Technical Reports of the Institute of Cetacean Research (2020) pp. 35–40

Technical Report (not peer reviewed)

Genetic tagging technique: basic concept and a case study by the Institute of Cetacean Research

Mioko TAGUCHI*

Institute of Cetacean Research, 4–5 Toyomi-cho, Chuo-ku, Tokyo 104–0055, Japan

**Contact e-mail: taguchi@cetacean.jp*

ABSTRACT

The studies on biological conservation require contemporary demographic estimates, e.g., migration rate, effective population size and stock structure. Although these estimates can be obtained through standard population genetics analyses, such studies could fail to infer them because of recent evolutionary and biogeographical events. Alternative approaches to obtain such parameters include individual identification and kinship inferences using genetic data, also called 'genetic tagging', which has become a common technique in the fields of ecology and conservation during the last several decades. This paper presents the basic concept of genetic individual identification and kinship inference. It subsequently outlines the parentage analysis for the western North Pacific common minke whale conducted by the Institute of Cetacean Research as a case study.

INTRODUCTION

The studies on biological conservation in animal species require information on contemporary population parameters, e.g., effective population size, migration rate and population structure. The estimates of these parameters can be obtained through traditional population genetics analyses. However, these analyses often fail to infer them in cases where the populations have diverged recently, or if changes in gene flow and effective population size have occurred. Therefore, such estimates have been mainly obtained by individual-based tagging research using conventional tagging or photo-identification (e.g., Calambokidis *et al.*, 2001; Johnson *et al.*, 2009; Mizroch and Rice, 2013; Urian *et al.*, 2015).

Individual animals can also be identified using genetic data, generally called 'genetic tagging' (Palsbøll, 1999). This has become a common technique to obtain the contemporary estimates in the ecological and conservation fields during the last several decades (e.g., humpback whales, *Megaptera novaeangliae*, Palsbøll *et al.*, 1997; North Pacific right whale*s, Eubalaena japonica*, Wade *et al.*, 2011). In addition, in recent years, the extended genetic tagging incorporating information on kinship among individuals is being used to investigate contemporary population dynamics (e.g., Kanda *et al.*, 2014; Bravington *et al.*, 2016; Ohashi *et al.*, *in-press*). The methods for abundance estimates based on two types of genetic tags (individual identification and kinship) are summarized by Takahashi (this issue).

This paper presents the basic concept of genetic individual identification and kinship inference, with a brief summary of a case study by the Institute of Cetacean Research (ICR), which is the parentage analysis for the North Pacific common minke whale (*Balaenoptera acutorostrata*).

BASIC CONCEPT OF GENETIC TAGGING TECHNIQUES

Individual identification

Sexual reproduction guarantees that, in the main, each individual has a unique genotype. There are some exceptions, however, such as identical twins. This characteristic allows each individual to be tagged and, as a consequence, it can be determined if two genetic samples come from the same individual. Figure 1 shows an example of individual identification using microsatellite genotype data. In this example, the four samples are derived from three whales. Samples A and B have the same genotypes and therefore the two samples are assumed from the same whale, while samples C and D have different genotypes suggesting they are from different whales.

Probability of identity

The probability of identity (*I*) is the probability that two unrelated individuals in the population share the same genotype. This estimate requires information on allele frequencies in the population, because common alleles

Sample	Locus 1	Locus 2	Locus 3	Locus 4	Matching DNA profile
A	120120	210214	099107	264264	B
B	120120	210214	099107	264264	A
C	120122	210210	101103	264267	No
D	122126	214218	099103	264264	No

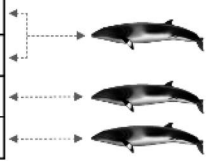


Figure 1. Example data involving four microsatellite loci genotyped for each of four whales. Genotypes are represented by 6-digit codes, and each allele is coded by 3 digits. The results of individual identification are shown in the column 'Matching DNA profile'.

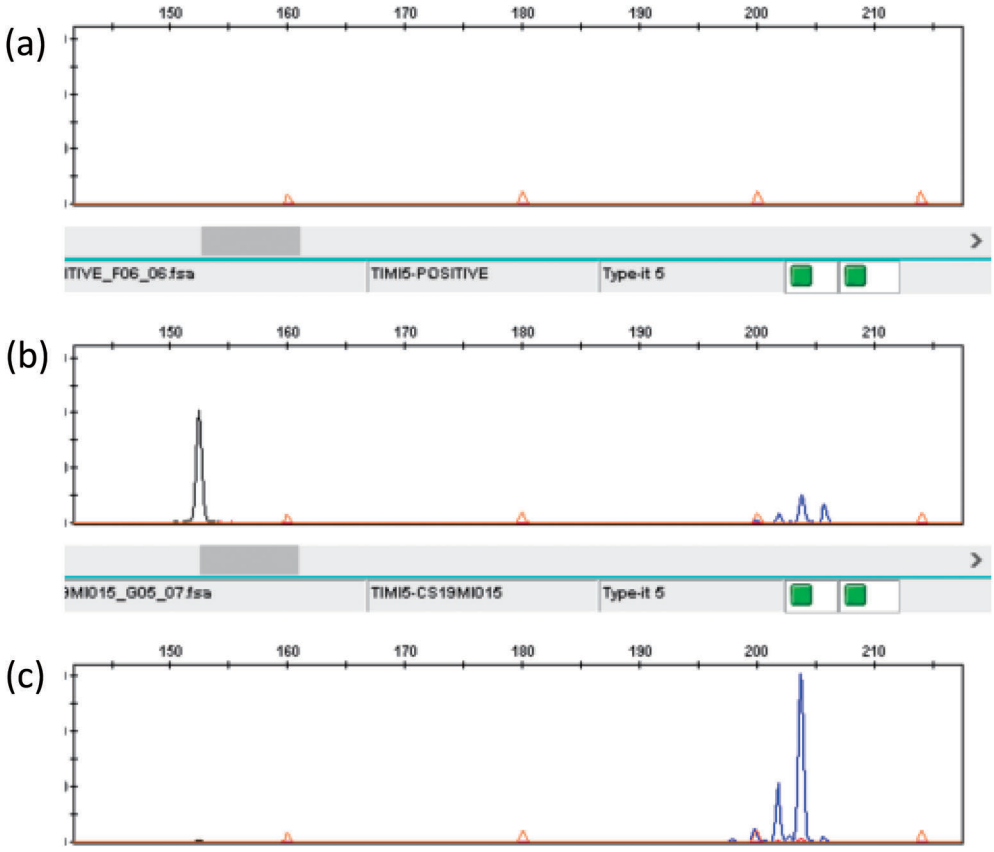


Figure 2. Example of sex determination in common minke whales through multiplex fluorescent PCR using the *SRY* locus (black) located on the Y chromosome and a microsatellite locus (blue) as the internal control: (a) failure to amplify PCR fragments; (b) male; (c) female.

are much more likely to be shared than rare ones. This can be estimated for a population with random mating according to the formula derived by Paetkau and Strobeck (1994):

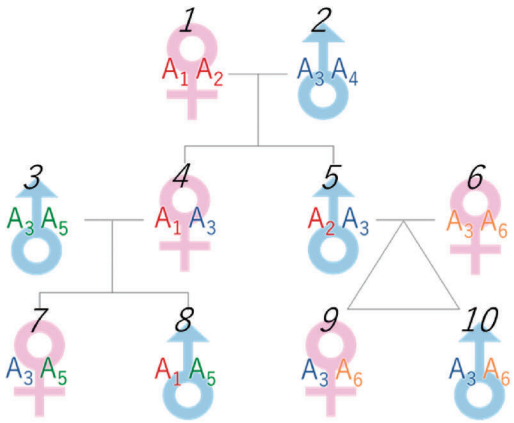
$$\hat{I} = \prod_k \left(\sum_i p_i^4 + \sum_i \sum_{j>i} (2 p_i p_j)^2 \right)$$

where, p_i and p_j are the frequencies of the ith and jth alleles at the kth locus in the population. In practice, it is desirable to aim for a probability of shared genotype identity between any two individuals (I) of <0.05.

Molecular sex determination

Sex information is useful not only to assist in individual identification but also for a biological interpretation of the results of analyses.

The PCR-based technique has been proposed for sex determination using the presence or absence of genes on the sex chromosomes. The testis-determining *SRY* genes are male-specific in mammals and can be detected by PCR amplification using specific primers. The ICR genetic team developed the primer set that amplifies the *SRY* gene to determine the sex of cetaceans, using a microsatellite locus as the internal control (Abe *et al.*, 2001).



Kinship category	Pairs of individuals	Φ	κ_0	κ_1	κ_2
Identical / Monozygotic twins	9-10	0.50	0	0	1.00
Parent-offspring	1,2-4,5 / 3,4-7,8 / 5,6-9,10	0.25	0	1.00	0
Full siblings	4-5 / 7-8	0.25	0.25	0.50	0.25
Half siblings	1,2-7,8,9,10	0.125	0.50	0.50	0.00
First cousins	7,8-9,10	0.0625	0.75	0.25	0
Unrelated	3-6	0	1.00	0	0

Figure 3. Example of pedigree containing ten non-inbred individuals. The figure shows the kinship categories and the values for three IBD coefficients (κ_0, κ_1 and κ_2). The three IBD coefficients can be also summarized by the kinship coefficients (Φ) which is the probability that a random allele from one individual is IBD to a random allele from the other. The Φ is also equivalent to the inbreeding coefficient of their offspring.

Figure 2 shows an example of genetic sex determination using this method. In case of male, a PCR fragment of *SRY* gene is observed between 150 and 155bps (Figure 2b), while the fragment is absent in case of female or PCR failure. We can distinguish female (Figure 2c) from PCR failure (Figure 2a) by the presence of the PCR fragment of one microsatellite locus as an internal control.

Kinship inference

Individuals that are biologically related share genes that are identical-by-descent (IBD), i.e., identical copies of a gene segregating from a common ancestor within the defined pedigree. IBD is a fundamental concept that underlies kinship inference (Thompson, 2013). Here, the basic concept of the IBD coefficient is explained, which provides an introduction to the ideas of IBD. It is followed by the simple principle of parentage analysis which is a unique application of kinship inferences.

IBD coefficients

Two non-inbred diploid individuals can share none, one, or two alleles IBD. In case of the pedigree shown in Figure 3, full sibling individuals *4* and *5* share one allele IBD (A_3 which is inherited from their father) while unrelated individuals *3* and *6* share no allele IBD. The monozygotic twins *9* and *10* share two alleles IBD (A_3 and A_6, which are inherited from father and mother, respectively). The probabilities of each of the events are summarized by the IBD coefficients, i.e., κ_0, κ_1 and κ_2 (Cockerham, 1971), which are shown in Figure 3.

Table 1

Patterns of allele sharing IBD for a full sibling pair (individuals *4* and *5* in Figure 3).

Possible genotypes	A_1A_3	A_1A_4	A_2A_3	A_2A_4
A_1A_3	2	1	1	0
A_1A_4	1	2	0	1
A_2A_3	1	0	2	1
A_2A_4	0	1	1	2

Table 1 shows an example on how to determine the IBD coefficients for a category of full siblings. Consider that the genotypes of the parents at a locus are A_1A_2 and A_3A_4 (individuals *1* and *2* in Figure 3). Each offspring can have one of the four following genotypes: A_1A_3, A_1A_4, A_2A_3 or A_2A_4 (individuals *4* and *5* in Figure 3). Out of the sixteen ways to pair two offsprings, the dyad can share 2 alleles that are IBD in four ways, 1 allele in eight ways and 0 alleles in four ways. Thus, the IBD coefficients, κ_0, κ_1 and κ_2, for full siblings are 0.25 (4/16), 0.50 (8/16) and 0.25 (4/16), respectively. The IBD coefficients for other kinship categories can be calculated in the same manner.

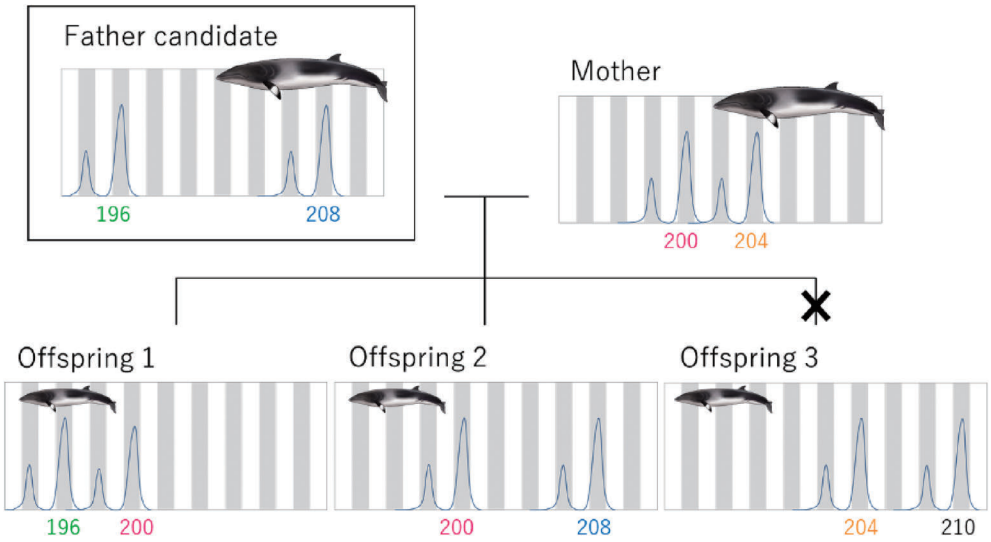


Figure 4. Principle of parentage analysis using the exclusion approach. The numbers under the peaks are the size of PCR fragment in base pairs which can be interpreted as alleles.

Parentage analyses

Parentage analysis is a unique application in which the most likely parents of a target offspring are sought amongst eligible candidates. The basic concepts of the parentage analysis are grouped into six categories according to Jones *et al.* (2010): (1) Exclusion, (2) Categorical Allocation, (3) Fractional Allocation, (4) Full Probability Parentage Analysis, (5) Parental Reconstruction, (6) Sibship Reconstruction. In this section, categories (1) (the conceptually simplest approach) and (2) (the most commonly used approach), are described.

Exclusion approach

This approach uses rules of Mendelian inheritance for diploid organisms wherein an offspring inherits one of two alleles at each locus from each of its parents. A simple exercise for the exclusion approach using a single microsatellite locus is shown in Figure 4.

In this exercise, the mother shows the genotype of 200/204, and the genotypes of her three offspring 1, 2 and 3 were 196/200, 200/208 and 204/210, respectively. Here, a male with genotype of 196/208 can be excluded as a father candidate for offspring 3 since no allele from the male is observed in offspring 3. On the other hand, the male is still a father candidate for offspring 1 because the alleles 196 and 200 can be found from the male and mother, respectively. The same is true for offspring 2.

This approach is powerful when there are few candidate parents and multiple highly polymorphic markers are available. In practice, most of the studies based on the exclusion approach actually require at least two mismatching loci between the candidate and the offspring to account for typing errors or mutations.

Categorical Allocation approach

As in the case of exclusion method, this approach requires at least one focal offspring and a set of candidate parents. The Categorical Allocation approach was developed to resolve situations in which complete exclusion may not be feasible (Meagher and Thompson, 1986). The main benefit of this approach is to choose the single most likely parent from a group of non-excluded putative parents. The logic stems from the observation that different genotypes of parent may differ in their probability of having produced the genotype of the focal offspring (Meagher and Thompson, 1986). This approach also has the advantage of handling genotyping errors or mutations.

Currently, most of the Categorical Allocation approaches use a likelihood approach (e.g., Marshall *et al.*, 1998). However, a Bayesian approach (e.g., Nielsen *et al.*, 2001) can also be used (see also Takahashi, this issue).

Genotyping errors and their effects

Microsatellite DNA marker is one of the most common genetic markers not only for individual identification and kinship analyses, but also for the standard analyses on population genetics. However, microsatellite DNA is known as the error-prone marker in its genotyping, which is recognized to have a serious impact on genetic individual identification as well as kinship inferences (reviewed in Hoffman and Amos, 2005).

The genotyping of a microsatellite allele depends on the microsatellite profile, and requires strict rules to be defined in advance. Typical microsatellite profiles are

characterized by a succession of peaks with growing intensity due to stutter band. Researchers often suffer from several types of ambiguous microsatellite profiles in genotyping, e.g., null alleles, overlapping of stutter bands, split peak, large allele dropout or short allele dominance and false alleles. Some factors are known to be involved in the ambiguous profiles, e.g., variation in DNA sequences, low quality or quantity of template DNA, biochemical artifact and human error (Hoffman and Amos, 2005).

Figure 5 illustrates two types of genotyping error that are likely to induce false paternity and consequently bias the biological conclusions. Consider in the example of Figure 5 that the male is the real father of the offspring. The real genotype for male is 196/208 without error: the offspring inherited allele 200 from his mother and allele 208 from his father. In the cases with genotyping errors, the male would incorrectly be excluded as the father of the offspring because allele 208 is mistyped as allele 202.

In order to avoid this type of incorrect paternity exclusion, it is better to allow one or several genetic mismatches in the parentage analysis depending on the probability of identity and the calculated error rate.

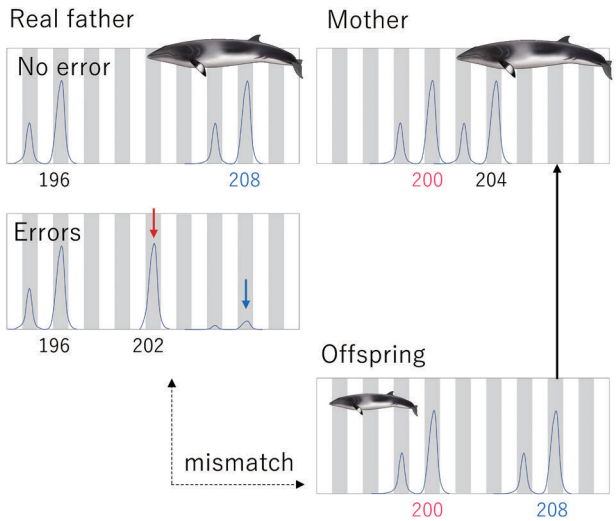


Figure 5. Simple example of parentage analysis by the Exclusion approach. Arrows show the genotyping errors due to false allele (red) and large allele dropout (blue).

A CASE STUDY ON PARENTAGE BY THE INSTITUTE OF CETACEAN RESEARCH

Parent-offspring inference in North Pacific common minke whales

The parent-offspring pair analysis was carried out based on a type of Categorical Allocation approach, using a total of 4,707 common minke whales in the western North Pacific (see Tiedemann *et al.* (2017) for the details of the analytical procedures). Each sample was genotyped at 16 microsatellite loci. The mitochondrial DNA control region haplotypes as well as biological information, e.g., sampling date and position, sex, sexual maturity and body length, were used, if available, to assist the interpretation of the results.

The parent-offspring analyses inferred a total of 40 and 13 parent-offspring pairs for the O and J stock (Goto *et al.*, 2017), respectively . The forty parent-offspring pairs of O stocks were widely distributed through the Pacific side of Japan (Figure 6). A total of 17 pairs of O stock animals were found between coastal and offshore waters. In most of the pairs, the offspring were found near the coast while their parents were in offshore waters. This is consistent with the pattern of migration with sexual segregation described previously for western North Pacific common minke whales (Hatanaka and Miyashita, 1997).

The thirteen parent-offspring pairs of J stocks were distributed within and between the Sea of Japan and the Pacific side of Japan (Figure 6). Four of them were found between the Sea of Japan side and the Pacific side. The

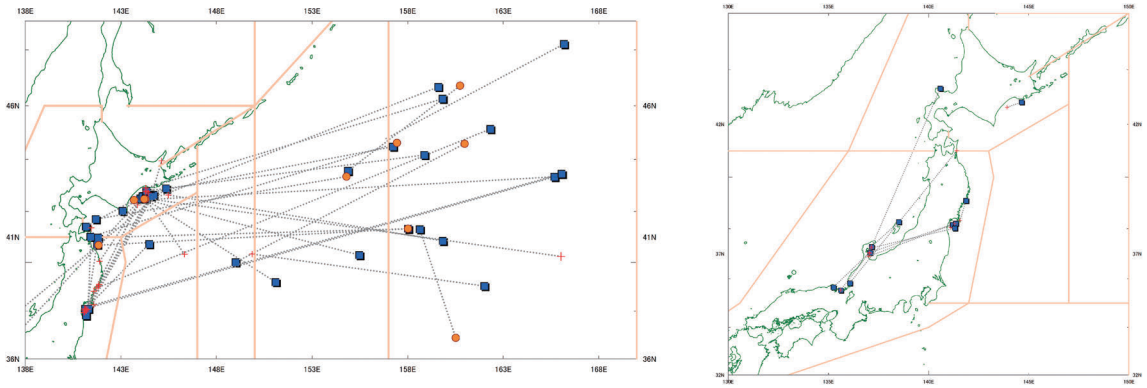


Figure 6. Distribution of parent-offspring pairs of O (left) and J (right) stocks in western North Pacific common minke whales. Blue square: parent; orange circle: mature offspring; red cross: immature or unknown maturity stage offspring.

results implied that the J stock individuals occurring in the Sea of Japan and the Pacific side of Japan were derived from the same stock.

Future work

The development of the new genetic marker which is less error-prone, that is, Single Nucleotide Polymorphisms (SNPs) marker is on-going at the ICR. The ICR genetic team will perform analyses using SNPs genotypes, not only for western North Pacific common minke whales, but also for other whale species such as Bryde's, fin, right and blue whales. The results of these analyses will aid in obtaining the contemporary demographic estimates of these whale species using genetic data.

REFERENCES

Abe, H., Goto, M., Pastene, L.A., Dewa, K. and Naito, E. 2001. Practical use of multiplex fluorescent PCR for cetacean sex identification. *Mar. Mam. Sci.* 17(3): 657–664.

Bravington, M.V., Skaug, H.J. and Anderson, E.C. 2016. Close-kin mark-recapture. *Statist. Sci.* 31(2): 259–274.

Calambokidis, J., Steiger, G.H., Straley, J.M., Herman, L.M., Cerchio, S., Salden, D.R., Jorge, U.R., Jacobsen, J.K., Ziegesar, O.V., Balcomb, K.C., Gabriele, C.M., Dahlheim, M.E., Uchida, S., Ellis, G., Mlyamura, Y., de guevara Paloma Ladrón, P., Yamaguchi, M., Sato, F., Mizroch, S.A., Schlender, L., Rasmussen, K., Barlow, J. and Ii, T.J.Q. 2001. Movements and population structure of humpback whales in the North Pacific. *Mar. Mam. Sci.* 17: 769–794.

Cockerham, C.C. 1971. Higher order probability functions of identity of alleles by descent. *Genetics* 69(2): 235–246.

Goto, M., Taguchi, M. and Pastene, L.A. 2017. Distribution and movement of 'O' and 'J' stock common minke whales in waters around Japan based on genetic assignment methods. *Technical Reports of the Institute of Cetacean Research (TEREP-ICR)* No.1: 37–43.

Hatanaka, H. and Miyashita, T. 1997. On the feeding migration of the Okhotsk Sea-West Pacific stock of minke whales, estimates based on length composition data. *Rep. int. Whal. Commn* 47: 557–564.

Hoffman, J.I. and Amos, W. 2005. Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Mol. Ecol.* 14: 599–612.

Johnson, M., Aguilar, de Soto N. and Madsen, P.T. 2009. Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: a review. *Mar. Ecol. Prog. Ser.* 395: 55–73.

Jones, A.G., Small, C.M., Paczolt, K.A. and Ratterman, N.L. 2010. A practical guide to methods of parentage analysis. *Mol. Ecol. Resour.* 10: 6–30.

Kanda, N., Goto, M. and Pastene, L.A. 2014. Paternity analysis on Antarctic minke whales using JARPA and JARPAII samples. Paper SC/F14/J7 presented to the IWC/SC Review Workshop of the Japanese Whale Research Program under Special Permit in the Antarctic-Phase II (JARPAII), February 2014 (unpublished). 4 pp. [Available from the IWC Secretariat].

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: 639–655.

Meagher, T.R. and Thompson, E. 1986. The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. *Theor. Popul. Biol.* 29(1): 87–106.

Mizroch, S.A. and Rice, D.W. 2013. Ocean nomads: Distribution and movements of sperm whales in the North Pacific shown by whaling data and Discovery marks. *Mar. Mam. Sci.* 29: E136–E165.

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: 1673–1682.

Ohashi, Y., Goto, M., Taguchi, M., Pastene, L.A. and Kitakado, T. Evaluation of a paternity method based on microsatellite DNA genotypes for estimating the abundance of the Antarctic minke whales, *Balaenoptera bonaerensis*, in the Indo-Pacific region of the Antarctic. *Cetacean Population Studies (in-press)*.

Palsbøll, P.J. 1999. Genetic tagging: contemporary molecular ecology. *Biol. J. Linn. Soc. Lond.* 68: 3–22.

Palsbøll, 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. 1997. Genetic tagging of humpback whales. *Nature* 388(6644): 767–769.

Paetkau, D. and Strobeck, C. 1994. Microsatellite analysis of genetic variation in black bear populations. *Mol. Ecol.* 3: 489–495.

Thompson, E.A. 2013. Identity by Descent: Variation in Meiosis, Across Genomes, and in Populations. *Genetics* 194(2): 301–326.

Tiedemann, R., Tiedemann, M.R., Goto, M., Taguchi, M. and Pastene, L.A. 2017. Finding parent-offspring pairs among western North Pacific common minke whales. Paper SC/67a/SDDNA1 presented to the IWC Scientific Committee, May 2017 (unpublished). 18 pp. [Available from the IWC Secretariat].

Urian, K., Gorgone, A., Read, A., Balmer, B., Wells, R.S., Berggren, P., Durban, J., Eguchi, T., Rayment, W. and Hammond, P.S. 2015. Recommendations for photo-identification methods used in capture-recapture models with cetaceans. *Mar. Mam. Sci.* 31: 298–321.

Wade, P.R., Kennedy, A., LeDuc, R., Barlow, J., Carretta, J., Shelden, K., Perryman, W., Pitman, R., Robertson, K., Rone, B., Salinas, J.C., Zerbini, A., Brownell, R.L. and Clapham, P.J. 2011. The world's smallest whale population? *Biol. Lett.* 7: 83–85.