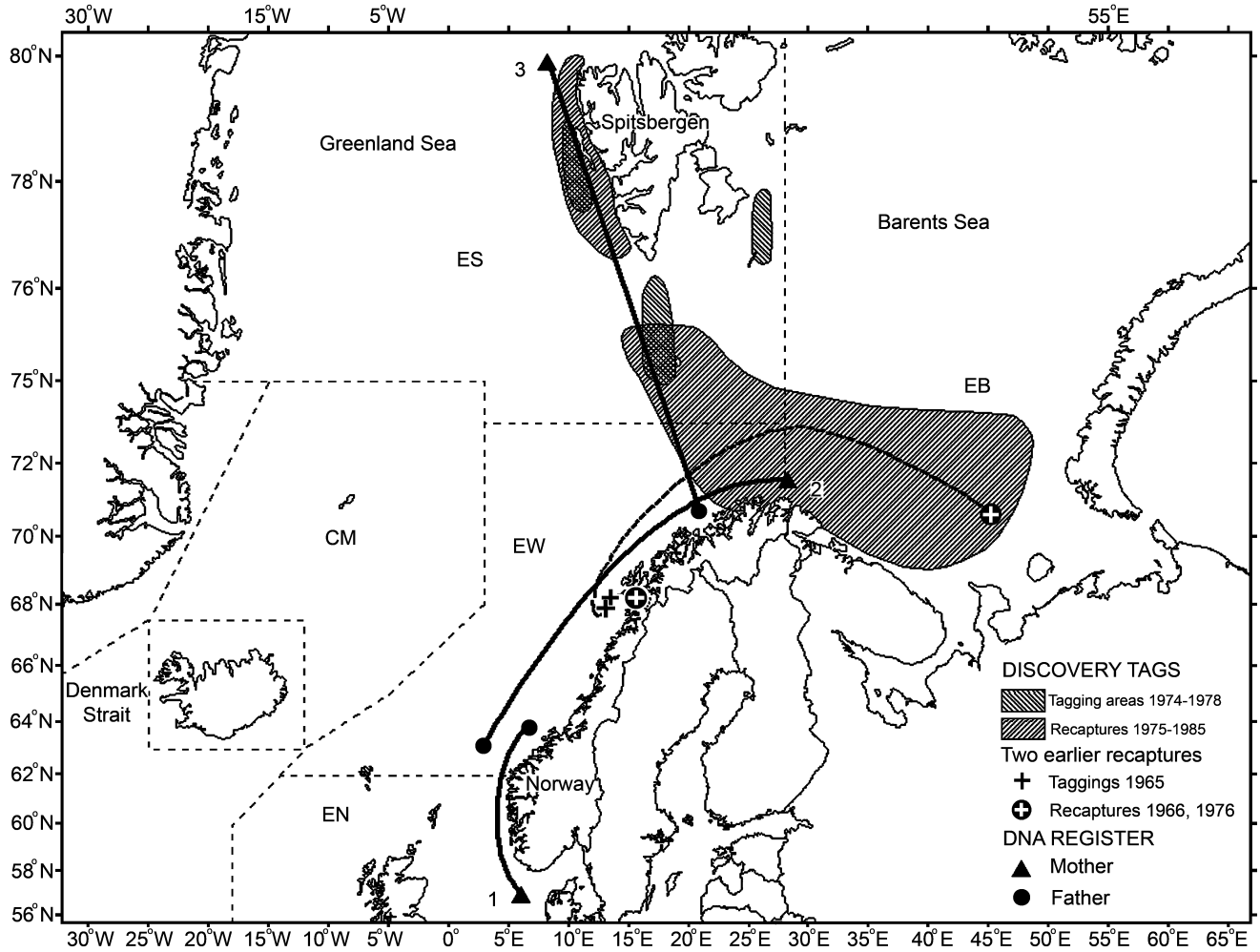

Fig. 1. Location of tagging and tag-recoveries, both for previously applied Discovery tags and for the genetic tags obtained in the present study. Dashed lines indicate borders between Small Areas. The numbers associated with mothers provide a link to Table 3.

biological information about common minke whales is rather limited. The goal was rather to report on the feasibility of the approach, and to point to potential applications if larger datasets should become available.

MATERIAL AND METHODS

Origin and nature of samples

The establishment of the Norwegian common minke whale DNA-register ensures that samples (muscle tissues) are taken from each animal caught under the Norwegian catch quota, and that a DNA-profile for each whale is established and stored in a database (Olaisen, 1997). The DNA-profile consists of 10 microsatellites, mtDNA and a sex-marker (Dupuy and Olaisen, 1998). In addition, for each animal, the register contains information about the time and geographical location of capture, as well as some biological parameters (length etc.). At the time of writing the register contained information (Table 1) on 3,301 individuals, out of the total of 3,392 individuals caught during the period 1997-2002. All individuals were caught in the season from April to September.

Starting from year 2000, foetal tissue samples have also been taken from pregnant females. In this study we have established the DNA-profiles of 288 foetuses, using the same protocol (Dupuy and Olaisen, 1998) and laboratory as has been used for the DNA-register. As mtDNA is maternally inherited, it does not carry information about the father and is only used for data checking, to guard against accidental sample switching during data collection.

Table 1

Norwegian catches of minke whales in the North Atlantic by Small Area for the period 1997-2002. The number of animals used in the present analysis is given in parenthesis.

	EN	EC	EB	ES	CM	Total
1997	57(53)	14(12)	283(280)	129(124)	20(19)	503(488)
1998	139(131)	15(14)	285(281)	129(126)	57(57)	625(609)
1999	122(116)	12(12)	287(277)	112(111)	58(55)	591(571)
2000	83(81)	16(8)	228(224)	103(101)	57(56)	487(470)
2001	128(124)	11(10)	262(257)	120(116)	31(31)	552(538)
2002	132(129)	13(13)	308(307)	146(141)	35(35)	634(625)
Total	661(634)	81(69)	1,653(1,626)	739(719)	258(253)	3,392(3,301)

Tag specificity and paternity probability

The specificity, p_{spec}, of a partial DNA-profile is a statement about how rare its constituent alleles are in the population. Thus, tag specificity can be calculated from maternal and foetal DNA-profiles alone, without consideration of any candidate fathers. Later, when given a database of candidate fathers, paternity probabilities can be assigned to each individual in the database according to formula (1) below, adopted from Nielsen *et al.* (2001).

The offspring and maternal DNA-profiles are denoted by O_i and M_i, respectively, for the *i*th foetus-female pair. Because the offspring inherits one allele from each parent, the part of O_i that is paternally inherited can be inferred by comparison of O_i with M_i (Table 2). At loci where the offspring is a heterozygote, it is not possible to determine which of the two alleles is inherited from the father (locus GT211 in Table 2). The partial DNA-profile (shaded part of the paternal-profile in Table 2) can be used to exclude potential fathers. Non-exclusion by this criterion cannot be taken as absolute evidence for paternity since consistency could occur by chance. The specificity, p_{spec}, of a genetic tag (a partial DNA-profile) is defined here as the probability that there are no males in the population, other than the true father, that are consistent with the tag. If f_l is the population frequency of the inferred allele at locus l then the probability that a given male (non-father) is consistent at L loci by chance is $p_0 = \Pi_{l=1}^{L}\left[1-\left(1-f_l\right)^2\right]$, and hence the specificity is given as: $p_{spec} = \left(1-p_0\right)^{N_m-1}$, where N_m is the number of reproductively active males in the population. A practical interpretation of p_{spec} is as a measure of 'tag quality', as a value of p_{spec} close to one means that the true father will be identified with near certainty if his profile is present in the database.

The DNA-profile of the *j*th male in the DNA-register is denoted by F_j. The paternity probability, $P_i(j)$, is defined as $P_i(j)$ = Pr(Male *j* is father of fetus *i*). Under the assumption that the J males contained in the DNA-register constitute a random sample from the N_m reproductively active males in the population, it follows (Nielsen *et al.*, 2001) that the paternity probability is given as:

$$P_i(j) = \frac{P(O_i \mid M_i, F_j)}{\sum_{k=1}^{J} P(O_i \mid M_i, F_k) + (N_m - J)P(O_i \mid M_i)} \qquad (1)$$

Here, $P(O_i \mid M_i, F_j)$ is the conditional probability of the offspring DNA-profile, given both maternal and paternal profiles. Similarly, $P(O_i \mid M_i)$ is the conditional probability of O_i given only the mother profile. Expressions for these probabilities can be derived from Mendel's law, together with the assumption that loci are independent, so that probabilities can be multiplied across loci. As pointed out by Nielsen *et al.* (2001), formula (1) can be interpreted as a Bayesian posterior probability. Formula (1) takes into account the fact that the DNA-register only covers a proportion J / N_m of the male population. The (posterior) probability that the DNA-register contains the true father of the *i*th foetus is $\sum_{j=1}^{J} P_i(j)$.

The year in which the *i*th mother-foetus pair was captured is denoted by y_i. Because the gestation period for common minke whales is suggested to be around 10 months (Horwood, 1990), only males in the DNA-register caught in year y_i or later were used to calculate $P_i(j)$. To emphasise this J is replaced by J_i, the number of males contained in the DNA-register caught in year y_i or later.

Abundance estimation

As with ordinary mark-recapture experiments, the data obtained in the present study can be used to estimate animal abundance, but for this case the uncertainty associated with the tag needed to be reflected. Expression (7) from Nielsen *et al.* (2001) was modified to obtain the log-likelihood function:

$$l(N_m) = \sum_{i=1}^{288} \log\left(\frac{N_m - J_i}{N_m} P(O_i \mid M_i) + \frac{1}{N_m}\sum_{j=1}^{J_i} P(O_i \mid M_i, F_j)\right)$$

An estimate of N_m was found by maximising $l(N_m)$, using a simple bisection algorithm.

RESULTS

Among the 288 genetic tags obtained in this study, five were consistent with two or more males in the DNA-register (matching at each of the 10 loci). All of these tags had very low specificities ($p_{spec} < 1\times10^{-8}$), showing that they were not useful as tags in a mark-recapture setting. There were 17 tags matching exactly one male in the DNA-register. A histogram of the specificities for all 288 tags is shown in Fig. 2. Sixty-eight tags had specificity higher than 0.9, and 127 tags were in the range 0.1-0.9. For calculation of p_{spec}, the value N_m = 36,000 was used. This number was derived from an abundance estimate of 107,200 minke whales (Skaug *et al.*, 2004), and the assumption that reproductively active males constitute 1/3 of these.

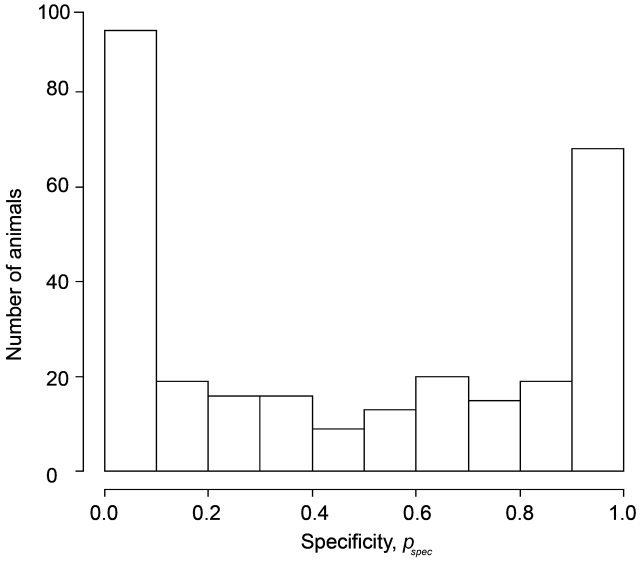

Fig. 2. Frequency distribution of tag-specificities for genetic tags obtained from the 288 minke whale mother-offspring pairs.

Table 2

DNA-profiles of a likely foetus-mother-father triplet. Alleles shared by both mother and offspring are shown in bold, while alleles shared by father and offspring are shaded grey. For locus GT211 it is not possible to infer which of the two alleles is inherited from the father. The column 'ID' gives the identification number used internally by the DNA-register.

	ID	GATA098	GT509	EV1	EV37	GT310	GT211	GT575	GT023	GATA028	GATA417
Foetus		95 **95**	**201** 205	**153** 157	201 **203**	**117** 121	104 106	154 **156**	**99** 99	**161** 206	**220** 236
Mother	104030	87 **95**	**201** 201	141 **153**	203 **203**	115 **117**	104 106	156 **156**	**99** 103	**161** 161	213 **220**
Father	201593	95 95	205 211	157 157	201 201	117 121	104 106	152 154	99 99	183 206	232 236

Four mother-father-offspring triplets had a paternity probability larger than 0.5 (Table 3). For each of these the positions and time points of capture were extracted from the database (Table 3). The male and female capture locations for the three triplets with a probability higher than 0.8 are shown in Fig. 1.

The log-likelihood as a function of N_m is shown in Fig. 3. The maximum likelihood estimate is $\hat{N}_m$ = 38,400, and the effective number of recaptures is m_e = 5. The lower bound in a 95% confidence interval is 13,000. The upper bound of the confidence interval is in practice infinity.

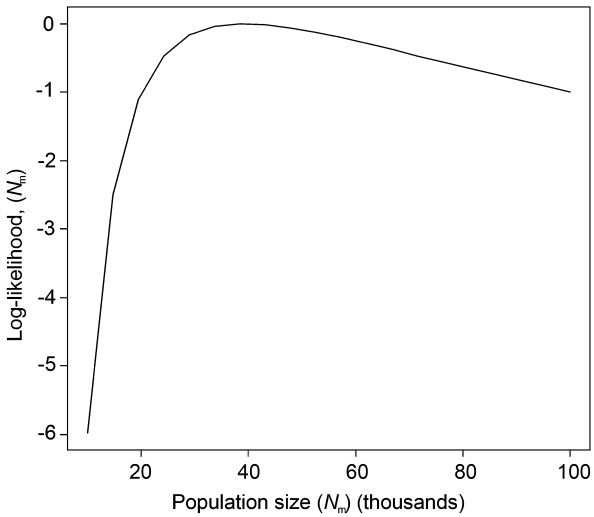

Fig. 3. Normalised log-likelihood as a function of male population size N_m (in thousands).

DISCUSSION

In Fig. 1, only mother-father-offspring triplets with paternity probabilities larger than 0.8 were plotted. The fourth triplet listed in Table 3 was excluded since its paternity probability was close to 0.5 and thus deemed not a 'certain' case. Nevertheless, paternity probabilities below 0.8 have various uses (Jones and Arden, 2004), one of them being abundance estimation.

In human genetics it has been estimated that the error rate in large-scale microsatellite screens of the type underlying the minke whale DNA-register is 0.25-2% per locus (Ewen *et al.*, 2000). Comparison of maternal and foetal DNA-profiles in the present study indicated that the error rate in the minke whale DNA-register is in the same range. Inconsistencies between maternal and foetal profiles discovered during this process were subsequently resolved by the genetics laboratory. However, errors in the unchecked half of the foetal profile, together with errors in the paternal profile when contained in the DNA-register, would cause the paternity probability to become zero, and hence lead to erroneous exclusion of the father. A simple sensitivity study was conducted to show that none of the excluded fathers in the database were likely to be excluded due to a typing error.

There is no evidence of monogamy in common minke whales, so it is very unlikely that the DNA-register would contain any full siblings of a given foetus. Half-siblings can be present, however, and we thus calculated the probability that a half sibling (same father as the foetus) is consistent with the inferred part of the father-profile by chance. This probability was found to be low (0.008), showing that half-siblings are unlikely to have caused problems in the study.

Based on the tagging programme carried out in the period 1974-1978, the Northeast Atlantic stock of common minke whales was estimated to be in the range 81,500 to 121,000 (Beddington *et al.*, 1984). More recent abundance estimates have been based on line transect methodology, and have given numbers in the same range (Skaug *et al.*, 2004). The estimate N_m = 38,400 for the reproductively active male population obtained in the present study is consistent with previous abundance estimates, but as it is based on few 'recaptures', the associated uncertainty is large.

Mark-recapture methods rely on the assumption that all individuals have the same probability of being marked, as well as having the same probability of being recaptured on later occasions. If there are differences in reproductive success among males the first part of this assumption will be violated for the present method, because 'super-breeders' will have an increased chance of being tagged. There is no external information available on variation in reproductive success in common minke whales. Note that breeding opportunities prior to 2000 (the first year of collection of foetal samples) are not relevant in this context, and hence difference in age among males does not introduce heterogeneity in the marking probability. When recaptures are being recorded over a long period relative to the life span of the species, it may be necessary to apply mark-recapture estimators appropriate for open populations. As the recapture period in this study was only three years, this was not a concern.

A further assumption of mark-recapture methods is that of a single population, which translates into the assumption that northeastern Atlantic common minke whales are panmictic (i.e. constitute a single breeding unit). The fact that mtDNA markers for males and microsatellite markers (both males and females) have uniform haplotype frequencies across sub-areas indicate that there is only a single breeding population (IWC, 2004). The fact that there are significant differences in mtDNA haplotype frequencies between the central and eastern part of the North Atlantic for females, are likely to be caused by a learning process where the calves follow their mothers, and does not constitute evidence against the assumption of a single breeding stock.

Table 3

The four most likely foetus-mother-father triplets, as measured by the paternity probability (rightmost column). Information about time and position of capture, together with zoological length, is also given. The first column provides a link to Fig. 1 which displays capture positions. The column 'SMA' shows the corresponding IWC Small Area (Fig. 1).

	Female and offspring					Male				
Fig. 1	Date	Map reference	SMA	Length (cm)		Date	Map reference	SMA	Length (cm)	$P_i(j)$
				Moth.	Fet.					
1	10.05.01	56°59'N, 06°04'E	EN	775	14	11.08.02	63°50'N, 06°37'E	EW	864	1.00
2	30.05.01	71°35'N, 28°17'E	EB	785	44	08.06.02	63°10'N, 03°00'E	EW	770	0.99
3	28.05.00	79°55'N, 08°18'E	ES	810	48	18.07.01	70°46'N, 20°53'E	EW	820	0.87
	23.06.01	57°15'N, 04°34'E	EN	836	62	04.08.01	64°14'N, 05°44'E	EW	734	0.58

It is not known where breeding and calving take place. Studies of North Atlantic foetal growth data indicate a prolonged mating season with conceptions occurring December-May, with February as the peak month (Horwood, 1990). Hence, the whaling operations in one year end (in September) before next year's reproduction begins. Gestation has been calculated to last about 10 months and thus calves are born over the period October to March with a peak in December in the North Atlantic. None of the probable fathers listed in Table 3 were caught prior to the seasonal catch date of the female. This may be in accordance with the general belief inferred from catch statistics that mature females migrate into the summer feeding areas earlier than males (Øien, 1988).

Based on the observations from several studies of minke whale foetal growth rates, it has been found that the foetus has an average growth of approximately 1cm day^{-1} (Horwood, 1990). With this assumption all the foetuses in Table 3 must have been conceived in April, probably the latter half of the month. Although information on travel speeds of minke whales are sparse, results from satellite tracking in Norwegian waters indicate travelling distances of the order 50-80km day^{-1} (Heide-Jorgensen *et al.*, 2001). This means that the female caught in the Skagerrak part of the North Sea on 10 May 2001 with a foetus of length 14cm must have conceived within the North Sea area, and consequently, the father must also have been present in the North Sea area in this period. Combining this with the fact that the father was caught in the Norwegian Sea the year after, there are two locations in space and time for the father. This type of multiple recapture data provides a means of studying site fidelity in minke whales.

The links as given in Table 3 also give insight to reproductive parameters. Studies of Northeast Atlantic minke whales have given estimated lengths at sexual maturity of 7.15m for females and 6.75m in males. Both females and males listed in Table 3 show lengths well above these, namely 7.75m-8.36m and 7.34m-8.64m for females and males, respectively.

To verify the hypothesised relationships in Table 3, particularly the last two, one could type the involved individuals at a number of additional loci. This could also be done for putative fathers with probabilities lower than 0.5 (not shown in Table 3), and would be a way of partly getting around the fact that a large proportion of the tags have low specificities (Fig. 2). Such a two-stage approach is both time and cost efficient compared to an approach where all the males in the DNA-register are typed at additional loci.

In conclusion, the various genetic tagging methods that exist have a large potential to yield important new information about cetacean demography. This is especially true for common minke whales, which are difficult to study by other means. From a management point of view, it seems necessary to monitor the level of relatedness in catches for harvested whale populations. Even if laboratory costs should prevent DNA-profiles being established for the full catch, it is vital that tissue samples are taken from all individuals, and stored for future analysis.

ACKNOWLEDGEMENTS

We are grateful to Charlotta Lindblom at IMR for handling the tissue samples, and to the Norwegian Directory of Fisheries for providing funding for the study.

REFERENCES

Amos, B., Schlötterer, C. and Tautz, D. 1993. Social structure of pilot whales revealed by analytical DNA profiling. *Science* 260(5108):670-2.

Beddington, J.R., Cooke, J.G., Christensen, I., Øritsland, T., Oien, N. and Rorvik, C.J. 1984. Assessments of the northeast Atlantic stock of minke whales. *Rep. int. Whal. Commn* 34:285-91.

Christensen, I. and Rorvik, C.J. 1978. Preliminary results from recent markings of minke whales in the North East Atlantic. *Rep. int. Whal. Commn* 28:183-4.

Clapham, P.J. and Palsbøll, P.J. 1997. Molecular analysis of paternity shows promiscuous mating in female humpback whales (*Megaptera novaenaglie*, Borowski). *Proc. R. Soc. Lond. Ser. B.* 264(1378):95-8.

Dupuy, B.M. and Olaisen, B. 1998. Typing procedure for the Norwegian minke whale DNA register. Department of Forensic Medicine, University of Oslo, Norway. 32pp.

Ewen, K.R., Bahlo, M., Treloar, S.A., Levinson, D.F., Mowry, B., Barlow, J.W. and Foote, S.J. 2000. Identification and analysis of error types in high-throughput genotyping. *Am. J. Hum. Genet.* 67(3):727-36.

Garrigue, C., Dodemont, R., Steel, D. and Baker, C.S. 2003. Organismal and 'gametic' capture-recapture using microsatellite genotyping confirm demographic closure and reproductive autonomy of a humpback whales wintering ground. *Mar. Ecol. Prog. Ser.* 274:251-62.

Heide-Jørgensen, M.P., Nordøy, E.S., Øien, N., Folkow, L.P., Kleivane, L., Blix, A.S., Jensen, M.V. and Laidre, K.L. 2001. Satellite tracking of minke whales (*Balaenoptera acutorostrata*) off the coast of northern Norway. *J. Cetacean Res. Manage.* 3(2):175-8.

Horwood, J.W. 1990. *Biology and Exploitation of the Minke Whale*. CRC Press, Boca Raton. 238pp.

International Whaling Commission. 2004. Report of the Scientific Committee. Annex D. Report of the sub-committee on the revised mamagement procedure. Appendix 14. Report of the working group on North Atlantic minke whales RMP *Implementation Review*. *J. Cetacean Res. Manage. (Suppl.)* 6:171-9.

Jones, A.G. and Arden, W.R. 2004. Methods of parentage analysis in natural populations. *Mol. Ecol.* 12(10):2511-23.

Jonsgård, Å. 1951. Studies on the little piked whale or minke whale (*Balaenoptera acutorostrata* Lacépède). Report on Norwegian investigations carried out in the years 1943-1950. *Norsk Hvalfangsttid.* 40:209-32.

Jonsgård, Å. 1955. Development of the modern Norwegian small whale industry. *Norsk Hvalfangsttid.* 44(12):697-718.

Marshall, T.C., Slate, J., Kruuk, L.E.B. and Pemberton, J.M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* 7(5):639-55.

Nielsen, R., Mattila, D.K., Clapham, P.J. and Palsbøll, P.J. 2001. Statistical approaches to paternity analysis in natural populations and applications to the North Atlantic humpback whale. *Genetics* 157(4):1673-82.

Øien, N. 1988. Length distributions in catches from the northeastern Atlantic stock of minke whales. *Rep. int. Whal. Commn* 38:289-95.

Olaisen, B. 1997. Proposed specifications for a Norwegian DNA database register for minke whales. Paper SC/49/NA1 presented to the IWC Scientific Committee, September 1997, Bournemouth (unpublished). 7pp. [Paper available from the Office of this Journal].

Skaug, H.J., Øien, N., Schweder, T. and Bøthun, G. 2004. Abundance of minke whales (*Balaenoptera acutorostrata*) in the northeastern Atlantic; variability in time and space. *Can. J. Fish. Aquat. Sci.* 61(6):870-86.

Date Received: November 2004
Date Accepted: April 2005