



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

Non-invasive genetic sampling and individual identification

PIERRE TABERLET* AND GORDON LUIKART

*Laboratoire de Biologie des Populations d'Altitude, CNRS UMR 5553,
Université Joseph Fourier, BP 53, F-38041 Grenoble Cedex 9, France*

Individual identification via non-invasive sampling is of prime importance in conservation genetics and in behavioural ecology. This approach allows for genetics studies of wild animals without having to catch them, or even to observe them. The material used as a source of DNA is usually faeces, shed hairs, or shed feathers. It has been recently shown that this material may lead to genotyping errors, mainly due to allelic dropout. In addition to these technical errors, there are problems with accurately estimating the probability of identity (PI, or the probability of two individuals having identical genotypes) because of the presence of close relatives in natural populations. As a consequence, before initiating an extensive study involving non-invasive sampling, we strongly suggest conducting a pilot study to assess both the technical difficulties and the PI for the genetic markers to be used. This pilot study could be carried out in three steps: (i) estimation of the PI using preliminary genetic data; (ii) simulations taking into account the PI and choosing the technical error rate that is sufficiently low for assessing the scientific question; (iii) polymerase chain reaction (PCR) experiments to check if it is technically possible to achieve this error rate.

© 1999 The Linnean Society of London

ADDITIONAL KEY WORDS:—hairs – faeces – feathers – allelic dropout – individual identification – conservation genetics – behavioural ecology – pilot study – microsatellites – probability of identity.

CONTENTS

Introduction	42
Technical challenges	43
Sample preservation	43
DNA extraction	43
DNA amplification	44
Contamination control	44
Genotyping errors	44
Probability of identity	47
Estimation and bias	47
Magnitude and consequences of bias	48
Avoiding bias problems	49
Perspectives and guidelines	50
Acknowledgements	53
References	53

* Corresponding author. Email: pierre.taberlet@ujf-grenoble.fr

INTRODUCTION

Assessing the genetics of wild animal populations twenty or thirty years ago required relatively large amounts of fresh tissue (often from several different organs) for protein electrophoresis studies (Lewontin, 1991; Murphy *et al.*, 1996). As a consequence, the animals were often killed for the scientific study. This destructive sampling technique was widely used for small animals like passerine birds, rodents, or insectivores. During that period, scientists were under pressure to avoid such destructive approaches, and to use tissues that could be collected non-destructively. For large animals, the first alternative was to use blood samples and to study only serum proteins (e.g. Schmutz *et al.*, 1987; Stratil *et al.*, 1990). For birds, attempts were made to use growing feathers as a source of proteins (Marsden & May, 1984). The situation did not change in 1979 when new genetic markers—restriction fragment length polymorphisms (RFLP) of mitochondrial DNA (mtDNA)—were introduced; these markers also often required destruction of the animals in order to extract mtDNA from fresh liver (Awise *et al.*, 1979; Awise, Lansman & Shade, 1979; Brown & Wright, 1979).

The context radically changed with the discovery of the Polymerase Chain Reaction (PCR) (Saiki *et al.*, 1985) using a thermostable polymerase (Mullis & Faloona, 1987; Saiki *et al.*, 1988). This new technique has become widely used by population geneticists (see reviews in Erlich, Gelfand & Sninsky, 1991; White, Arnheim & Erlich, 1989; Wright & Wynford-Thomas, 1990). PCR allows amplification of DNA from minute amounts of fresh, alcohol-preserved, or even dry tissues. To the general satisfaction of population biologists, the destructive sampling approach became unnecessary, and a new era of non-destructive genetic sampling began, at least for vertebrates. Today, blood samples, small biopsies, and even plucked hairs or feathers can provide a ready source of template DNA for PCR and for the subsequent genetic analyses.

Another advancement was demonstrating that shed hairs, shed feathers, sloughed or shed skin, faeces or even saliva could also provide a valuable source of DNA (see review in Morin & Woodruff, 1996; see also Briker *et al.*, 1996; Valsecchi *et al.*, 1998). Scientists working in conservation biology and in ethology were particularly interested in non-invasive sampling techniques, because the animal does not have to be captured, disturbed, or even observed, making genetic sampling much easier. This interest was corroborated by the publication of many preliminary reports (e.g. Taberlet & Bouvet, 1992; Morin *et al.*, 1993; Tikel, Blair & Marsh, 1996; Reed *et al.*, 1997), the dominant idea being that non-invasive sampling has the same potential as blood or biopsy samples. However, after 1995, some studies revealed a risk of genotyping errors associated with the use of hair or faeces (Gerloff *et al.*, 1995; Taberlet *et al.*, 1996, 1997; Gagneux, Boesch & Woodruff, 1997; Goossens, Waits & Taberlet, 1998). These technical difficulties might explain why, during the six or seven years after the initial development of these non-invasive methods, only a few relatively comprehensive studies have been published. Two of them (Morin *et al.*, 1994; Gagneux *et al.*, 1997) deal primarily with the degree of relatedness among individuals in communities of wild chimpanzees using hairs collected in abandoned sleeping nests. Apparently, the same protocol used in these two studies can produce a substantial level of genotyping errors (31%) when applied to wild chimpanzees (Gagneux *et al.*, 1997). This does not mean that the conclusions in these two studies were wrong, rather that some genotyping errors might have gone undetected, as

could occur in any study where the error rate is not quantified. The three other most comprehensive studies concern bears (Taberlet *et al.*, 1997; Woods *et al.*, 1999) and coyotes (Kohn *et al.*, 1999) and present different methods for detecting genotyping errors. In addition to technical difficulties with producing reliable genotype data, there are also problems with the statistics used to quantify the probability of individuals having identical genotypes (PI or probability of identity) in natural populations.

In many research fields, such as conservation genetics, behavioural ecology, forensics, and demography, the genetic identification of individuals (DNA ‘fingerprinting’) can be of prime importance. This explains why the number of research projects using non-invasive sampling techniques is dramatically increasing. But can non-invasive genetic sampling always provide reliable genotypes for identifying individuals? In which conditions will non-invasive sampling supplant more conventional approaches? The purpose of this paper, based on the current literature, is to assess the potential and the limitations of non-invasive genetic sampling (of shed hairs, shed feathers, faeces) for identifying individuals, both from a technical and from a statistical point of view. We also suggest guidelines and a pilot study approach for quantifying and avoiding errors associated with non-invasive sampling and individual identification.

TECHNICAL CHALLENGES

Sample preservation

The first problem facing field biologists is choosing a preservation method that will not compromise the genetic analysis several weeks or months after the sampling. Up to now, only two comparative studies have been conducted on (bear and baboon) faeces (Frantzen *et al.*, 1998; Wasser *et al.*, 1997). It appeared that the best faeces preservation methods involve: (1) the use of a 4:1 ratio of desiccating silica beads (Sigma, St Louis, MO, USA) either at room temperature or at -20°C ; or (2) storage at room temperature in a DMSO/EDTA/Tris/salt solution (Seutin, White & Boag, 1991). Concerning hairs and feathers, it seems that they can be well preserved dry at room temperature, as shown by many studies (e.g. Morin *et al.*, 1994; Taberlet & Bouvet, 1994). More rigorous studies with multiple replicates and controls are needed to thoroughly assess the performance of various preservation protocols on a diverse range of sample types.

DNA extraction

Conventional DNA extraction methods using phenol/chloroform purification and ethanol precipitation usually give poor results, due to either the low DNA content of shed hairs or feathers, or to the copurification of PCR inhibitors when using faeces. Fortunately, these technical difficulties have been overcome. The chelex extraction method for hairs (Singer-Sam, Tanguay & Riggs, 1989; Walsh, Metzger & Higuchi, 1991), and the silica method for faeces give very good results (Boom *et al.*, 1990; Höss & Pääbo, 1993; Kohn *et al.*, 1995). More recently, commercial kits

for DNA extraction have been developed based on the silica method (e.g. QIAamp Kit, Qiagen). The comparative study of Wasser *et al.* (1997) showed that kits perform as well as the previous more complex silica-based methods in removing PCR inhibitors. It is very difficult to assess the usefulness of the DNA extracted by quantifying it using conventional methods (e.g. fluorometry) because (i) the DNA quantity is too low, (ii) it could be too degraded for amplification even if large amounts were extracted, (iii) much of the DNA could be from microbes when using scats.

DNA amplification

The amount of DNA extracted from shed hairs or feathers, or from faeces, is often only a few nanograms. Some samples provide less than one nanogram. As a consequence, when only a small fraction of such extracts is used as template, the number of PCR cycles has to be increased in order to get a PCR product that can be detected and analysed. Theoretically, if the template DNA contains only one target molecular, and if the PCR efficiency is 0.7 (proportion of molecules amplified during each cycle), then 44 amplification cycles are necessary to obtain 10 ng of target DNA at the end of the PCR (Ramekers, Hummel & Herrmann, 1997). The number of amplification cycles can be lower if a very sensitive detection method is used (radioactivity, fluorescence). Recent improvements have provided a *Taq* polymerase that is active only after a 10 min incubation at 95°C (AmpliTaq Gold, Perkin-Elmer) and that reduces non-target amplifications via a hot start PCR (Birch, 1996; Kebelmann-Betzing *et al.*, 1998). This allows for more PCR cycles without problems, which means that a single target molecule can be easily detected if the PCR conditions are optimized.

Contamination control

Along with the possibility of detecting a single target molecule, there is also a possibility of detecting a single contaminant molecule. Therefore, extreme care must be taken to avoid contaminations, either by PCR products, or by concentrated genomic DNA (Kwok, 1990). From this point of view, working with non-invasive genetic sampling is similar to ancient DNA studies (e.g. Béraud-Colomb *et al.*, 1995; Stoneking, 1995), and the same rules must be applied: (i) physical separation of the laboratory rooms where pre- and post-PCR experiments are carried out, (ii) avoidance of concentrated DNA extracts in the extraction room, (iii) use of dedicated pipettes with aerosol-resistant pipette tips, (iv) continuous monitoring of all reagents for DNA contamination (negative controls), etc. Respecting all these rules to avoid contamination is much more time-consuming and expensive than conventional studies.

Genotyping errors

When using hairs, feathers, or faeces from free ranging animals, the total amount of DNA available for the genetic typing can be very low and is often in the picogram

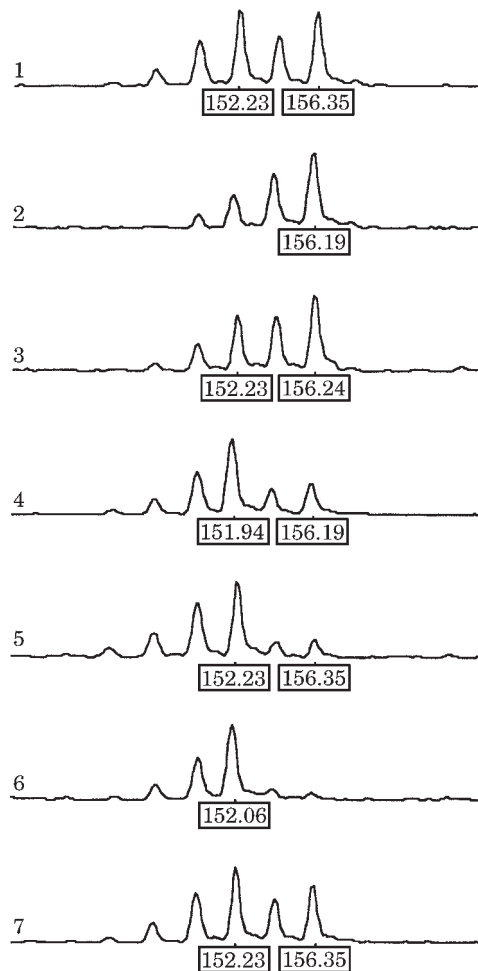


Figure 1. Example of allelic dropout producing false homozygotes (from Taberlet *et al.*, 1997). Results of seven independent genotyping experiments (multiple-tube approach) at microsatellite locus G10L. For the seven experiments, the same DNA extraction was used from a shed brown bear hair collected in the field. The PCR products were detected and analysed on a ABI 377 prism automated sequencer. This bear is a heterozygote with alleles 152 and 156 (bp). PCRs 1 and 7 detect both alleles. PCRs 2 and 6 show examples of allelic dropout. PCRs 3, 4, and 5 exhibit ambiguous results as one of the two alleles is much weaker than the other.

range. Under these circumstances, and if the PCR allows detection of a single target molecule, only one allele of a heterozygous individual is often detected (Gerloff *et al.*, 1995; Foucault *et al.*, 1996; Taberlet *et al.*, 1996, 1997; Gagneux *et al.*, 1997; Goossens *et al.*, 1998). This type of error, called 'allelic dropout', produces false homozygotes (Fig. 1), and can be explained by sampling stochasticity: when pipetting template DNA in a very dilute DNA extract, sometimes only one of the two alleles is pipetted, amplified, and detected (Taberlet *et al.*, 1996). Allelic dropout might also be related to extreme DNA degradation, or differential amplification of one of the two alleles. A mathematical model has been developed to account for these stochastic

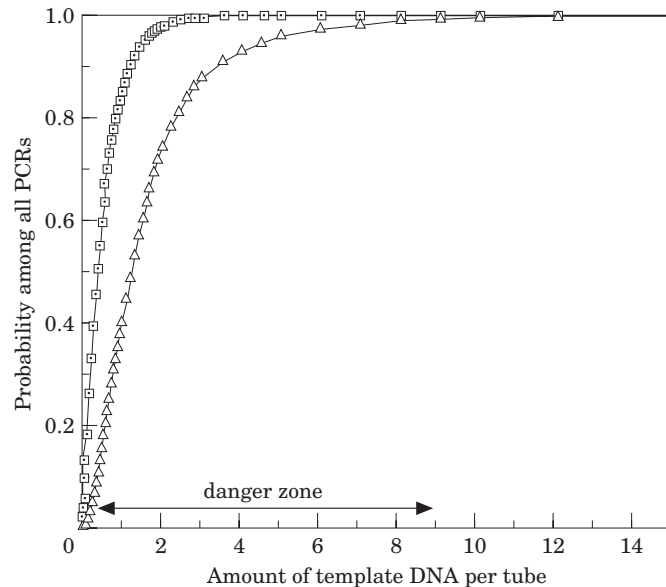


Figure 2. Results of the simulation concerning the genetic typing of a heterozygous individual bearing alleles A and B, and accounting for stochastic events when pipetting template DNA in a diluted DNA extract. Probability among all PCRs of obtaining a positive PCR (allele A or B; □), a correct genotype (alleles A and B; △), according to the amount of template DNA (one unit of template DNA corresponds to the equivalent of the DNA content of one diploid cell). The danger zone is the range of template DNA concentrations where allelic dropout can occur. See Taberlet *et al.* (1996) for further details.

events, and computer simulations have been performed to quantify error rates (Taberlet *et al.*, 1996) (Fig. 2). The probability of obtaining a correct genotype at the 99% confidence level requires the use of template DNA equivalent to the content of eight diploid cells, which is about 56 pg of DNA for mammals. However, a 99% probability of at least some PCR product (e.g. 1 of 2 alleles in a heterozygote) is reached using as template the equivalent of the DNA content in only 2.4 cells (about 17 pg). Therefore, when using between 17 and 56 pg of template DNA, there is a high risk of allelic dropout, despite obtaining positive PCRs in almost all experiments.

In our experience, when working with shed hairs or faeces collected in the field, the amount of template DNA used per PCR is highly variable, and often in the range where allelic dropout occurs. In order to avoid this genotyping error, an ideal solution would be to quantify the DNA concentration of the extract, and to adjust the amount of template DNA per PCR. But such an estimation is either technically very difficult due to the very low DNA content of hairs or feathers, or even impossible when using faeces because the DNA of the species under study is copurified with large amount of prey or bacterial DNA. Furthermore, even if large quantities of DNA are obtained, it may be highly degraded and the number of intact target molecules may be in the range leading to allelic dropout. This could be especially problematic for long PCR products.

Another kind of genotyping error has been detected when using very little template DNA and when amplifying dinucleotide microsatellites: artefacts generated during the amplification process and then misinterpreted as true alleles because they have

the same characteristic shadow band profile. If such a ‘false allele’ occurs in a homozygous individual, then this individual can be recorded as a heterozygote, and if it occurs in a heterozygous individual, then the presence of three ‘alleles’ will allow the detection of the error. These artefacts generating false alleles are easily confused with sporadic contaminations. They generally occur in less than 5% of the PCRs (Taberlet *et al.*, 1996), but should not be disregarded as they can lead to erroneous genotypes.

Researchers have proposed a multiple tubes approach (Navidi, Arnheim & Waterman, 1992; Taberlet *et al.*, 1996) to overcome all potential sources of genotyping errors (allelic dropout, false alleles, sporadic contaminations). In this approach, each genotype amplification is repeated independently several times, and the genotype is deduced by analysing the set of all experiments (see Taberlet *et al.* [1996] for precise guidelines).

PROBABILITY OF IDENTITY

In studies requiring individual identification (i.e. DNA ‘fingerprinting’), it is important to quantify the power or ability of molecular markers to resolve between different individuals. The statistic most commonly used is the probability of identity (PI), or the probability of obtaining identical genotypes given certain allele frequency distributions. In this section we (i) explain the problems (biases) associated with estimating PI, (ii) give examples illustrating the magnitude and possible consequences of the bias observed in natural populations, and (iii) provide guidelines for avoiding problems that the bias can cause.

Estimation and bias

To understand the causes of the bias, it will help to first compute the theoretical probability of identity for an example data set and explain the assumptions involved. We could compute PI as either the probability of (i) obtaining a second genotype identical to one we already have *in hand* (e.g. one from DNA found in the field), or (ii) obtaining two identical genotypes in a given population. An example application of the first PI would be in wildlife forensics where we may first obtain a DNA ‘fingerprint’ for a blood stain from an illegally killed bear, and then ask the question: what is the probability of finding a second identical DNA ‘fingerprint’ among all the bears that exist? An application of the second PI would be in planning a study requiring accurate individual identification. In this case, we could estimate the number of loci (with a certain heterozygosity) that would be needed to achieve a sufficiently low PI to resolve all individuals.

We now compute the (second) PI to illustrate the potential bias associated with the estimation of the theoretical PI. Consider a population sample for which we analyse 3 loci having 2, 4, and 8 alleles, respectively, and having equal frequencies (0.5, 0.25, and 0.12 for each locus respectively). We then estimate PI in two steps. First we compute PI for each locus using the following equation (Paetkau & Strobeck, 1994):

$$PI = \sum p_i^4 + \sum (2p_i p_j)^2 \quad (1)$$

where p_i and p_j are the frequencies of the i th and j th alleles. Note that the following less biased equation exists for correcting for small samples of individuals (Kendall & Stewart, 1977):

$$\text{PI} = [n^3(2a_2^2 - a_4) - 2n^2(a_3 + 2a_2) + n(9a_2 + 2) - 6] / (n-1)(n-2)(n-3) \quad (2)$$

where n is the sample size, a_i equals $\sum p_j^i$ and p_j is the frequency of the j th allele.

Using equation 1, the PI for the first locus (with 2 alleles) is $(0.5^4 + 0.5^4) + (2 \cdot 0.5 \cdot 0.5)^2 = 0.375$. Thus, approximately 37 out of 100 individuals should theoretically have the same genotype at the first locus. For the second and third locus we obtain $\text{PI} = 0.1094$ and 0.0293 , respectively. The computation of PI for each locus assumes independence among alleles at the locus, and thus, that genotypes are in Hardy–Weinberg proportions. This assumption may be violated if mating is not random or if strong selection occurs.

Second, we use the product rule to multiply probabilities across loci to obtain a multi-locus PI. Thus, in our example, PI equals $0.375 \cdot 0.1094 \cdot 0.0293 \cong 0.001$. This second step assumes independence among alleles from different loci. This assumption may be violated in natural populations because associations among loci are often generated by new mutations, immigration (of individuals with distinct alleles), population substructure (due to isolation by distance or social behaviour), genetic drift (especially population bottlenecks), and natural selection. If non-random associations exist among alleles at or between loci, the actual observed PI will generally be higher than the theoretical PI computed with the equations assuming independence.

Another way to view the problem of non-independence among alleles and loci is through shared ancestry. Multi-locus genotypes within small natural populations may not be independent because of shared ancestry within families or within partially isolated subpopulations. For example, individuals within a family group are more likely to have identical genotypes than individuals drawn randomly from the entire population. If many large families (or subpopulations) exist within a population, then many more individuals will have identical genotypes than is estimated from the PI equations using the product rule and assuming no shared ancestry. Thus, in many natural populations the theoretical equations for PI will underestimate the true probability of finding identical genotypes.

Magnitude and consequences of bias

The potential magnitude of the underestimation bias has been quantified (Waits, Luikart & Taberlet, unpublished data) using large microsatellite data sets from 64 Montana wolves (*Canis lupus*; Forbes & Boyd 1997), 84 Scandinavian brown bears (*Ursus arctos*; Waits *et al.*, unpublished data), and 50 endangered Australian wombats (*Lasiorynus krefftii*; Taylor *et al.*, 1994; Taylor, 1995). For these data sets, the actual observed PI was computed as the proportion of all pairwise individual genotypes that were identical using a given number of loci (1, 2, 3, etc.). When using 3–10 loci, a bias of 1–3 orders of magnitude was observed in the data sets. Far less bias was observed in three other wolf data sets and in three other bear data sets. The bias mostly arose from close relatives (e.g. sibs) sharing identical genotypes at numerous loci. The samples in these studies come from populations that are apparently in

Hardy–Weinberg proportions. The wombat population consists of only ~ 100 individuals that remain in this highly endangered species. The Montana wolf population was recently established through recolonization by individuals from Canada (probably from multiple source populations). The bears are from one of four interconnected populations that have increased in size since a decline in the late 1800s.

The consequences of the bias can be severe in wildlife forensics cases where, for example, a biased PI estimator could underestimate (e.g. 0.0001 instead of 0.01) the probability that a hunter/collector possesses an illegally killed animal (or animal tissue). In this example, an innocent hunter could be fined or jailed. In a second example, researchers may want to estimate a population’s census size using DNA ‘fingerprints’ as genetic tags, and traditional capture-recapture methods. Such estimation is feasible without ever seeing the animals of interest, thanks to non-invasive sampling (Woods *et al.*, 1999; Kohn *et al.*, 1999). However, if a population contains close relatives, researchers could mistakenly identify a new capture as a recapture when, for example, an individual trapped during the recapture trapping period has a genotype identical to a different individual marked during the initial capture period. The potential impact of such errors on estimates of population size (N_c) has been quantified in a computer simulation study by Mills *et al.* (in press). They found that the failure to resolve individuals can cause an underestimation of N_c and can change confidence intervals on the estimate.

Avoiding bias problems

One solution for insuring that all individuals are identified is to use enough highly-polymorphic loci that the PI will be sufficiently small, even when substantial bias in PI exists (due to shared ancestry). The problem of bias is most likely to occur in highly substructured populations and especially in populations containing many siblings (Donnelly, 1995). Thus, a guideline for the number of loci needed to achieve a sufficiently low PI can be obtained by estimating the PI for sibs. The PI for sibs represents the upper limit on the possible range of PIs in a population. It also gives an upper limit for the number of loci needed to have a high probability of resolving all individuals. The equation for the PI among randomly sampled sibs is:

$$PI_{sibs} = 0.25 + (0.5 \sum p_i^2) + [0.5(\sum p_i^2)^2] - (0.25 \sum p_i^4) \quad (3)$$

This equation, like equations 1 & 2, assumes random sampling of individuals and independence among alleles within and between loci and no shared ancestry. Figure 3 suggests that approximately 14 loci (with $H=0.6$) are required to achieve a very low probability (0.0001) of finding two sibs with identical genotypes, while approximately seven loci are required to achieve a very low probability of finding two random individuals with identical genotypes. For capture-recapture studies, a sufficiently low PI may be only 0.01. In this case, 4–8 loci should be generally sufficient for resolving random individuals and sibs, respectively (if $H=0.6$). Using eight loci would be conservative for populations that do not contain many sibs. The Montana wolf population contains numerous sibs and we found that the theoretical PI (equations 1 & 2) was biased by up to three orders of magnitude (Waits, Luikart & Taberlet, submitted). In all the data sets from wolves, bears and wombats, the

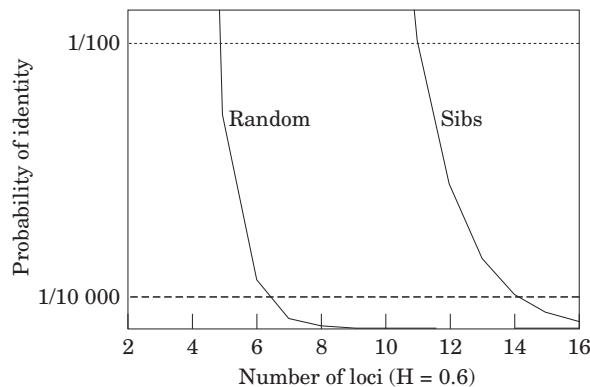


Figure 3. Relationship between the number of loci ($H=0.6$) and the probability of finding two identical genotypes. All loci have five alleles with frequencies 0.59, 0.2, 0.1, 0.07, and 0.04. These allele frequencies were chosen to achieve $H=0.6$, which is typical of microsatellite loci in many species. (Note: if we use loci with three or 10 alleles, the PI is very similar, as long as $H \cong 0.6$.)

actual *observed* PI was always between the two theoretical PIs (PI and PI_{sibs}). Determining the PI that is ‘sufficiently low’ depends on (i) the abundance of sibs (or shared ancestry) in the population and (ii) the severity of the consequences of failing to differentiate individuals. For capture-recapture studies, the consequences of not differentiating all individuals will generally be less severe than in forensics investigations which may require a PI of ≤ 0.0001 .

A second solution to the bias problem would be to calculate the PI using a formula that accounts for shared ancestry. Such formulae have been derived from mathematical models (Balding & Nichols, 1994). These formulae are widely applicable but require estimates of shared ancestry. The amount of shared ancestry can be estimated for many natural populations. But the estimation requires large sample sizes, and it may still be imprecise (Balding & Nichols, 1994).

PERSPECTIVES AND GUIDELINES

In conservation genetics and behavioural ecology, there is strong pressure on field biologists to use non-invasive genetic sampling methods. These methods do not require capturing the animal, thus reducing the risk of injuries to an endangered species, and the risk of disturbing the social group in behavioural studies. These arguments are sufficient to predict increasing use of non-invasive methods in the near future. However, recent studies have demonstrated that non-invasive sampling methods provide not only advantages, but also potentially serious drawbacks in terms of reliability, if applied without extreme care. Moreover, there are substantial differences among species concerning the amount and the quality of the DNA that can be extracted from hair follicles or faeces. For instance, in our laboratory, wolf faeces provides much more target DNA than bear faeces. This makes generalizations between species hazardous. Therefore, before conducting extensive non-invasive sampling, we strongly advise testing if this sampling method is compatible with the scientific goal—if, in other words, it will provide enough amplifiable DNA for the

desired genetic analyses. Not assessing the potential of non-invasive sampling prior to extensive analysis can lead to the collection of useless samples, and delay the achievement of the study by many years. A pilot study should assess both the technical difficulties, and the PI, and could be carried out in three steps: (i) estimation of the PI; (ii) simulations accounting for the PI in order to determine the genotyping error rate that is compatible with the research; (iii) laboratory experiments to check if it is technically possible to reach such a error rate.

The first step of the pilot study is to estimate the PI in the population under study. We suggest estimating both the observed PI (if >50 multilocus genotypes are available) and the theoretical expected PI for sibs, which represents the upper limit on the possible range of PIs in a population (see equation 3). For this purpose, it is necessary to know at least the allele frequencies or the heterozygosity of the genetic markers in the population. As allele frequencies and heterozygosity can vary greatly among populations, it could be risky to apply the observed heterozygosity of one population to another, particularly if the population under study is small and has experienced genetic drift. The ideal solution would be to estimate the level of heterozygosity for each marker by analysing 20 to 50 individuals belonging to the population under study.

The second step of the pilot study is to quantify the combined error rate (via simulations) of both the technical error and statistical bias to assess their impact on the results. For this, a simulated population is generated with a known PI, a realistic mating system, and a known level of heterozygosity per locus. Random samples of the simulated populations are drawn and the corresponding genotypes modified according to the technical error rates. These modified genotypes are then analysed using the statistical analyses of interest, and compared with the results obtained without introducing errors. The bias caused by the technical errors can then be estimated, and a threshold deduced for the technical error rate that is compatible with the scientific question. For example, a genotyping error rate of 5% may be compatible with estimating population size via non-invasive mark-recapture, whereas an error rate of less than 1% may be required for studies of relatedness (e.g. paternity exclusion).

The third step is to estimate the actual technical error rate via laboratory PCR experiments. This is the most time-consuming and difficult part of a pilot study. The main purpose is to prove that the protocol used provides multi-locus genotypes with an error rate below the acceptable threshold deduced from the simulations. Until now, scientists simply have checked that a few genotypes obtained with a non-invasive approach match the genotype obtained from a blood or a tissue sample. This strategy does not allow the detection of low error rates and thus cannot be considered as a pilot study. An error rate of about 1% per locus may correspond to an error rate of 10% for identifying individuals with a ten locus genotype. Consequently, the pilot study should allow precise estimation of very small per-locus error rates, and for this purpose several hundred PCR amplifications must be conducted. If the single tube error rate is above the threshold deduced from the pilot study, then a multiple tube approach can be implemented (Taberlet *et al.*, 1996) to reduce the error rate to an acceptable level. If the technical error rate cannot be below the threshold deduced from the simulations, then the use of a non-invasive approach is not appropriate for the scientific study. As a consequence, either the non-invasive approach is simply abandoned, or more technical adjustments are required to reduce the error rate. These technical adjustments could involve

multiplexing during the PCR in order to use more efficiently all the template DNA, to extract DNA from more material, or to choose other genetic markers with a higher level of heterozygosity. Another goal of the pilot study is to assess the respective occurrence of both types of genotyping errors, allelic dropout and false alleles. Indeed, these two error types have different impacts on individual identification and on the subsequent analyses.

In studies where there is a risk of genotyping errors or of misidentifying individuals, it is important to validate the genetic identification by using independent data for at least some individuals. The reliability of identification (e.g. genetic tagging) can be assessed by data on gender (using Y-specific markers or hormones in scats), natural markings, or track sizes observed in the field. For example, the same genotype should never be obtained for two individuals with different track sizes or of different sexes.

The choice of the genetic markers is extremely important, as it has consequences for all subsequent analyses. The ideal genetic marker should exhibit high heterozygosity (e.g. 0.6–0.8), should provide good DNA amplification, and should be easy to score on the electrophoresis gels. By choosing genetic markers with higher heterozygosity, the total number of loci required to reach the desired PI can be reduced, as can the cost of the analysis and the risk of erroneous individual identification. Some loci provide enough PCR products when using large amounts of template DNA, but are difficult to optimize in extreme conditions with little template DNA, and are often resistant to multiplexing. Dinucleotide microsatellites are useful because they have relatively short alleles, and are therefore easier to amplify from degraded DNA. However, tri- or tetranucleotide microsatellite loci are easier to score on gels than dinucleotide loci because of the reduced number and intensity of shadow bands below each allele. Indeed, shadow bands make difficult the distinction of adjacent alleles (alleles differing by only one repeat), particularly when using little template DNA, as the intensity of each allele can vary from one experiment to another (see Goossens, Waits & Taberlet, 1998; Taberlet *et al.*, 1996, 1997). Furthermore, the occurrence of false alleles (see above) seems to be connected with the intensity of the shadow bands. Consequently, the use of tri- or tetranucleotide microsatellites should considerably reduce this type of error.

The large variance in DNA content among samples represents another difficulty of non-invasive genetic sampling. If some DNA extracts provide enough DNA for obtaining a PCR product, but not enough DNA for obtaining a correct genotype (Fig. 2), then these samples should be either disregarded or analysed with the multiple-tube approach. To identify such problematic samples (and to quantify the variance in DNA quality among samples), researchers should perform a multiple tube analysis (Taberlet *et al.*, 1996) on at least one heterozygous locus (with relatively long PCR products). If allelic dropout is observed during this experiment, or if the relative intensity of the two alleles is not constant (Goossens *et al.*, 1998), then the results obtained at other loci might not be reliable.

In light of the rapidly growing interest in non-invasive genetic sampling, it is imperative that researchers fully understand the difficulties that were somewhat unexpected only a few years ago. The future challenges of non-invasive genetic sampling are monitoring for errors and precisely quantifying the confidence level of the results. Any results obtained via a non-invasive sampling should be considered as wrong, unless an appropriate pilot study has demonstrated that they are reliable. This calls for a different behaviour among molecular biologists, who may sometimes

be too confident in results obtained on a gel. Fortunately, the scientific community is beginning to appreciate the difficulties of non-invasive sampling, and to propose viable solutions. The future of non-invasive genetic sampling is thus quite promising.

ACKNOWLEDGEMENTS

We thank J.-M. Cornuet for deriving equation 3, and J. Bouvet, L. Gielly, C. Dubois-Paganon, V. Curri, S. Luche, H. Mattock, C. Beckhelling, S. Griffin, E. Uhrès, S. Questiau, N. Escaravage, V. Manceau, B. Goossens, L. P. Waits, M.-P. Biju-Duval, O. Disson, and J.-J. Camarra who helped us develop non-invasive genetic sampling techniques. GL was supported by a grant from NSF/NATO.

REFERENCES

- Awise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA. 1979.** Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences, U.S.A.* **76**: 6694–6698.
- Awise JC, Lansman RA, Shade RO. 1979.** The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus *Peromyscus*. *Genetics* **92**: 279–295.
- Balding D, Nichols R. 1994.** DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* **64**: 125–140.
- Béraud-Colomb E, Roubin R, Martin J, Maroc N, Gardeisen A, Trabuchet G, Goossens M. 1995.** Human β -globin gene polymorphism characterized in DNA extracted from ancient bones 12,000 years old. *American Journal of Human Genetics* **57**: 1267–1274.
- Birch DE. 1996.** Simplified hot start PCR. *Nature* **381**: 445–446.
- Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, Van Der Noordaa J. 1990.** Rapid and simple method for purification of nucleic acids. *Journal of Clinical Microbiology* **28**: 495–503.
- Briker J, et al. 1996.** Purification of high quality DNA from shed skin. *Herpetology Review* **27**: 133.
- Brown WM, Wright JW. 1979.** Mitochondrial DNA analyses and the origin and relation age of parthenogenetic lizards (genus *Cnemidophorus*). *Science* **203**: 1247–1249.
- Donnelly P. 1995.** Nonindependence of matches at different loci in DNA profiles: quantifying the effect of close relatives on the match probability. *Heredity* **75**: 26–34.
- Erlich HA, Gelfand D, Sninsky JJ. 1991.** Recent advances in the polymerase chain reaction. *Science* **252**: 1643–1651.
- Forbes SH, Boyd DK. 1997.** Genetic structure and migration in native and reintroduced Rocky Mountain wolf populations. *Conservation Biology* **11**: 1226–1234.
- Foucault F, Praz F, Jaulin C, Amor-Gueret M. 1996.** Experimental limits of PCR analysis of (CA)_n repeat alterations. *Trends in Genetics* **12**: 450–452.
- Frantzen M, Silk J, Fergusson J, Wayne R, Kohn M. 1998.** Empirical evaluation of preservation methods for faecal DNA. *Molecular Ecology* **7**: 1423–1428.
- Gagneux P, Boesch C, Woodruff DS. 1997.** Furtive mating in female chimpanzees. *Nature* **387**: 358–359.
- Gagneux P, Woodruff DS, Boesch C. 1997.** Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology* **6**: 861–868.
- Gerloff U, Schlötterer C, Rassmann K, Rambold I, Hohmann G, Fruth B, Tautz D. 1995.** Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology* **4**: 515–518.
- Goossens B, Waits LP, Taberlet P. 1998.** Plucked hair samples as a source of DNA: reliability of dinucleotide microsatellite genotyping. *Molecular Ecology* **7**: 1237–1241.

- Höss M, Pääbo S. 1993.** DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research* **21**: 3913–3914.
- Kebelmann-Betzing C, Seeger K, Dragon S, Schmitt G, Moricke A, Schild TA, Henze G, Beyermann B. 1998.** Advantages of a new *Taq* DNA polymerase in multiplex PCR and time-release PCR. *Biotechniques* **24**: 154–158.
- Kendall M, Stewart A. 1977.** *The advanced theory of statistics*. Volume 1. New York: Macmillan.
- Kohn M, Knauer F, Stoffella A, Schröder W, Pääbo S. 1995.** Conservation genetics of the European brown bear—a study using excremental PCR of nuclear and mitochondrial sequences. *Molecular Ecology* **4**: 95–103.
- Kohn MH, York EC, Karmradt DA, Haught G, Sauvajot RM, Wayne RK. 1999.** Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London* **B266**: in press.
- Kwok L. 1990.** Procedures to minimize PCR-product carry-over In: Innis MA, Gelfand DH, Sninski JJ, White TJ, eds. *PCR protocols, a guide to methods and applications* San Diego, California: Academic Press, 142–145.
- Lewontin R. 1991.** 25 years ago in genetics—electrophoresis in the development of evolutionary genetics—milestone or millstone. *Genetics* **128**: 657–662.
- Marsden JE, May B. 1984.** Feather pulp: a non-destructive sampling technique for electrophoretic studies of birds. *The Auk* **101**: 173–175.
- Mills L, Citta J, Lair K, Schwartz M, Tallmon D. 1999.** Estimating demographic rates of rare or elusive species using non-invasive DNA sampling. *Ecological Applications* (in press).
- Morin PA, Moore JJ, Chakraborty R, Jin L, Goodall J, Woodruff DS. 1994.** Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**: 1193–1201.
- Morin PA, Wallis J, Moore JJ, Chakraborty R, Woodruff D. 1993.** Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees. *Primates* **34**: 347–356.
- Morin PA, Woodruff DS. 1996.** Noninvasive genotyping for vertebrate conservation. In: Wayne RK, Smith TB, eds. *Molecular genetic approaches in conservation*. Oxford: Oxford University Press, 298–313.
- Mullis KB, Faloona FA. 1987.** Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods in Enzymology* **155**: 335–350.
- Murphy R, Sites JWJ, Buth D, Haufler C. 1996.** Proteins: isozyme electrophoresis. In: Hillis D, Moritz C, Mable B, eds. *Molecular Systematics* Sunderland, MA: Sinauer Associates, Inc, 51–120.
- Navidi W, Arnheim N, Waterman MS. 1992.** A multiple-tube approach for accurate genotyping of very small DNA samples by using PCR: statistical considerations. *American Journal of Human Genetics* **50**: 347–359.
- Paetkau D, Strobeck C. 1994.** Microsatellite analysis of genetic variation in black bear populations. *Molecular Ecology* **3**: 489–495.
- Ramekers J, Hummel S, Herrmann B. 1997.** How many cycles does a PCR need? Determination of cycle numbers depending of the number of targets and the reaction efficiency factor. *Naturwissenschaften* **84**: 259–262.
- Reed J, Tollit D, Thompson P, Amos W. 1997.** Molecular scatology: the use of molecular genetic analysis to assign species, sex and individual identity to seal faeces. *Molecular Ecology* **6**: 225–234.
- Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988.** Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Saiki RK, Scharf SF, Mullis KB, Horn GT, Erlich HA, Arnheim N. 1985.** Enzymatic amplification of β -globin genomic sequences and restriction site analysis for diagnostic of sickle cell anemia. *Science* **230**: 1350–1354.
- Schmutz S, Schmutz J, Simpson N, Tabel H. 1987.** Serum protein polymorphism in common eiders detected by isoelectric focusing. *Biochemical Genetics* **25**: 191–195.
- Seutin G, White B, Boag P. 1991.** Preservation of avian blood and tissue samples for DNA analysis. *Canadian Journal of Zoology* **69**: 82–90.
- Singer-Sam J, Tanguay RL, Riggs AD. 1989.** Use of Chelex to improve the PCR signal from a small number of cells. *Amplifications* **3**: 11.
- Stoneking M. 1995.** Ancient DNA: how do you know when you have it and what can you do with it? *American Journal of Human Genetics* **57**: 1259–1262.
- Stratil A, Glasnak V, Bobak P, Cizova D, Gabrisova E, Kalab P. 1990.** Variation of some serum proteins in red deer, *Cervus elaphus* L. *Animal Genetics* **21**: 285–293.

- Taberlet P, Bouvet J. 1992.** Bear conservation genetics. *Nature* **358**: 197.
- Taberlet P, Bouvet J. 1994.** Mitochondrial DNA polymorphism, phylogeography, and conservation genetics of the brown bear (*Ursus arctos*) in Europe. *Proceedings of the Royal Society of London* **B255**: 195–200.
- Taberlet P, Camarra J-J, Griffin S, Uhrès E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J. 1997.** Non-invasive 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* **26**: 3189–3194.
- Taylor AC, Sherwin WB, Wayne RK. 1994.** Genetic variation of microsatellite loci in a bottlenecked species: the hairy-nosed wombat (*Lasiornhinus krefftii*). *Molecular Ecology* **3**: 277–290.
- Taylor AC. 1995.** Molecular ecology of the endangered northern hairy-nosed wombat *Lasiornhinus krefftii*, and applications to conservation and management. Ph.D. Dissertation, University of New South Wales, Sydney, Australia.
- Tikel D, Blair D, Marsh H. 1996.** Marine mammal faeces as a source of DNA. *Molecular Ecology* **5**: 456–457.
- Valsecchi E, Glockner-Ferrari D, Ferrari M, Amos W. 1998.** Molecular analysis of the efficiency of sloughed skin sampling in whale population genetics. *Molecular Ecology* **7**: 1419–1422.
- Walsh PS, Metzger DA, Higuchi R. 1991.** Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques* **10**: 506–513.
- Wasser S, Houston C, Koehler G, Cadd G, Fain S. 1997.** Techniques for application of faecal DNA methods to field studies in Ursids. *Molecular Ecology* **6**: 1091–1097.
- White TJ, Arnheim N, Erlich HA. 1989.** The polymerase chain reaction. *Trends in Genetics* **5**: 185–189.
- Woods J, Paetkau D, Lewis D, McLellan B, Proctor M, Strobeck C. 1999.** Genetic tagging free-ranging black and brown bears. *Wildlife Society Bulletin* (in press).
- Wright PA, Wynford-Thomas D. 1990.** The polymerase chain reaction: miracle or mirage? A critical review of its uses and limitations in diagnosis and research. *Journal of Pathology* **162**: 99–117.