



*Molecular genetics in animal ecology*. Edited by P. A. Racey,  
P. J. Bacon, J. F. Dallas and S. B. Piertney

## Genetic tagging: contemporary molecular ecology

PER J. PALSBØLL\*

*Unit of Evolutionary Genetics, Department of Molecular Biology, Free University of Brussels, CP 244 Boulevard du Triomphe, B-1050 Brussels, Belgium; and School of Biological Sciences, University of Wales Bangor, Deiniol Road, Bangor, Gwynedd LL57 2UW*

Population genetic analyses have been highly successful in deciphering inter- and intra-specific evolutionary relationships, levels of gene flow, genetic divergence and effective population sizes. Parameters estimated by traditional population genetic analyses are evolutionary averages and thus not necessarily relevant for contemporary ecological or conservation issues. Molecular data can, however, also provide insight into contemporary patterns of divergence, population size and gene flow when a sufficient number of variable loci are analysed to focus subsequent data analyses on individuals rather than populations. Genetic tagging of individuals is an example of such individual-based approaches and recent studies have shown it to be a viable alternative to traditional tagging methods. Owing to the ubiquitous presence of hyper-variable DNA sequences in eukaryote genomes it is *in principle* possible to tag any eukaryote species and the required DNA can be obtained indirectly from substrates such as faeces, sloughed skin and hair. The purpose of this paper is to present the concept of genetic tagging and to further advocate the extension of individual-based genetic analyses beyond the identification of individuals to other kinds of relationships, such as parent-offspring relations, which more fully exploit the genetic nature of the data.

© 1999 The Linnean Society of London

Keywords:—microsatellite – individual identification – parent-offspring detection – population genetics – Cetaceae.

### CONTENTS

Introduction . . . . .	4
Genetic tagging of individuals . . . . .	5
Genetic and conventional tagging techniques . . . . .	5
Different methods for genetic individual identification . . . . .	9
Genetic tagging of North Atlantic humpback whales . . . . .	10
Indirect sampling of black and brown bears for genetic tagging . . . . .	11
Identification of parent-offspring relations . . . . .	12
Obscure gene flow between pilot whale pods . . . . .	13
Identification of parent-offspring relations . . . . .	13
How many loci are necessary for reliable detection of parent-offspring relations? . . . . .	14
Inferring contemporary population structure from parent-offspring relations . . . . .	18
Conclusions . . . . .	18
Acknowledgements . . . . .	19
References . . . . .	19

\* Correspondence to University of Wales Bangor. Email: [p.palsboll@bangor.ac.uk](mailto:p.palsboll@bangor.ac.uk)

## INTRODUCTION

Estimation of genetic divergence and gene flow among sub-populations and effective population sizes is central to molecular ecology. Such parameters are typically estimated from one or several loci either from haplotype counts only (e.g. Weir & Cockerham, 1984) or from haplotype divergence and haplotype frequencies (e.g. Hudson *et al.*, 1992). Population divergence, gene flow and effective population size estimated in this manner are evolutionary averages, which may not equal the contemporary values of these parameters. The divergence between haplotype sequences and frequencies among populations are generated by mutation and genetic drift, each slow processes from a human perspective, and significant levels of genetic divergence among sub-populations are usually only obtained after many generations (Fig. 1) even for rapid evolving loci, such as microsatellites.

In contrast, the objectives of ecological and conservation-related research are usually to obtain contemporary estimates; by relying on traditional evolutionary approaches these may fail to detect, for instance, recent population divergence, or recent changes in gene flow and effective population size. Hence, contemporary estimates of migration, abundance and structure have usually been obtained by studying individuals, each identified by some sort of tag (e.g. Seber, 1982).

Individual animals can also be identified from genetic data, e.g. by the composite genotype at multiple polymorphic VNTR (variable number of tandem repeats) loci (Jeffreys, 1985). Such a genetic ‘tag’ is in principle similar to conventional tags and enables tracking of individuals in a ‘real-time mode’, i.e. on a temporal and spatial scale relevant to ecological and conservation issues (e.g. Palsbøll *et al.*, 1997; Taberlet *et al.*, 1997). Genetic tagging based upon the composite genotype at multiple microsatellite loci (VNTR loci with 1–5 nucleotide long repeats) is in principle

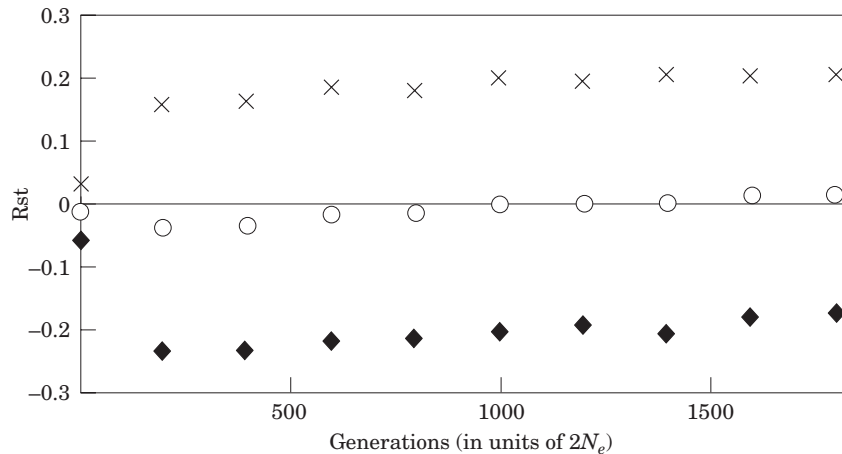


Figure 1. The rate of divergence between two equal size populations expressed as  $R_{ST}$ . Coalescence simulations were carried out under a single-step mutation model as described by Hudson (1990) and employed in Palsbøll (1999). For each data point 1000 coalescence simulations were conducted with sample sizes of 50 diploid individuals,  $4N_e\mu = 10$  (where  $N_e$  is the effective population size and  $\mu$  the number of mutations per generation) and six loci. The degree of divergence was estimated as  $R_{ST}$  as defined by Slatkin (1995). The lower (◆) and upper (×) 95% confidence limit were calculated from the variance. (○) Mean value of  $R_{ST}$ .

applicable to all eukaryotic organisms, as microsatellites appear to be a general feature of the eukaryotic genome (Tautz & Renz, 1984). The critical parameter for the feasibility of genetic tagging is the degree of polymorphism within a population, which in turn is the product of the effective population size and mutation rate. For example, populations with a small effective population size, e.g. due to a recent bottleneck, may contain insufficient variation at microsatellite loci to enable reliable individual identification (e.g. northern elephant seals; Hoelzel *et al.*, 1993).

Although only a few studies have yet been conducted employing genetic tagging, this approach has been shown to constitute a viable alternative to conventional methods and has been applied successfully in several cases (Palsbøll *et al.*, 1997; Taberlet *et al.*, 1997; Woods *et al.*, in press).

This change in the use and analysis of genetic data from populations of samples to individuals—which has been greatly facilitated by the discovery of microsatellite loci (Tautz, 1989; Weber & May, 1989)—opens novel opportunities to address contemporary patterns of population structure and level of gene flow (Bossart & Pashley Prowell, 1998). However, genetic tagging of individuals only makes limited use of the genetic content in each tag. The level of consanguinity among individuals is readily estimated from multi-locus genotypes, the accuracy of which is negatively correlated with the degree of relatedness. In such an analysis the OTU (operational taxonomic unit) becomes the individual as opposed to the population; estimates of gene flow, for instance, will include only a single or few generations.

The aim of this paper is to present the concept of genetic tagging illustrated by examples provided by two recent studies. In addition, the potential of extending genetic identification of individuals to parent-offspring relations is discussed in terms of obtaining contemporary estimates of divergence, gene flow and population size. A simple simulation approach to estimate the approximate number of loci necessary for reliable identification of parent-offspring relations among unrelated and half siblings, from a small preliminary data set, is presented as well.

The intent of this communication is *not* to review the application of microsatellite analysis in natural populations for the purpose of investigating social structure, mating strategies etc., nor conventional tagging methods, all of which have been reviewed elsewhere (e.g. Hammond *et al.*, 1990; Queller *et al.*, 1993; Schlötterer & Pemberton, 1994).

#### GENETIC TAGGING OF INDIVIDUALS

##### *Genetic and conventional tagging techniques*

A cornerstone in ecological research is the ability to identify and track the movements of individuals (e.g. Hammond *et al.*, 1990). Identification of individual animals in their natural environment relies either on man-made tags (e.g. Kaye, 1960), variations in natural markings (Pennycuik, 1978) or genetic markers (e.g. Palsbøll *et al.*, 1997; Taberlet *et al.*, 1997). Ideally, any individual identification technique should possess several basic characteristics, such as:

- (1) Universal applicability.
- (2) Tagging conducted remotely and preferably non-invasively.
- (3) No significant loss of tags over time.

- (4) Unambiguous identification of individuals.
- (5) Efficient and rapid matching of tags.

Conventional techniques of individual identification usually employ human applied tags or natural markings. Human applied tags consist of either attachment of man-made tags, e.g. bird banding (see Marion & Shamus, 1977), or branding (Clark, 1968). When human applied tags are employed, capture is usually a pre-requisite for marking either during the initial marking session or at both the marking and re-capture sessions. Although imposing a severe restraint to most studies, capture enables collection of additional information, such as physical and serological measurements or tracking of subsequent movements if radio transmitters are attached (e.g. narwhals; see Dietz & Heide-Jørgensen, 1995). In addition to the restraints imposed by the necessity of capture prior to marking is the fact that physical tags are not universally applicable although the available technology is constantly being refined (e.g. small tropical fish; see Beukers *et al.*, 1995). Although employed to a lesser extent, embedded tags can in most cases only be recovered by lethal re-capture. Such marking techniques are typically employed in studies of small and common species such as fish.

Individual identification by use of natural markings has the advantage that marking as well as recapture is conducted remotely, i.e. usually a photograph of the relevant part of the individual constitutes the marking/recapture event. Hence, the approach is non-invasive, which can be a deciding factor for studies aimed at endangered or threatened species. Naturally, the level of auxiliary information obtained is limited to what can be recorded remotely. The main obstacle to identification by natural markings is sufficient, accessible and lasting variation in natural markings to reliably identify individuals. Proper field conditions when documenting a sighting, e.g. by photography, is another considerable variable and the subsequent analysis of captures by matching photographs can be laborious and time consuming. In many instances matching is not categorical, such as matching numbered man-made tags, which can have implications in terms of matching reliability and has been shown to be correlated with the experience of the personnel as well as the quality of photographs (Carlson *et al.*, 1990). Attempts have been made to computerize the matching procedure (as with human fingerprints), the main obstacle being the three-dimensional nature of the object, often photographed from different angles and under different light conditions. Efficient software has been developed for species, which has obvious fix points guiding the alignment of images prior to matching (e.g. grey seals; Hiby & Lovell, 1990). The concept of human fingerprints has recently been employed to identify individual large cats by their tracks (Grigione *et al.*, 1999). Hence, individual animals can be identified and followed without observing the individual itself. Such an approach is naturally dependent on sufficient track size and individual variation in tracks as well as a substrate that leaves a clear footprint. The advantage is the relative ease with which tracks subsequently can be digitized and the matching effort computerized. Each foot provides a separate identification, and thus multiple independent 'tags' per individual.

A concern for conventional marking techniques as well as individual identification by natural markings is 'tag-loss'. Double-marking individuals can circumvent the problem, e.g. by documenting the scarring of the dorsal fin as well as the pigmentation pattern of the right side of the head on fin whales (Agler *et al.*, 1990).

Common for individual identification by human applied tags or natural markings is that the tag in itself contains little or no information beyond identification of the individual. The exception is some natural markings, such as pigmentation patterns, which can be considered a phenotypic expression of one or more loci, and thus potentially contain ‘auxiliary’ genetic information. An example of the use of such auxiliary information is the pattern of black and white pigmentation on the ventral side of the fluke in the humpback whale, *Megaptera novaeangliae*, which is used to identify individuals. Rosenbaum *et al.* (1995) analysed the relative proportions of black and white pigmentation on the fluke among humpback whale populations in a worldwide study and detected significant differences in the degree of pigmentation.

Genetic tagging is, in principle, universal as most eukaryote organisms appear to possess VNTR loci (Tautz & Renz, 1984; Tautz *et al.*, 1986). However, the level of polymorphism at microsatellite loci in a given population is not only a function of mutation rate, but also the effective population size. Hence, populations with small effective population size may not contain sufficient polymorphism at the DNA level for genetic tagging purposes. Although rare overall, such populations do exist, e.g., the northern elephant seal (Hoelzel *et al.*, 1993). In some species, such as *Drosophila*, microsatellite loci appear to evolve at a generally slower rate than mammals (Schug *et al.*, 1997), which of course impacts the level of intra-specific variation as well.

Pre-requisite to genetic tagging is tissue from which genomic DNA can be extracted and analysed. Tissue can be collected in an invasive but non-lethal manner, e.g. skin biopsies (Lambertsen, 1987) or by drawing blood (Arctander, 1988). Tissue samples can also be collected in a non-invasive, indirect manner from a variety of sources such as hair (Morin & Woodruff, 1992), feathers (Smith *et al.*, 1992), sloughed skin (Amos *et al.*, 1992), eggshells (Pearce *et al.*, 1997) or even faeces (Constable *et al.*, 1995). Indirect sampling strategies are often random, which may be desirable depending on the objective, but have the advantage that individuals can be sampled without ever being observed.

Obtaining a reliable genotype from low content, degraded genomic DNA has been shown to be problematic due to ‘allelic drop out’ and appearance of ‘spurious alleles’ (Taberlet *et al.*, 1996). These problems appear mainly to be associated with non-invasive sampling strategies where the overall concentration of genomic DNA is typically low and/or highly degraded, such as DNA extracted from sloughed hair or faecal samples (Taberlet *et al.*, 1996). Reliable genotyping from such samples requires multiple extractions and repeated genotyping of the same sample and can thus be quite laborious.

A main feature of a genetic tag is, naturally, the genetic information that it contains. The information can be used for additional phylogenetic and population genetic analyses, and also enable estimation of the probability of a random match (Paetkau & Strobeck, 1994), which in turn can be used to detect possible laboratory errors during genotyping (Palsbøll *et al.*, 1997).

The probability that two unrelated individuals have an identical genotype (i.e. an identical genetic tag) is negatively correlated with the number of loci analysed and degree of variation at each locus. The probability that two unrelated individuals from the same panmictic population have an identical composite genotype (a random match) is simply the probability of identity ( $I$ ), which is estimated as

$$\hat{I} = \prod_k (\sum_i p_i^4 + \sum_i \sum_{j>i} (2p_i p_j)^2) \text{ (Paetkau \& Strobeck, 1994),}$$

where  $p_i$  and  $p_j$  are the frequencies of the  $i$ th and  $j$ th alleles at the  $k$ th locus and  $\hat{I}$  is the estimate of  $I$ . The probability that two related individuals have identical genotypes is naturally higher and can in principle be estimated for any degree of relatedness from the observed allele frequencies. The difficulty in assessing the overall expected number of random matches lies in determining the total contribution from related individuals in a sample. The expected number random matches is simply the product of  $\hat{I}$  (estimated as shown above) and the number of pairwise comparisons (i.e.  $(n(n-1))/2$ , where  $n$  is the number of samples typed). For related individuals this estimation is complicated by the fact that the number of pairs of individuals in the sample that are related at a specific level, e.g. full siblings, is usually unknown. For related individuals the probability of identity ( $I$ ) is higher than for unrelated individuals, but the total number of such pairs is typically lower. For instance, if we assume constant population size, overlapping generations and promiscuous mating, each individual will on average be related to four individuals either as parent or offspring (sharing at least one allele per locus) i.e. the two parents as well as, on average, two offspring. Thus, in a population of 100 individuals there are 4950 pairwise comparisons in total but only approximately 400 parent and offspring relations. Hence, even if the probability of identity for parent-offspring pairs is ten times higher, the absolute contribution of random matches from individuals that are related either as parent or offspring will only equal that from unrelated individuals.

The probability of identity is obviously negatively correlated with the number of loci analysed, which in turn can be utilized to detect possible handling errors during the genotyping process and thus recover missed matches due to laboratory errors. If a genetic tagging experiment is to avoid random matches, a sufficient number of loci has to be analysed to ensure that the expected number of random matches is less than one. At this level of resolution only a few individuals will match at all but one locus. An error during laboratory analysis, caused by pipetting or data recording errors, will typically result in a match at all but one locus (where the error occurred). Barring excessively high error rates, the number of samples that match at all but one locus will thus be few and this subset of samples can be readily re-analysed at the potentially discrepant loci. This feature should, in principle, allow for all matches to be detected.

The chance of missed or random matches during marking experiments employing conventional methods is slim or non-existent provided double marking is employed and that sightings are recorded (e.g. documented photographically) in a manner that allows re-checking in case of inconsistencies. In the absence of double marking, no options for detection of inconsistencies among matches exist, and thus no possibilities for recovery of missed or random matches is available. The situation for individual identification by natural markings is similar to that of conventional marking techniques in this respect.

The main disadvantage of genetic tagging is that a tissue sample is required for each identification, rendering the approach impractical if multiple identifications are required over a short time-scale (e.g. during a single day), especially if a directed, and thus usually invasive, sampling strategy is employed. Table 1 briefly summarizes and compares the three principal marking methods discussed above.

TABLE 1. Comparison of genetic to conventional tagging methods

Tagging technique	Characteristics							
	Universal	Remote marking	Non-invasive	Assessment of random match	Permanency	Automated matching	Error finding	Additional information <sup>b</sup>
Genetic	+/-	+/-	+/-	+	+	+	+	+
Natural markings	-	+	+	+/- <sup>a</sup>	+/-	+/-	+	+
Human tags	-	+/-	-	+/- <sup>a</sup>	+/-	+	+	-

Notes: + = possible, - = not possible, +/- = possible in some instances, but not all. <sup>a</sup>If double marked. <sup>b</sup>Additional information in the tag itself, i.e. no additional measurements when tagging.

### *Different methods for genetic individual identification*

There are multiple genetic techniques available that permit reliable individual identification. Such techniques include single-locus mini- or microsatellite genotyping (Edwards *et al.*, 1991; Jeffreys *et al.*, 1985; Tautz, 1989; Weber & May, 1989), as well as multi-locus methods such as multi-locus fingerprinting (Jeffreys *et al.*, 1985), AFLPs (amplified fragment length polymorphism; Vos *et al.*, 1995) and RAPDs (random amplified polymorphic DNA; Welsh & McClelland, 1990; Williams *et al.*, 1990).

The desirable features of an appropriate method for genetic screening of multiple individuals include:

- (1) PCR (Polymerase Chain Reaction; Mullis & Faloona, 1987) enabling analysis of diluted and/or degraded DNA, e.g. from faeces.
- (2) Simple and direct comparison of data across experiments and laboratories.
- (3) Co-dominant data (for population genetic analyses).

The currently most widespread DNA-based method that fulfils the above criteria is genotyping of microsatellite loci (Edwards *et al.*, 1991; Tautz, 1989; Weber & May, 1989). The microsatellite arrays that are targeted for isolation and subsequent analysis are usually less than a few hundred nucleotides in length (Rassmann *et al.*, 1991) and thus even degraded DNA can be analysed using PCR (see for example Kohn & Wayne, 1997). Amplification products of such length can be resolved on a standard size polyacrylamide, sequencing matrix and the length of each allele is thus determined with the precision of a single nucleotide. Hence, it is straightforward to compare genotypes across experiments and laboratories thereby circumventing the need for additional parallel analyses of samples to validate a putative match, as would be the case for multi-locus fingerprinting.

Microsatellite alleles are encoded as simple integers (i.e. the fragment length of the amplification product) and hence any commercially available database software can be used to store and match genotypes. Although microsatellite data are co-dominant and thus, in principle, readily analysed using traditional population genetic approaches, problems arise as the evolutionary distance increases between the sample partitionings because of the stepwise mutation mode at microsatellite loci (Di Rienzo *et al.*, 1994; Goldstein *et al.*, 1995; Kimmel *et al.*, 1996; Slatkin, 1995).

This complication is only of concern when evolutionary parameters are estimated between distant taxa (populations or species), when mutation becomes a significant factor relative to genetic drift (Tomiuk *et al.*, 1998).

#### *Genetic tagging of North Atlantic humpback whales*

Palsbøll *et al.* (1997) recently presented a study of the North Atlantic humpback whale, *Megaptera novaeangliae*, aimed at investigating the feasibility of genetic tagging. Individual humpback whales are identified by their natural markings, in particular the pigmentation and scarring pattern on the ventral side of the fluke (Katona & Whitehead, 1981). More than two decades of individual identification in this manner have provided an accurate picture of the annual migration cycle of the whales in the North Atlantic (Katona & Beard, 1991). In brief, the whales forage in the summer at high latitudes off eastern North America, West Greenland, Iceland, and Jan Mayen as well as Bear Island in the Barents Sea (Katona & Beard, 1991). The winter constitutes the breeding and calving season and the principal breeding grounds in the North Atlantic are located in the Lesser Antilles (Martin *et al.*, 1984; Mattila *et al.*, 1989; Stevick *et al.*, 1998).

Individual humpback whales return to the same summer feeding area every year. This site-fidelity to a specific summer feeding ground is maternally directed, i.e. individual humpback whales return to the summer feeding ground to which they accompanied their mother during their first year (Clapham & Mayo, 1987). This maternally directed site-fidelity to a specific feeding ground is reflected in the distribution of the sequence variation of the mitochondrial control region amongst individuals across the North Atlantic, suggesting that this behavioral mechanism may have been maintained on an evolutionary time scale (Larsen *et al.*, 1996; Palsbøll *et al.* 1995, 1998).

Palsbøll *et al.* (1997) determined the genotype at six microsatellite loci in 3060 samples collected from free-ranging humpback whales (as either skin biopsies or sloughed skin) during the period from 1988 to 1995 across the entire North Atlantic. The expected number of random matches for unrelated individuals among all 3060 samples was estimated at 0.59 (i.e. the product of number of pairwise comparisons and the probability of identity). Since 692 samples had genotypes already detected once, such samples were inferred as duplicate samples from the individuals previously sampled. This inference was supported by the fact that all samples with identical genotypes also were of identical sex.

The pattern of matches observed in the study by Palsbøll *et al.* (1997) confirmed the present assumptions regarding migration patterns and individual site-fidelity towards specific summer feeding grounds (Table 2). On the summer feeding grounds only four percent of the recaptures were individuals that had been tagged on another summer feeding area, consistent with the observed site-fidelity to specific summer feeding grounds (Table 2).

The study identified 114 matches between a summer feeding area and the winter breeding grounds in the West Indies. Matches from all summer feeding areas were detected and in equal proportions including the eastern North Atlantic, from where observations based upon natural markings are scarce (Stevick *et al.*, 1998). This result indicates that at least some humpback whales in the eastern part of the North Atlantic also breed in the West Indies.



TABLE 2. Summary of recaptures in the study of North Atlantic humpback whales by Palsbøll and co-workers (1997)

Sampling area	Recaptures		Total
	Intra-annual	Inter-annual	
Summer feeding grounds <sup>1</sup>	110	97	207
Winter breeding grounds <sup>2</sup>	291	71	362
Breeding to feeding ground	–	–	114
Feeding to feeding ground <sup>3</sup>	3	6	9

Notes: <sup>1</sup> Gulf of Maine, Gulf of St. Lawrence, off Newfoundland/Labrador, off West Greenland, off Iceland/Jan Mayen, and the Barents Sea. <sup>2</sup>West Indies. <sup>3</sup>Gulf of St. Lawrence and Newfoundland/Labrador (8 recaptures), Jan Mayen/Iceland and the Barents Sea (1 recapture).

The study employed genetic tags to estimate the abundance of males and females, respectively, within the breeding grounds using standard mark-recapture techniques (Seber, 1982). From the samples collected during 1992 and 1993 in the West Indies, Palsbøll *et al.*, (1997) estimated the numbers to be 4894 males (95% confidence interval (CI): 3374–7123) and 2804 females (95% CI: 1776–4463), respectively. The combined estimate of 7698 whales was much higher (but not significantly so) than the most recent previous estimate of some 5505 whales (no 95% CI provided by Katona & Beard, 1991). The statistically significant difference in relative abundance of males and females was unexpected, given the overall equal sex ratio in the population. The most likely explanation is some level of either spatial or temporal segregation amongst females.

The study by Palsbøll *et al.* (1997) demonstrated that genetic tagging is a viable alternative to conventional tagging techniques by applying the approach within the context of a well-known population. The results obtained from the genetic tagging data were consistent with what is known about this population from earlier studies based upon identification of individuals from their natural markings, distribution and whaling records. The work also showed that genetic tagging is feasible even for an evasive and long-ranging species such as the humpback whale.

Genetic tagging depends on generation of correct genotypes, i.e. a low rate of handling errors during collection and laboratory analyses. The rate of handling errors was estimated by Palsbøll *et al.* (1997) at approximately one in a thousand genotypes. As mentioned above, handling errors during laboratory analysis can relatively easily be identified and corrected with a minor additional effort.

#### *Indirect sampling of black and brown bears for genetic tagging*

An elegant and innovative example of how indirect sampling strategies can be employed in connection with genetic tagging is the work by Woods *et al.* (in press) on the black bear, *Ursus americanus*, and brown bear, *U. arctos*, in western Canada.

The objective of the study was to estimate the abundance of the two species of bear, in particular the brown bear, by mark-recapture methods using genetic tagging data. Proper estimation of abundance by mark-recapture methods requires an unbiased sample, i.e. the tagging and recapture of individuals should preferably be conducted in a random manner and with an even distribution of effort in time and

space. Several earlier studies have demonstrated the use of shed hairs from bears for genetic tagging (Taberlet *et al.*, 1997). The remarkable feature of the work by Woods *et al.* (in press) is the sampling design, which fully exploits the virtues of indirect sampling but in a simple, systematic and proactive manner.

Hair samples from free-ranging bears were collected by use of 'hair traps'. Several different models were tested. The model used for the large-scale tagging experiment was relatively simple but effective. Each trap consisted of a 5 m (in diameter) circle of a single barbed wire attached to trees approximately 50 cm above ground level. At the center of each circle was placed a scent lure (fish smell). Prior to setting up the hair traps, several hair collecting devices, including barbed wire, were tested on immobilized bears in order to assess their effectiveness as well as ensuring that the procedures had no adverse effects.

In order to ensure an even distribution of effort and a systematic sampling scheme, hair traps were placed in a 64 km  $\times$  64 km grid, with one trap per 8 km  $\times$  8 km square. Hair collected by the barbed wire was removed and the hair traps were moved within each grid every 10th day over a 40 day period.

More than 1750 hair samples were collected during the large-scale sampling. The species identity of each sample was determined by analysis of the mitochondrial control region, as well as the genotype at six microsatellite loci and the sex of each unique composite microsatellite genotype. Pairs of samples that matched at all loci but one were subsequently re-typed at the discrepant locus. Two samples with an identical genotype were assumed to be from the same individual if the conditional probability that a full sibling would have the same genotype as the individual was less than 0.05.

The approach employed by Woods *et al.* (in press) constitutes a simple but highly efficient sampling design, which should be applicable to a wide variety of species. The sampling scheme is especially suitable to estimates of abundance by mark-recapture and migration rates, as it is easy to ensure a uniform distribution of effort over a wide area. Similar work has been conducted with Grizzly bears as well (Mowat & Strobeck, unpublished data).

#### IDENTIFICATION OF PARENT-OFFSPRING RELATIONS

The above studies have illustrated that genetic tagging is a viable alternative to conventional marking methods and that indirect sampling can be employed in a systematic and proactive manner. However, the strength of mark-recapture experiments based upon individual identification is in terms of estimating abundance. Mark-recapture data are also used to map the range of movements by individuals, which will ultimately provide insights into contemporary population structure and migration rates. However, obtaining reasonable accurate estimates of migration rates and thus population structure by mapping the range of individuals requires a substantive increase in effort relative to, e.g. estimation of abundance. Each mark-recapture session only detects those individuals that were tagged in the marking as well as the re-capture session and is thus limited in scope. An alternative strategy to obtain estimates of contemporary levels of migration/gene flow is to utilize the genetic content of a genetic tag by extending the number of microsatellite loci analysed to enable reliable detection of pairs of samples that are related as parent

and offspring. In this manner all incidences of gene flow contained in the sample will be detected, even if they happened prior to the mark-recapture experiment. In fact, this kind of analysis requires only a single ‘marking’ session, as it does not rely upon recapture. Extending genetic tagging to include the identification of parent-offspring relations will thus enable detection of population structure at a level relevant to ecology and conservation studies. Hence, if the rate of movement or gene flow is less between than within sample partitionings this should, in principle, be detectable even when divergence is too recent or gene flow too high to yield significant levels of genetic heterogeneity among sub-populations detectable by traditional evolutionary statistics, such as Weir’s  $\theta$  (Weir & Cockerham, 1984) or Slatkin’s  $R_{ST}$  (Slatkin, 1995).

#### *Obscure gene flow between pilot whale pods*

An example of how illusory movements of individuals can be regarding estimation of gene flow is the long finned pilot whale, *Globicephala melaena*. Pilot whales are structured in tight social units termed pods. Microsatellite profiling by Amos *et al.* (1991, 1993) of entire pilot whale pods revealed that calves belonging to the same cohort within a pod were sired by one or a few closely related males, and that none of the mature males were likely to have sired any of the calves within their own pod. The most likely explanation for these observations, proposed by the authors, is that pods merge briefly during the mating season when mature males from one pod mate with females from the other pod after which they return to their natal pod (Amos *et al.*, 1991, 1993). This hypothesis is supported by occasional sightings of very large aggregations of pilot whales of several hundred individuals, much larger than the ‘average’ pod size. The study is an example of how poorly migration rates and gene flow may correlate, in this case because mature individuals stay within their natal pod and only ‘migrate’ for a brief period during the mating season. This behaviour would most likely have gone undetected during a mark-recapture experiment but was detected by use of individual-based microsatellite analyses, which allowed the identification of parent-offspring relations.

#### *Identification of parent-offspring relations*

Two individuals related as parent and offspring will share at minimum one allele at each locus. This level of identity (one allele at each locus) could also happen by chance for a pair of unrelated individuals, and is of course even more likely for siblings as well as grandparents and grandchildren. In order to minimize the number of such ‘random matches’, i.e. pairs of individuals that share minimum one allele at each locus but are not parent and offspring, more loci have to be analysed than is required for reliable identification of individuals (see below).

At first, the extra effort may seem costly in terms of the additional laboratory time required, but this has to be viewed against the additional field and laboratory activity necessary to obtain accurate estimates of migration/gene flow among sub-populations from mapping the movements of individuals by mark-recapture. For species with overlapping generations, such as most vertebrates, any random collection of samples will contain more pairs of samples that are related as parent or offspring

than duplicate samples from individuals. An unrealistic but illustrative example would be if an entire population was sampled once. Such a sample would contain few or no recaptures of individuals but all parent-offspring relations. Thus isolating and analysing additional microsatellite loci, both of which are now routine tasks in most molecular laboratories (Rassmann *et al.*, 1991), are likely to represent a lesser effort overall. In addition, oligonucleotide primers will often amplify the homologous locus in closely related species (e.g. Schlötterer *et al.*, 1991; FitzSimmons *et al.*, 1995).

*How many loci are necessary for reliable detection of parent-offspring relations?*

Reliable identification of individuals that are related as parent and offspring (i.e. share a minimum of one allele per locus) from an array of microsatellite genotypes hinges upon the number of analysed loci. If an insufficient number is analysed, some individuals (that are not related as parent or offspring) will share at least one allele per locus simply by chance (a random match), and be erroneously included as related in subsequent analyses. The number of loci necessary to eliminate such random matches depends on the degree of genetic variability, which in turn is a function of the mutation rate and population size, as well as the sample size. A simple explorative approach, which generates an approximate estimate of the number of loci required to avoid multiple random matches, is to resample data from an existing sample of microsatellite genotypes. Typically, a few loci will already have been analysed for population genetic or individual identification purposes, which can serve as the base data for such estimation.

The basic assumption in this kind of estimation, which employs bootstrapped samples of the base data, is that the degree of polymorphism uncovered by the microsatellite loci already analysed is representative for the additional loci that will be analysed to detect parent-offspring relations and that each locus segregates independently. The approach is illustrated by estimations from microsatellite data collected from humpback whales off West Greenland, a North Atlantic summer feeding area. In total, 189 samples were collected between 1988 and 1994 and for each sample the genotype was determined at six microsatellite loci (R. Sponer, unpublished data; Palsbøll *et al.*, 1997). Among the samples 149 unique genotypes were detected, each assumed to represent an individual humpback whale (Palsbøll *et al.*, 1997). The mean number and the upper percentile of the 95% distribution of random parent-offspring matches was estimated for 7–20 loci from bootstrap samples generated from the six analysed loci.

During each simulation the number of loci (e.g. ten) was re-sampled with replacement from the six loci in the original data set. After the required number was sampled, alleles within each locus were randomized amongst individuals and the numbers of pairs of individuals that shared minimum one allele per locus counted. The mean and upper 95% confidence limits were estimated from the variance among 1000 such simulations. The mean and upper limits of the 95% confidence interval should thus include the variances due to the sampling of loci as well as individuals. The results of the simulations are shown in Figures 2 and 3.

The first test is an estimate of the null-distribution and expectation of the current data set (the six original loci). During this test, no re-sampling of loci was performed but alleles were randomized among individuals at each locus. After the randomization

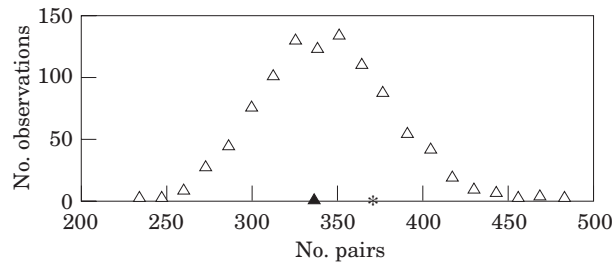


Figure 2. Estimation of the null-distribution and null-expectation of pairs of samples that match on at least one allele per locus in a sample of West Greenland humpback whales. ( $\Delta$ ) frequency distribution of pairs of samples that match on at least one allele per locus in 1000 simulations. ( $\blacktriangle$ ) Mean number of pairs of samples that match on at least one allele per locus. (\*) Observed number of pairs of samples that match on at least one allele per locus in the sample. See text for details on simulation procedures.

of genotypes the number of individuals that shared at least one allele per locus was counted. A total of 1000 such randomizations was performed. The resulting distribution and mean (thus constituting an estimate of the null-distribution and expectation) is illustrated in Figure 2 along with the observed value of individuals that shared at minimum one allele per locus. The sample of 149 individuals contained 371 pairs that shared a minimum of one allele at all six loci. The 1000 simulations yielded a mean of 337 such pairs and the probability of the observed value of 371 (or more) pairs of individuals that shared a minimum of one allele per locus was estimated at 0.19 (Fig. 2). This latter result implies an 19% probability of obtaining the observed number of individuals that share minimum one allele per locus by chance and hence it is likely that our sample contains true parent-offspring pairs, although it is not improbable that all ‘parent-offspring pairs’ are simply random matches.

The above example demonstrates quite clearly that additional loci are necessary to avoid random matches. The purpose of the second estimation is to assess, by simulation, approximately how many loci have to be analysed to reduce the number of such random matches to a negligible level. During each simulation, a new data set was generated by randomly sampling (with replacement) a number of loci (e.g. 15) from the original six loci. After sampling, simulations and estimations were carried out as described above. For each number of loci 1000 simulations were conducted, each with between six and twenty loci. The results are listed in Figure 3. They show that for six loci we observed a mean of close to 410 matches, with an upper limit for the 95% confidence interval of approx. 590 matches.

When the number of bootstrapped loci was increased to 17 the average number of random matches decreased to approximately one, and the upper limit of the 95% confidence interval was reduced to 3.2. Thus our simulations indicated that it will probably be necessary to analyse an additional 11 + loci in addition to the six loci already analysed, in order to ensure that samples which match at a minimum one allele per locus indeed represent true parent-offspring relations (Fig. 3).

The above simulation assumes that all individuals are unrelated. In this particular example this assumption is likely to be violated. Behavioral as well as genetic studies has shown that North Atlantic humpback whales display a maternally directed site-fidelity to specific summer feeding areas, such as West Greenland (Clapham *et al.*,

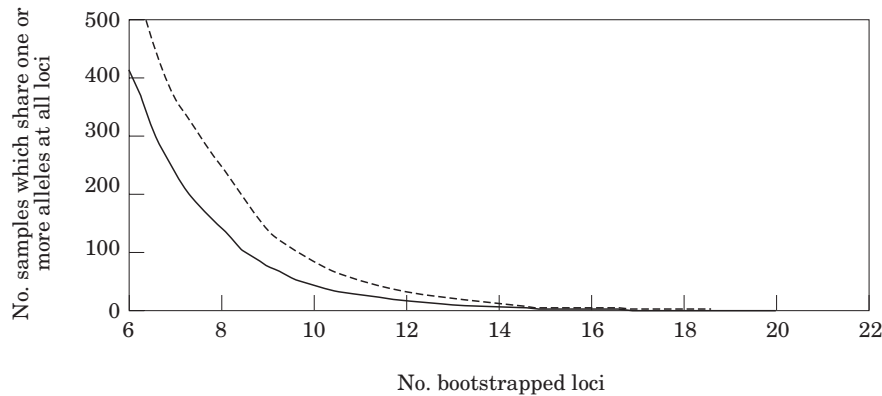


Figure 3. Estimation of the number of random pairs that match at one or more allele per locus. (—) Mean number of random samples that match on at least one allele per locus estimated from 1000 simulations. (----) Upper limit of the 95% confidence interval, estimated from the variance in 1000 simulations. See text for details of simulation procedures.

1993), where the summer feeding aggregation has been estimated at approximately 430 individuals (Finn Larsen, pers. comm.). Mating occurs during the winter in the West Indies where whales from all the North Atlantic summer feeding areas appear to congregate (Mattila *et al.*, 1989). As the mating system of humpback whales is promiscuous (Clapham & Palsbøll, 1997; Mattila *et al.*, 1989) only a few (if any) individuals sampled on the same summer feeding ground are likely to be paternal half-siblings, whereas it is most likely that several individuals are maternal half-siblings due to the maternal-directed site fidelity to summer feeding grounds.

The probability that a pair of half-siblings share at least one allele per locus was also estimated by simulations as above, albeit in a slightly different manner. After sampling (with replacement) the required number of loci from the original sample of six loci, 1000 half-sibling pairs were generated from the simulated data set by sampling alleles by replacement to generate the parents (one shared and two non-shared parents). The genotypes of the half siblings were then generated from the 'parents'. For each simulation 1000 independent half-sibling pairs were generated and the probability of sharing one or more alleles at each locus was estimated as the proportion of the 1000 half-sibling pairs that shared a minimum of one allele at each locus. For each number of loci, 1000 such simulations were conducted. Figure 4 shows the results of the estimation of the probability that half-siblings share one or more alleles at each locus for simulations with 6 to 25 loci. In each instance the mean probability and the upper limit of the 95% confidence interval estimated from 1000 simulations are shown. For 17 loci (which was the number of loci the previous simulations estimated as necessary to avoid random matches for unrelated individuals) the mean probability that a pair of half-siblings share at minimum one allele per locus was estimated at 0.019 and the upper limit of the 95% confidence interval at 0.036. How these probabilities translate into an absolute number of pairs of half-siblings that share one or more alleles at each locus is not straightforward as we do not know the average number of half-siblings per individual. As a simplistic approximation we could assume a constant population size (which may not be true for North Atlantic humpback whales; see Barlow & Clapham, 1997) as well as a

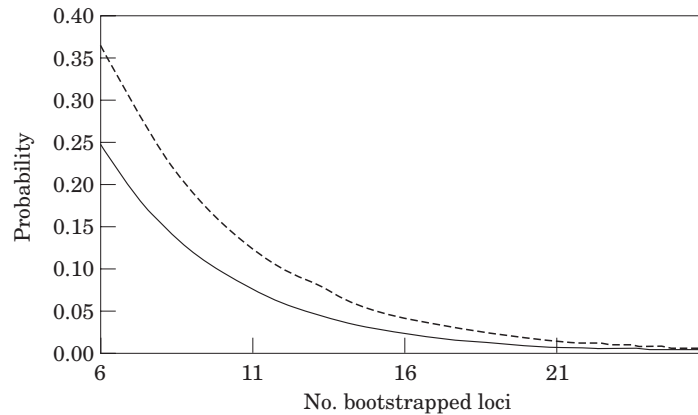


Figure 4. Estimation of the probability that half-siblings match at one or more alleles per locus. (—) Mean number of random samples that match on at least one allele per locus estimated from 1000 simulations. (----) Upper limit of the 95% confidence interval, estimated from the variance in 1000 simulations. See text for details of simulation procedures.

relatively uniform reproductive success among females. Such assumptions would imply that every female on average has two descendants, which thus are half-siblings. The entire North Atlantic humpback whale population has recently been estimated at approximately 10 000 individuals (Smith *et al.*, 1999). Hence, it appears highly unlikely that any two of the *c.* 430 West Greenland humpback whales share the same father, given the common breeding ground and promiscuous mating system. Thus an approximate number of pairs of half-siblings in our sample would be the product of the number of females in the sample and the average number of offspring per female. An average of two offspring per female and 72 sampled females (Palsbøll *et al.*, 1997) yield a mean and upper 95% confidence limit of half-sibling pairs that share at least one allele per locus of 2.8 and 5.2, respectively. Hence, although half-siblings have a higher probability of sharing alleles, the total contribution from such related individuals is low due to the relatively few such pairs in a sample.

The estimations obtained from the above simulations reflect the somewhat simple and crude methods used to explore the data. It is likely that more powerful tests based upon likelihood ratios will increase the accuracy substantially (i.e. as implemented in the program Kinship by Goodnight & Queller [in press]). However, the effect of a less powerful test is likely to be an overestimate of the number of random matches and thus the simulation approach should yield a conservative estimate of the necessary number of loci required for reliable identification of parent-offspring relations. As the number of loci and pairwise comparisons increases so does the likelihood of the inclusion of novel alleles due to mutation. This is especially true for loci with mutation rates as high as those observed at microsatellite loci. Hence, a mismatch at a single locus in an array of, say, 17 loci could be due to a germ line mutation. Thus (as is the case for individual identification as well) it is advisable to analyse sufficient loci so that sharing minimum one allele at all but one locus is improbable as well.

*Inferring contemporary population structure from parent-offspring relations*

Once parent-offspring relations are detected, relatively simple standard statistics (such as a standard G-test) can be employed to detect population structure among sample partitionings. Indication of structure, i.e. reduced migration or gene flow between two sample partitionings would be evident as significantly fewer inter-population parent-offspring relations relative to the intra-population parent-offspring relations. The more extreme the difference in this proportion, the lower the degree of gene flow and thus the higher the degree of divergence, as with evolutionary estimates. The relative contributions of male- and female-mediated gene flows can be estimated if relative age can be assigned to individuals. If no information on age is available, such estimation is restricted to male-male and female-female parent-offspring relations only. It is of course not possible to distinguish between migration and gene flow in this kind of analysis.

The proportion of parent-offspring relations in a random sample of individuals from a single sub-population is related to population census size. Hence, an estimate of relative abundance in sub-populations can be derived from the proportion of parent-offspring relations in each population sample. If census size is known for one of the sub-populations the remainder can thus be extrapolated from this number. Obviously, the accuracy of these estimates hinges upon the proportion of the population that has been sampled. At some (unknown) lower threshold few or no parent-offspring relations are contained in a random collection of individuals and thus no direct inference regarding contemporary population structure is possible. The relatively large number of microsatellite loci, which has to be analysed in order to reliably identify parent-offspring relations, might be sufficient to estimate a genealogy, where the OTU is the individual and the genetic distance between OTUs simply one minus the proportion of alleles shared (see e.g. Bowcock *et al.*, 1994). This approach has the additional advantage that it circumvents the need for an *a priori* partitioning sample and the fit of topology with different evolutionary models simply assessed by estimation of branch support. Other individual-based analyses, such as relative reproductive success and social structure, can also be conducted with such data as well as traditional evolutionary analyses.

## CONCLUSIONS

The ease with which multiple co-dominant and highly variable nuclear microsatellite loci can be analysed in many samples enables analyses of individuals rather than populations. Such a change in focus will facilitate estimation of contemporary (as opposed to evolutionary) estimates of population structure and gene flow, which traditional evolutionary approaches cannot address. Migration rates and population size can be estimated directly by genetic identification of individuals. Such 'genetic tagging' is in principle widely applicable and has been shown to constitute a feasible alternative to conventional methods of individual identification. Genetic tags can be generated from blood, hair, skin and faecal material and thus individuals can be tagged and tracked directly as well as indirectly. The genetic content of such individual genetic tags can also serve as basis for traditional population genetic analyses in an evolutionary context.



A relatively modest increase in experimental effort, i.e. analysis of additional microsatellite loci per sample, will enable reliable detection of related individuals, such as parent-offspring relations. Extending the genetic analysis from identification of individuals to related individuals exploits the genetic content of the data more fully and will in principle provide insight into contemporary population genetic and ecological aspects, such as abundance and gene flow, and with less effort in the field than is necessary for studies based solely upon identification of individuals.

## ACKNOWLEDGEMENTS

The work upon which this paper is based would not have been possible without the tremendous and dedicated effort by numerous people and organizations in the field, office and laboratory. Much of the work was conducted under the auspices of the international collaborative project YoNAH (Years of the North Atlantic Humpback whale), which included participants from Canada, Denmark, the Dominican Republic, Iceland, Greenland, Norway, the United Kingdom and the U.S.A. Thanks are also due to Drs Peter Arctander, Richard Hudson and Michel Milinkovitch for support. Dr Curtis Strobeck is thanked for providing the unpublished information on the work on the bears. The paper was partly written while supported by a visiting scientist grant from Université Libre de Bruxelles. The programs used for the simulations are available upon request from the author (compiled for Microsoft Windows 95).

## REFERENCES

- Agler BA, Beard JA, Bowman RS, Corbett HD, Frohock SE, Hawvermale MP, Katona SK, Sadove SS, Seipt IE. 1990.** Fin whale (*Balaenoptera physalus*) photographic identification: methodology and preliminary results from the western North Atlantic. *Report of the International Whaling Commission Special Issue 12*: 349–356.
- Amos B, Barrett J, Dover GA. 1991.** Breeding behaviour of pilot whales revealed by DNA fingerprinting. *Heredity 67*: 49–55.
- Amos B, Schlötterer C, Tautz D. 1993.** Social structure of pilot whales revealed by analytical DNA profiling. *Science 260*: 670–672.
- Amos W, Whitehead H, Ferrari MJ, Glockner-Ferrari DA, Payne R, Gordon J. 1992.** Restrictable DNA from sloughed cetacean skin; its potential for use in population analysis. *Marine Mammal Science 8*: 275–283.
- Arctander P. 1988.** Comparative studies of avian DNA by restriction fragment length polymorphism analysis: Convenient procedures based on blood samples from live birds. *Journal für Ornithologie 129*: 205–216.
- Barlow J, Clapham PJ. 1997.** A new birth-interval approach to estimating demographic parameters of humpback whales. *Ecology 78*: 535–546.
- Beukers JS, Jones GP, Buckley RM. 1995.** Use of implant microtags for studies of populations of small reef fish. *Marine Ecology Progress Series 125*: 61–66.
- Bossart JL, Pashley Prowell D. 1998.** Genetic estimates of population structure and gene flow: limitations, lessons and new directions. *Trends in Ecology and Evolution 13*: 202–206.
- Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994.** High resolution of human evolutionary trees with polymorphic microsatellites. *Nature 368*: 455–457.
- Carlson CA, Mayo CA, Whitehead H. 1990.** Changes in the ventral fluke pattern of the humpback whale (*Megaptera novaeangliae*), and its effects on matching; evaluation of its significance to photo-identification research. *Report of the International Whaling Commission Special Issue 12*: 105–111.

- Clapham PJ, Baraff LS, Carlson CA, Christian MA, Mattila DK, Mayo CA, Murphy MA, Pittman S. 1993.** Seasonal occurrence and annual return of humpback whales in the southern Gulf of Maine. *Canadian Journal of Zoology* **71**: 440–443.
- Clapham PJ, Mayo CA. 1987.** Reproduction and recruitment of individually identified humpback whales, *Megaptera novaeangliae*, observed in Massachusetts Bay, 1979–1985. *Canadian Journal of Zoology* **65**: 2853–2863.
- Clapham PJ, Palsbøll PJ. 1997.** Molecular analysis of paternity shows promiscuous mating in female humpback whales (*Megaptera novaeangliae*, Borowski). *Proceedings of the Royal Society, London* **B264**: 95–98.
- Clark DR. 1968.** Branding as a marking technique for amphibians and reptiles. *Copeia* **1968**:
- Constable JJ, Packer C, Collins DA, Pusey AE. 1995.** Nuclear DNA from primate dung. *Nature* **373**: 393.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB. 1994.** Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA* **91**: 3166–3170.
- Dietz R, Heide-Jørgensen M-P. 1995.** Movements and swimming speed of narwhals, *Monodon monoceros*, equipped with satellite transmitters in Melville Bay, northwest Greenland. *Canadian Journal of Zoology* **73**: 2106–2119.
- Edwards A, Civitello A, Hammond HA, Caskey CT. 1991.** DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American Journal of Human Genetics* **49**: 746–756.
- FitzSimmons NN, Moritz C, Moore SS. 1995.** Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution. *Molecular Biology and Evolution* **12**: 432–440.
- Goldstein DB, Linares AR, Cavelli-Sforza LL, Feldman MW. 1995.** An evaluation of genetic distances for use with microsatellite loci. *Genetics* **139**: 463–471.
- Goodknight KF, Queller DC.** in press. Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology*.
- Grigione MM, Burman P, Bleich VC, Pierce BM. 1999.** Identifying individual mountain lions *Felis concolor* by their tracks: refinement of an innovative technique. *Biological Conservation* **88**: 25–32.
- Hammond PS, Mizroch SA, Donovan GP. 1990.** Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters. *Report of the International Whaling Commission Special Issue*.
- Hiby L, Lovell P. 1990.** Computer aided matching of natural markings: a prototype system for grey seals. *Report of the International Whaling Commission Special Issue No. 12*, 57–62.
- Hoelzel AR, Halley J, O'Brien SJ, Campagna C, Arnbom T, Le Boeuf B, Ralls K, Dover GA. 1993.** Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *Heredity* **84**: 443–449.
- Hudson RR. 1990.** Gene genealogies and the coalescent process. In: Futuyma DJ, Antonovics J, eds. *Oxford Surveys in Evolutionary Biology* Oxford: Oxford University Press, 1–44.
- Hudson RR, Boos DD, Kaplan NL. 1992.** A statistical test for detecting geographic subdivision. *Molecular Biology and Evolution* **9**: 138–151.
- Jeffreys AJ, Wilson V, Thein SL. 1985.** Individual-specific 'fingerprints' of human DNA. *Nature* **316**: 76–79.
- Katona SK, Beard JA. 1991.** Humpback whales (*Megaptera novaeangliae*) in the western North Atlantic Ocean. *Memoirs of the Queensland Museum* **30**: 307–321.
- Katona SK, Whitehead HP. 1981.** Identifying humpback whales using their natural markings. *Polar Record* **20**: 439–444.
- Kaye SV. 1960.** Gold-198 wires used to study movements of small mammals. *Science* **13**: 824.
- Kimmel M, Chakraborty R, Stivers DN, Deka R. 1996.** Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143**: 549–555.
- Kohn MH, Wayne RK. 1997.** Facts from faeces revisited. *Trends in Ecology and Evolution* **12**: 223–227.
- Lambertsen RH. 1987.** A biopsy system for large whales and its use for cytogenetics. *Journal of Mammalogy* **68**: 443–445.
- Larsen AH, Sigurjónsson J, Øien N, Vikingsson G, Palsbøll PJ. 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. *Proceedings of the Royal Society, London* **B263**: 1611–1618.

- Marion WR, Shamus JD. 1977.** An annotated bibliography of bird-marking techniques. *Bird-Banding* **48**: 42–61.
- Martin AR, Katona SK, Mattila D, Hembree D, Waters TD. 1984.** Migration of humpback whales between the Caribbean and Iceland. *Journal of Mammalogy* **65**: 330–333.
- Mattila DK, Clapham PJ, Katona SK, Stone GS. 1989.** Population composition of humpback whales, *Megaptera novaeangliae*, on Silver Bank, 1984. *Canadian Journal of Zoology* **67**: 281–285.
- Morin PA, Woodruff DS. 1992.** Paternity exclusion using multiple hypervariable microsatellite loci amplified from nuclear DNA of hair cells. In Martin RD, Dixson AF, Wickings EJ, eds. *Paternity in primates: Genetic tests and theories* Basel: Karger, 63–81.
- Mullis KB, Faloona F. 1987.** Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**: 335–350.
- Paetkau D, Strobeck C. 1994.** Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* **3**: 489–495.
- Palsbøll PJ, Allen J, Bérubé M, Clapham PJ, Feddersen TP, Hammond P, Hudson RR, Jørgensen H, Katona S, Larsen AH, Larsen F, Lien J, Mattila DK, Sigurjónsson J, Sears R, Smith T, Sponer R, Stevick PT, Øien N. 1997.** Genetic tagging of humpback whales. *Nature* **388**: 676–679.
- Palsbøll PJ, Bérubé M, Jørgensen H. 1999.** Multiple levels of single-strand slippage at cetacean tri- and tetranucleotide repeat microsatellite loci. *Genetics* **151**: 285–296.
- Palsbøll PJ, Clapham PJ, Jørgensen H, Larsen F, Mattila DK, Sears RSJ, Vasquez O. 1998.** The value of parallel analysis of uni- and bi-parental inherited loci: the North Atlantic humpback whale (*Megaptera novaeangliae*). In: Karp A, Isaac PG, Ingram DS, eds. *Molecular tools for screening biodiversity: Plants and animals*. London: Chapman and Hall, 426–430.
- Palsbøll PJ, Clapham PJ, Mattila DK, Larsen F, Sears R, Siegismund HR, Sigurjónsson J, Vasquez O, Arctander P. 1995.** Distribution of mtDNA haplotypes in North Atlantic humpback whales: the influence of behaviour on population structure. *Marine Ecology Progress Series* **116**: 1–10.
- Pearce JM, Fields RL, Scribner KT. 1997.** Nest materials as a source of genetic data for avian ecological studies. *Journal of Field Ornithology* **68**: 471–481.
- Pennycuik CJ. 1978.** Identification using natural markings. In: Stonehouse B, ed. *Animal marking – Recognition marking in animals in research*. London: MacMillan, 147–159.
- Queller DC, Strassmann JE, Hughes CR. 1993.** Microsatellites and kinship. *Trends in Ecology and Evolution* **8**: 285–288.
- Rassmann K, Schlötterer C, Tautz D. 1991.** Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting. *Electrophoresis* **12**: 113–118.
- Rosenbaum HC, Clapham PJ, Allen J, Nicole-Jenner M, Jenner C, Florez-González L, Unrbám RJ, Ladrón GP, Mori K, Yamaguchi M, Backer CS. 1995.** Geographic variation in ventral fluke pigmentation of humpback whale *Megaptera novaeangliae* population worldwide. *Marine Ecology Progress Series* **124**: 1–7.
- Schlötterer C, Amos B, Tautz D. 1991.** Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**: 63–65.
- Schlötterer C, Pemberton J. 1994.** The use of microsatellites for genetic analysis of natural populations. In: Schierwater B, Streit B, Wagner GP, DeSalle R, eds. *Molecular ecology and evolution: Approaches and applications*. Basel, Switzerland: Birkhäuser, 203–214.
- Schug MD, Mackay TFC, Aquadro CF. 1997.** Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genetics* **15**: 99–102.
- Seber GAF. 1982.** *The estimation of animal abundance and related parameters*. London: Charles Griffin.
- Slatkin M. 1995.** A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**: 457–462.
- Smith EFG, Arctander P, Fjeldsá J, Amir OG. 1992.** A new species of shrike (Laniidae: *Lanius*) from Somalia, verified by DNA sequence data from the only known individual. *Ibis* **133**: 227–235.
- Smith TD, Allen J, Clapham PJ, Friday N, Hammond PS, Katona S, Larsen F, Lien J, Mattila D, Palsbøll PJ, Sigurjónsson J, Stevick PT, Øien N. 1999.** An ocean-basin-wide mark-recapture study of the North Atlantic humpback whale (*Megaptera novaeangliae*). *Marine Mammal Science* **15**: 1–32.
- Stevick PT, Øien N, Mattila DK. 1998.** Migration of a humpback whale (*Megaptera novaeangliae*) between Norway and the West Indies. *Marine Mammal Science* **14**: 162–166.
- Taberlet P, Camarra J-J, Griffin S, Uhrès E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J. 1997.** Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology* **6**: 869–876.

- Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J. 1996.** Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**: 3189–3194.
- Tautz D. 1989.** Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* **17**: 6463–6471.
- Tautz D, Renz M. 1984.** Simple sequences are ubiquitous repetitive components of eukaryote genomes. *Nucleic Acids Research* **12**: 4127–4138.
- Tautz D, Trick M, Dover GA. 1986.** Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**: 652–656.
- Tomiuk J, Guldbrandtsen B, Loeschcke V. 1998.** Population differentiation through mutation and drift—a comparison of genetic distance measures. *Genetica* **102/103**: 545–558.
- Vos P, Hogers R, Bleker M, Rijans M, Vande Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M, 1995.** AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407–4414.
- Weber JL, May PE. 1989.** Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**: 388–396.
- Weir BS, Cockerham CC. 1984.** Estimating  $F$ -statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.
- Welsh J, McClelland M. 1990.** Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**: 7213–7218.
- Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV. 1990.** DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**: 6531–6535.
- Woods JG, Paetkau D, Lewis A, McLellan BN, Proctor M, Strobeck C. In press.** Genetic tagging free-ranging black and brown bears. *Wildlife Society Bulletin*.