Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion

J. I. HOFFMAN and W. AMOS

Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

Abstract

Microsatellite genotyping errors will be present in all but the smallest data sets and have the potential to undermine the conclusions of most downstream analyses. Despite this, little rigorous effort has been made to quantify the size of the problem and to identify the commonest sources of error. Here, we use a large data set comprising almost 2000 Antarctic fur seals *Arctocephalus gazella* genotyped at nine hypervariable microsatellite loci to explore error detection methods, common sources of error and the consequences of errors on paternal exclusion. We found good concordance among a range of contrasting approaches to error-rate estimation, our range being 0.0013 to 0.0074 per single locus PCR (polymerase chain reaction). The best approach probably involves blind repeat-genotyping, but this is also the most labour-intensive. We show that several other approaches are also effective at detecting errors, although the most convenient alternative, namely mother–offspring comparisons, yielded the lowest estimate of the error rate. In total, we found 75 errors, emphasizing their ubiquitous presence. The most common errors involved the mis-interpretation of allele banding patterns (*n* = 60, 80%) and of these, over a third (*n* = 22, 36.7%) were due to confusion between homozygote and adjacent allele heterozygote genotypes. A specific test for whether a data set contains the expected number of adjacent allele heterozygotes could provide a useful tool with which workers can assess the likely size of the problem. Error rates are also positively correlated with both locus polymorphism and product size, again indicating aspects where extra effort at error reduction should be directed. Finally, we conducted simulations to explore the potential impact of genotyping errors on paternity exclusion. Error rates as low as 0.01 per allele resulted in a rate of false paternity exclusion exceeding 20%. Errors also led to reduced estimates of male reproductive skew and increases in the numbers of pups that matched more than one candidate male. Because even modest error rates can be strongly influential, we recommend that error rates should be routinely published and that researchers make an attempt to calculate how robust their analyses are to errors.

Keywords: adjacent allele heterozygote, allelic dropout, Antarctic fur seal, *Arctocephalus gazella*, genotyping error rate, paternity, pinniped, scoring error

Received 13 September 2004; revision received 27 October 2004; accepted 2 November 2004

Introduction

Microsatellite markers are a prominent tool in ecological, medical and forensic genetics (Queller *et al.* 1993; Jarne & Lagoda 1996; Luikart & England 1999). However, microsatellite genotyping can be error-prone and consequently few data sets, especially those typical of large-scale studies

Correspondence: Joseph I. Hoffman, Fax: +44 1223 336676; E-mail: jih24@cam.ac.uk

of natural populations, are likely to be perfect (Marshall *et al.* 1998). With a number of recent studies showing that even modest error rates can seriously perturb estimates of genetic diversity, population size and structure, migration rates, kinship and parentage (Marshall *et al.* 1998; Taberlet *et al.* 1999; Creel *et al.* 2003; Piggott & Taylor 2003; Waits & Leberg 2003), the importance of genotyping errors has become increasingly recognized. Despite this, we know of very few studies that have looked systematically at the origin and frequency of genotyping errors.

Microsatellite genotyping errors may arise in a number of ways. When the template DNA is of low quantity and/or quality, as is typical of studies employing noninvasive tissue-sampling, PCR (polymerase chain reaction) amplification can become unreliable (Gerloff *et al*. 1995; Taberlet *et al*. 1996; Gagneux *et al*. 1997a). Here, a common problem is the stochastic failure of one allele to amplify, leading to heterozygotes appearing to carry only one allele, referred to as 'allelic dropout' (Navidi *et al*. 1992; Walsh *et al*. 1992; Gerloff *et al*. 1995; Taberlet *et al*. 1996; Gagneux *et al*. 1997a). Another source of artefact is 'misprinting', in which amplification products are generated that can be misinterpreted as true alleles (Taberlet *et al*. 1996; Goossens *et al*. 1998; Bradley & Vigilant 2002). The frequency of such errors can exceed 0.25 per reaction (Taberlet *et al*. 1996; Gagneux *et al*. 1997a; Morin *et al*. 2001; Frantz *et al*. 2003 although see also Wasser *et al*. 1997; Kohn *et al*. 1999; Ernest *et al*. 2000; Sloane *et al*. 2000; Parsons 2001; Segelbacher 2002; Fernando *et al*. 2003). Consequently, numerous quality control protocols have been developed, including the adoption of multiple tube approaches where DNA samples are amplified independently several times (Navidi *et al*. 1992; Taberlet *et al*. 1996; Frantz *et al*. 2003), comparison of genotypes obtained with those from matched blood or tissue (Wasser *et al*. 1997; Kohn *et al*. 1999; Ernest *et al*. 2000; Sloane *et al*. 2000; Parsons 2001; Fernando *et al*. 2003), strategic re-amplification at loci likely to harbour errors (Miller *et al*. 2002), prescreening of samples for DNA quantity (Morin *et al*. 2001; Segelbacher 2002) and the use of pilot studies (Taberlet & Luikart 1999) and simulations (Taberlet *et al*. 1996; Valiere *et al*. 2002).

Even when large quantities of high-quality DNA are extracted from blood or tissue, genotyping errors still occur. These include allele non-amplification resulting from primer binding site mutation ('null alleles' Callen *et al*. 1993; Pemberton *et al*. 1995; Dakin & Avise 2004), errors due to electrophoresis artefacts (Fernando *et al*. 2001; Davidson & Chiba 2003), mis-scoring of allele banding patterns, and data entry and other clerical errors. Of these, the most important source of error is probably the incorrect calling of alleles on autoradiographs or fluorescent profiles. In particular, the presence of 'stutter bands', generated by slippage of *Taq* polymerase during PCR, can make it difficult to score alleles reliably (Litt *et al*. 1993; Ginot *et al*. 1996; Harker 2001; Johansson *et al*. 2003), especially when there are large signal intensity differences between alleles and/or the lengths of two alleles in a heterozygous individual differ by only a few nucleotides. Unfortunately, these problems remain even when genotyping is automated, as genotypes are either accepted blind or, more usually, corrected manually (Ginot *et al*. 1996; Ewen *et al*. 2000).

Given the considerable scope for errors to arise under most circumstances, it is perhaps surprising that few studies beyond those utilizing noninvasive sampling have sought to determine either the prevalence or sources of microsatellite genotyping errors. Exceptions come largely from the field of human medical genetics and include Brzustowicz *et al*. (1993), Ghosh *et al*. (1997), Ewen *et al*. (2000), Sobel *et al*. (2002) and a small number of among-lab reproducibility trials (e.g. Jones *et al*. 1997; Weeks *et al*. 2002). Furthermore, error rates are rarely published and where they are expressed the terminology varies greatly, from the proportion of PCR reactions yielding at least one incorrect allele to the proportion of alleles that are incorrect, and from the error rate per locus to that across multiple loci (e.g. Ewen *et al*. 2000; Sobel *et al*. 2002; Hoffman *et al*. 2003). These and other studies (Brzustowicz *et al*. 1993; Ginot *et al*. 1996; Ghosh *et al*. 1997; Palsbøll *et al*. 1997; Weeks *et al*. 2002) report error rates between 0.001 and 0.127 per reaction.

Modest error rates may seem inconsequential, but are they? Simple calculations suggest that the problem may be far from trivial. For example, a 1% error rate in allele calling would lead to almost a quarter of 12-locus genotypes containing at least one error. Worse, the point of generating genotypes is usually to compare them with others. With this error rate, only about 62% of comparisons between the same individual typed twice would show genotype identity. With a 2% error rate, the probability of obtaining the same genotype twice for the same individual falls to below 40%. Similarly, estimates of relatedness, especially parentage, can be heavily impacted by errors, especially when candidate fathers are excluded on the basis of only a single mismatch (Marshall *et al*. 1998; Taberlet *et al*. 1999). Consequently, whenever unexpectedly few paternities are assigned, leading to the inference of extra pair, group or population matings, it is important that error rates are determined and modelled so that their impact can be assessed. Examples of where genotyping errors are thought to have led to misleading conclusions include a study of chimpanzees by Gagneux *et al*. (1997b) that reported frequent 'furtive' matings by females outside their social groups, and a study of polygynous Antarctic fur seals *Arctocephalus gazella* (Gemmell *et al*. 2001) that assigned paternity to fewer than a quarter of offspring. In both cases, later studies suggest that these unexpected findings were likely due, at least in part, to poor quality genotype data (Constable *et al*. 2001; Gagneux *et al*. 2001; Vigilant *et al*. 2001; Hoffman *et al*. 2003).

Fortunately, there are many ways to identify genotyping errors and estimate their rates. Most obviously, a subset of individuals can be regenotyped and compared, although to obtain a sufficiently representative sample this can involve a large extra experimental effort (Brzustowicz *et al*. 1993; Ghosh *et al*. 1997; Ewen *et al*. 2000). More economical therefore is to conduct statistical tests on the data that already exist. One commonly used test is for deviations from Hardy–Weinberg equilibrium (Gomes *et al*. 1999) that

reveal the homozygous excess resulting from either null alleles or allelic dropout. Further verification can be achieved by comparing known mother–offspring pairs and looking for mismatches (Marshall *et al*. 1998), although this may be complicated by the possibility of fostering or egg-dumping, particularly if practised by close kin. Where resampling of animals is common, close but imperfect matches may be so unlikely that they are best explained as genotyping errors and hence can be examined to identify problems (Palsbøll *et al*. 1997). Two further approaches are less often used. First, in the same way that close identity matches can be rechecked on the original autoradiographs/traces, so too can paternity matches. Second, because one of the most likely sources of mis-scoring involves confusion between homozygotes and heterozygotes where the alleles differ by a single repeat unit (Fernando *et al*. 2003), these particular genotypes can be scrutinized.

Here we use a large data set comprising almost 2000 Antarctic fur seals genotyped at nine highly variable microsatellite loci to explore methods of error detection, identify common sources of error, and explore the consequences of errors for paternal exclusion analysis.

Materials and methods

Study site, animal identification and tissue sampling

Tissue samples were collected during the austral summers of 1994/1995–2000/2001 at a small fur seal breeding beach located at Bird Island, South Georgia (54°00′-S, 38°02′-W), where a scaffold walkway (Doidge *et al*. 1984) enabled all animals to be observed and tissue-sampled with minimum disturbance. Adult females, identified using plastic tags (Dalton Supplies) placed in the trailing edge of the fore-flipper, were tissue-sampled from the interdigital margin of the foreflipper using piglet-ear notching pliers (Majluf & Goebel 1992). Pups born to tagged females were sampled in the same way. Adult males occupying territories on the beach and on the surrounding rocks were too large and aggressive to be captured, and were instead individually marked using small patches of gloss paint (Arnould & Duck 1997) and remotely sampled using a biopsy dart system (Gemmell & Majluf 1997). All sampling equipment was sterilized using ethanol between uses. A total of 1763 tissue samples were collected as part of an ongoing study of paternity, comprising 375 adult females, 718 pups and 670 adult males (Hoffman *et al*. 2003). These samples were stored individually in the preservative buffer 20% dimethyl sulphoxide (DMSO) saturated with salt (Amos & Hoelzel 1991) and stored at −20 °C.

DNA extraction and microsatellite amplification

Total genomic DNA was extracted using an adapted Chelex 100 protocol (Walsh *et al*. 1991). All samples were then genotyped using a panel of nine dinucleotide-repeat microsatellite loci, previously isolated from a variety of pinniped species (Table 1). These loci were chosen because

Table 1 Summary of microsatellite loci used in this study, including literature sources and polymorphism characteristics for the Antarctic fur seals genotyped

Locus	Isolated from species	Reference	Number of alleles	Size range (bp)	H_E	T_1 (°C)	T_2 (°C)
Aa4	South American fur seal *Arctocephalus australis*	(Gemmell *et al*. 1997)	7	204–220	0.747	46	48
Hg1.3	Grey seal *Halichoerus grypus*	(Gemmell *et al*. 1997)	13	234–276	0.868	42	46
Hg6.3	Grey seal *Halichoerus grypus*	(Allen *et al*. 1995)	16	213–245	0.858	46	48
Hg8.10	Grey seal *Halichoerus grypus*	(Allen *et al*. 1995)	5	162–184	0.450	42	46
Lw10	Weddell seal *Leptonychotes weddellii*	(Davis *et al*. 2002)	18	100–140	0.906	46	48
M11a	Southern elephant seal *Mirounga leonina*	(Hoelzel *et al*. 1999)	18	148–184	0.921	46	48
Pv9	Grey seal *Halichoerus grypus*	(Allen *et al*. 1995)	11	162–186	0.771	48	52
PvcA	Harbor seal *Phoca vitulina*	(Coltman *et al*. 1996)	9	137–153	0.774	46	48
PvcE	Harbor seal *Phoca vitulina*	(Coltman *et al*. 1996)	15	94–148	0.872	45	50
Overall			12.4	94–276	0.796	45	50

they exhibited clear banding patterns, did not deviate significantly from Hardy–Weinberg equilibrium and were highly polymorphic, yielding up to 18 alleles per locus. PCR reactions were carried out in 10 μL reaction volumes containing 1 μL template DNA, 1× Thermalase buffer (10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl$_2$, 0.1% Tween 20, 0.1% gelatine, 0.1% IGEPAL, Sigma, 60 mM tetra-methylammonium chloride (TMAC), 2.5% formamide, 0.1 mM dGTP, 0.1 mM dATP, 0.1 mM dTTP, 0.02 mM dCTP, 4 pmol of each primer, 0.25 units of *Taq* polymerase and 0.01 μCi [α^{32}P]-dCTP. Loci were amplified using the following PCR profile: one cycle of 120 s at 94 °C, 45 s at T$_1$, 50 s at 72 °C; 10 cycles of 30 s at 94 °C, 45 s at T$_1$, 50 s at 72 °C; 25 cycles of 30 s at 89 °C, 45 s at T$_2$, 50 s at 72 °C; and one final cycle of 5 min at 72 °C (see Table 1 for T$_1$ and T$_2$). To enable the rapid screening of large numbers of samples, 96-well microtitre plates were used. Each plate contained 88 individual samples, together with a pair of negative controls (no DNA) and three pairs of positive controls (standards with alleles of known size). PCR products were resolved by electrophoresis on standard 6% polyacrylamide sequencing gels, and detected by autoradiography. Exposed X-ray films were assessed and if required, a second exposure was made for an adjusted time period.

Microsatellite scoring

Microsatellites can sometimes be difficult to score because of the presence of stutter bands (Ginot *et al.* 1996; Harker 2001). For example, a potentially ambiguous pattern can be produced when two alleles of a heterozygote differ in size by one repeat unit, causing the stutter bands to overlap and produce adjacent bands of similar intensity. Here, we distinguished these 'adjacent allele heterozygotes' from homozygotes by comparing them with the banding profiles of known single alleles. Another potential source of error when scoring autoradiographs is to read them in the wrong orientation. For example, because autoradiographs are transparent, it is possible to mistakenly read samples from right to left instead of left to right. To guard against this, positive controls were loaded onto gels in asymmetric locations. Allele sizes were then determined by comparison with these standards. Genotypes were entered manually into a Microsoft Excel spreadsheet and were systematically double-checked with reference to the original autoradiographs.

Incorporation of existing data, gap-filling and quality control

Data were incorporated from a previous study (animals sampled during 1994 and 1995, genotyped at loci Aa4, Hg1.3, Hg8.10, M11a, PvcA and PvcE by Gemmell *et al.* 2001). For these data, the original autoradiographs were

rescored and whenever an uncertain score was encountered, the sample was genotyped again at that locus. Approximately 10 additional reactions from each plate were also regenotyped in order to determine unambiguously the orientation and location of samples on autoradiographs. Then, for the combined data set, any reactions that failed were repeated up to two times. In addition, all reactions yielding uncertain genotypes (e.g. with faint or unclear bands) were repeated. Finally, samples that failed to amplify at two or more loci (*n* = four pups and one adult male) were excluded from the data set. The final frequency of missing single locus genotypes in the data set was 0.0021.

Identification and quantification of genotyping errors

Repeat-genotyping.

First, we randomly selected approximately 10% of all samples (*n* = 204 individuals) and independently repeat-genotyped these at all nine loci. Unsuccessful reactions were not regenotyped, and consequently, the average number of samples that yielded duplicate genotypes was 190. Discrepancies between the two sets of genotypes were then examined case-by-case, by comparing the two sets of autoradiographs directly. These were then classified as follows: (i) 'adjacent allele heterozygote scoring error', resulting from confusion between homozygote and adjacent allele heterozygote genotypes; (ii) 'other scoring error' (the various types found are detailed in the results section); (iii) 'data input', where a mistake was made in transcribing genotypes into the excel spreadsheet; (iv) 'allele dropout', where a mismatch between the two genotypes was compatible with the non-amplification of a single allele; and (v) 'unknown', where the error could not be attributed to any of the above causes and could, for example, have resulted from sample mix-up, pipetting mistakes or contamination.

Where genotyping error rates are expressed in the literature, the terminology varies greatly, from the number of errors per allele to the number of errors per reaction, and from the error rate per locus to that across multiple loci. To enable comparison with other studies, we expressed error rates wherever possible as both the number of errors per allele and per reaction, and summarized these for each locus individually and across all loci. The great majority of genotyping errors involved scoring or typographical errors, which after correction yielded the same genotype as the duplicate. Thus, a specific genotype could usually be designated 'correct' or 'incorrect'. Consequently, we calculated error rate per reaction as the number of incorrect genotypes divided by the total number of reactions used for comparison. Similarly, error rate per allele was calculated as the number of incorrect alleles divided by the total number of alleles. When a discrepancy between two genotypes was consistent with the dropout of a single allele, we

counted this single allele as being mistyped, and when no cause could be found for the discrepancy, the number of discrepant alleles was given as the number that were mistyped. However, because it is possible that such comparisons could involve errors at more than one of the duplicate genotypes, this approach may slightly underestimate the true genotyping error rate.

Deliberately resampled individuals We next examined concordance among the genotypes of individuals that were deliberately resampled. These numbered 40 adult females that were deliberately resampled because of illegible, broken or lost flipper tags, and 50 adult males that were resampled between one and three times each (yielding a total of 107 samples) to verify identity because of fading paint marks. In all of these cases, because there was a strong a priori expectation that duplicate genotypes represented identical individuals, we were able to identify not only scoring and typographical errors but also a small number of cases that could be attributed to allele dropout or unknown causes.

Unintentionally resampled individuals Following the approach of Palsbøll *et al.* (1997), we also examined concordance among the genotypes of animals that were likely to have been unintentionally sampled more than once during the course of the study. Because it is possible for different individuals to have identical multilocus genotypes if an insufficient number of loci have been used, we first calculated the probability of identity (P_{ID}, Paetkau & Strobeck 1994) across all individuals and all loci. This was very low (1.354×10^{-12}), indicating that identical genotypes almost certainly represented individuals that were resampled. Because of the possibility of relatives being present in the population, we also took the conservative measure of calculating the P_{ID} among siblings ($P_{ID\text{-}Sib}$, Evett & Weir 1998). This was sufficiently low (1.20×10^{-4}) to distinguish even siblings with high confidence. These probabilities also remained low when a single locus was removed from the calculation (P_{ID} range = 3.31×10^{-12}–1.11×10^{-10}, $P_{ID\text{-}Sib}$ range = 1.94×10^{-4}–4.12×10^{-3}) suggesting that very few genotypes were expected to match by chance at all but one locus. Consequently, we identified duplicate genotypes, allowing one mismatch, using the program IDENTITY (Allen *et al.* 1995). However, in contrast to comparisons involving deliberately resampled individuals, there was no a priori expectation that certain individuals had been resampled. Therefore, we conservatively counted discrepancies as genotyping errors only when they could be attributed to scoring or data entry errors, and we did not include cases that invoked allele dropout or unknown causes.

Mother–offspring pairs Where relationships between individuals are known from field observational data, it is also possible to identify genotyping errors from mismatches (i.e. genotypes that do not share a common allele) between mothers and their offspring (Marshall *et al.* 1998). We checked 718 mother–pup pairs for mismatches using the program NEWPAT XL (Worthington Wilmer *et al.* 1999). As with deliberately resampled animals, because there was an a priori expectation that mother–offspring pairs mismatching at only a single locus were genuine, mismatching duplicate genotypes that could not be clearly ascribed to scoring or typographical error were assigned to the 'allele dropout' or 'unknown' categories. However, because a further 58 mother–offspring pairs genuinely mismatched at multiple loci due to the wrong pup being sampled in the field, these animals were excluded from the analysis. Defining an error as the replacement of the true genotype with a genotype selected at random under Hardy–Weinberg assumptions, the per-reaction error rate e_l for locus *l* can be estimated as follows:

$$e_l \approx \frac{1}{2P_l} \cdot \frac{m_l}{M_l}$$

where m_l is the number of mother–offspring mismatches, M_l is the number of mother–offspring pairs compared and P_l is the exclusion probability at that locus (Marshall *et al.* 1998). The underlying error rate may then be estimated as the average across *n* loci:

$$e \approx \frac{1}{n} \sum_{l=1}^{n} e_l$$

To avoid overestimation of the error rate, all mother–offspring pairs must be independent, and no single individual should be included both as an offspring and parent. Therefore, because our data set included a number of mothers with multiple offspring, we randomly chose a single pup from each of these individuals so that no mother was included more than once. To provide different estimates of the genotyping error rate, we first counted only mother–offspring pairs that mismatched at up to one locus in the analysis ($n = 258$). Then, to determine how the inclusion of genuinely mismatching mother–offspring pairs influenced the error rate, we included mother–offspring pairs that mismatched at multiple loci in the calculation (total $n = 298$).

Mismatches between pups and putative fathers In the same way that mismatches between mothers and their offspring can be rechecked on autoradiographs/traces, so too can mismatches between fathers and offspring. Consequently, we examined instances in which pups and candidate males mismatched at only one or two loci. The paternity analysis was conducted as described by Hoffman *et al.* (2003) using the program NEWPAT XL (Worthington Wilmer *et al.* 1999).

Resulting mismatches were checked by reference to the original autoradiographs, and those that could be attributed to scoring or typographical errors were classified as genotyping errors.

Influence of error rate on paternity assignment

To explore the potential impact of genotyping errors on the rate of paternity assignment, we conducted simulations using our data set. All of the errors that we identified in this study were rectified, effectively reducing the error rate to a negligible level. The resulting data file comprised 660 matching mother–pup pairs and 415 unique adult male genotypes, with 388 pups being assigned paternity. Into these data we then introduced errors by selecting alleles and replacing them with alleles selected randomly from the underlying allele frequency distribution. If the replacement allele was by chance the same as the original allele, further selections were not made. We then conducted a paternity analysis using NEWPAT XL (Worthington Wilmer *et al.* 1999), allowing a maximum of one unscored locus and no mismatches. For each run, we recorded the number of mother–offspring pairs that mismatched and the number of pups assigned a paternity. To determine how errors influenced our estimate of the variance in male reproductive success, we also recorded the number of pups assigned to each male. Error rates ranging from 0.001 to 0.02 per allele were examined, chosen to reflect values reported in the literature.

Results

To identify genotyping errors and estimate the underlying rate of error for our data set, we used the following methods.

Repeat-genotyping

Approximately 10% of all samples were regenotyped at all nine loci. Unsuccessful reactions were not repeated, yielding an average of 190 duplicate genotypes at each locus. The genotyping error rate combined over all loci and all samples was low, at 0.0038 per reaction or 0.0022 per allele, the breakdown being 13 reactions mistyped, accounting for 15 out of 6848 alleles incorrect (Table 2). Of these errors, seven were found in the original data set (two adult females, one adult male and four pups), with the remainder being found among the duplicate genotypes (three adult females, one adult male and two pups). Mistyping was largely the result of autoradiograph scoring and data entry errors (*n* = five and two cases respectively, see Table 5). Four additional instances were consistent with the dropout of a single allele during PCR amplification, yielding an estimate of the allele dropout rate of 0.0012

per reaction. The remaining two mismatches could not be explained, and could have arisen, for example, from sample mix-up, pipetting error or contamination.

Deliberately resampled individuals

We next examined concordance among the genotypes of individuals that were deliberately sampled more than once. First, we examined duplicate genotypes belonging to adult females that were resampled because of illegible, broken or lost flipper tags (*n* = 80 samples). Among these comparisons, we found two mistyped alleles out of 1440, providing a second estimate of the error rate of 0.0028 per reaction or 0.0014 per allele (Table 2). Both of these mistyped alleles were the result of errors made in the scoring of autoradiographs. A number of adult males were also deliberately resampled, both within and among seasons, to verify their identities because of fading paint marks. Again, we found a small number of mismatches (four alleles mistyped out of 959, error rate = 0.0042 per reaction or 0.0021 per allele, Table 2), all of which were due to scoring error.

Unintentionally resampled individuals

Using the program IDENTITY (Allen *et al.* 1995), we then identified animals that were likely to have been sampled more than once during the study. For each genotypic comparison, we allowed a mismatch at up to one locus. Of 670 adult male genotypes analysed, in addition to the 57 intentional matches previously described, we identified a further 198 unintentional resamplings. These involved 147 different individuals that were each resampled between one and four times, yielding a total of 345 genotypes. Among these genotypes, we found 23 mistyped alleles, equivalent to an error rate of 0.0074 per reaction or 0.0037 per allele (Table 2). No duplicates were found among the genotypes of either pups or adult females.

Parent–offspring error-checking

Next, we examined single-locus mismatches among mother–offspring pairs. Of 718 putative pairs, 645 matched perfectly, 15 were putative typing errors that mismatched at only a single locus (Table 3) and the remainder mismatched at multiple loci, probably representing cases where the pup genuinely did not belong to the supposed mother. On examination of the 15 possible typing errors, 10 were attributed to scoring error, one to a data entry error and another four were consistent with the non-amplification of a single allele (Table 5). In each of these cases, the error was found in the genotype of the pup.

    To estimate the genotyping error rate based upon mismatches between known mother–offspring pairs, we used

Table 2 Locus-specific and overall genotyping error rates estimated from concordance among duplicate genotypes. Owing to small numbers of missing single-locus genotypes, the number of reactions differs among loci

Method	Locus	Number of reactions	Number of mistyped reactions	Number of mistyped alleles	Error rate per reaction	Error rate per allele
Repeat-genotyping	Aa4	378	1	2	0.0026	0.0026
	Hg1.3	370	0	0	0	0
	Hg6.3	364	2	2	0.0055	0.0027
	Hg8.10	380	0	0	0	0
	Lw10	386	1	1	0.0026	0.0013
	M11a	382	2	2	0.0052	0.0026
	Pv9	400	3	4	0.0075	0.0050
	PvcA	380	2	2	0.0053	0.0026
	PvcE	384	2	2	0.0052	0.0026
	Overall	3424	13	15	0.0038	0.0022
Deliberately resampled females	Aa4	80	0	0	0	0
	Hg1.3	80	1	1	0.0125	0.0063
	Hg6.3	80	0	0	0	0
	Hg8.10	80	1	1	0.0125	0.0063
	Lw10	80	0	0	0	0
	M11a	80	0	0	0	0
	Pv9	80	0	0	0	0
	PvcA	80	0	0	0	0
	PvcE	80	0	0	0	0
	Overall	720	2	2	0.0028	0.0014
Deliberately resampled males	Aa4	107	1	1	0.0093	0.0047
	Hg1.3	107	1	1	0.0093	0.0047
	Hg6.3	105	0	0	0	0
	Hg8.10	107	1	1	0.0093	0.0047
	Lw10	107	0	0	0	0
	M11a	105	1	1	0.0095	0.0048
	Pv9	107	0	0	0	0
	PvcA	107	0	0	0	0
	PvcE	107	0	0	0	0
	Overall	959	4	4	0.0042	0.0021
Unintentionally resampled males	Aa4	345	2	2	0.0058	0.0029
	Hg1.3	345	6	6	0.0174	0.0087
	Hg6.3	338	5	5	0.0148	0.0074
	Hg8.10	341	2	2	0.0059	0.0029
	Lw10	345	3	3	0.0087	0.0043
	M11a	343	2	2	0.0058	0.0029
	Pv9	345	1	1	0.0029	0.0014
	PvcA	345	1	1	0.0029	0.0014
	PvcE	345	1	1	0.0029	0.0014
	Overall	3092	23	23	0.0074	0.0037

the program CERVUS (Marshall *et al*. 1998). Because individuals should not be counted more than once in the analysis, a single pup was chosen at random from each mother with multiple offspring. Considering only mother–offspring pairs that mismatched at one or fewer loci (n = 258), the error rate was estimated at 0.0013 per reaction (Table 4). However, by selecting single pups randomly, only three out of the 15 mismatches previously identified were included in the analysis. Therefore, to give the upper limit of the error rate, we included all 15 mismatching pups in the analysis while counting each mother only once, providing an estimate of 0.0064 per reaction. Finally, to investigate how the inclusion of genuinely mismatching mother–offspring pairs influenced the estimated error rate, we included mother-offspring pairs that mismatched at multiple loci in the calculation (total n = 298). The estimated error rate was this time substantially higher (0.0635 per reaction, Table 4), indicating that this approach is sensitive to mistakes made in the identification of mother–offspring pairs in the field.

Table 3 Locus-specific and overall numbers of single-locus mismatches among 660 mother–pup pairs

Locus	Number of reactions	Number of mistyped reactions	Number of mistyped alleles
Aa4	1316	1	1
Hg1.3	1316	2	2
Hg6.3	1296	1	1
Hg8.10	1320	0	0
Lw10	1316	2	2
M11a	1318	1	1
Pv9	1318	4	4
PvcA	1318	1	1
PvcE	1318	3	3
Overall	11 836	15	15

Next we examined mismatches between pups and their putative fathers. The paternity analysis (Hoffman *et al.* 2003) yielded 369 mismatch-free paternity assignments, plus 19 assignments that invoked genotyping errors (for breakdown, see Table 5). Because there was no a priori expectation of a given male matching any pup, we only counted instances where scoring or typographical errors were implicated, with the exception of a single case previously identified as a probable instance of allele dropout by repeat-genotyping. The genotypic mismatches that we identified corresponded to typing errors in the genotypes of 14 different pups, four adult males and a single female (allele dropout).

To what extent do different methods identify the same errors?

The above methods yielded a total of 75 mistyped reactions and 78 mistyped alleles. To determine how many of these errors were detected by more than one of the methods employed, errors were not initially rectified upon discovery. Only a single error was found by more than one method (a probable instance of allele dropout that was revealed by both repeat-genotyping and paternity mismatching). More subtly however, errors in the genotypes of adult males caused the number of unique candidate male genotypes to be overestimated. This was because, when an individual was sampled more than once and one of the genotypes contained an error, two genotypes were generated rather than one. Consequently, in paternity analysis using error-prone data where some males have been accidentally resampled, some pups will match the same male represented by two different genotypes. In our study, 21 pups (7.5%) matched more than one candidate male, but more than half of these ($n = 13$) disappeared following error rectification. Hence, when the probability of paternal exclusion is high, it may be worthwhile to inspect cases where offspring match more than one candidate male for genotyping errors.

Sources of genotyping error

Sources of genotyping error and their frequencies are summarized in Table 5. The majority of errors ($n = 72, 96\%$)

Table 4 Locus-specific and overall genotyping error rates estimated from mismatches between adult females and their putative offspring using the program CERVUS. The upper section of the table shows error rates estimated from comparisons among 258 mother–pup pairs, which yielded a small number of genuine scoring errors. For comparison, the bottom section shows error rates estimated using 298 putative mother-offspring pairs, which included genuinely mismatching pairs

	Locus	Number of genotypes compared	Number of alleles mismatching	Detection probability	Estimated error rate per reaction
Mother-offspring pairs mismatching at one or fewer loci ($n = 258$ pairs)	Aa4	257	0	0.3448	0
	Hg1.3	258	0	0.5859	0
	Hg6.3	252	0	0.5611	0
	Hg8.10	258	0	0.1019	0
	Lw10	257	0	0.6782	0
	M11a	257	1	0.7254	0.0027
	Pv9	257	2	0.4253	0.0091
	PvcA	257	0	0.3981	0
	PvcE	258	0	0.5896	0
	Overall		3		0.0013
All mother-offspring pairs ($n = 298$ pairs)	Aa4	297	9	0.3448	0.0439
	Hg1.3	296	17	0.5859	0.0490
	Hg6.3	290	23	0.5611	0.0707
	Hg8.10	298	5	0.1019	0.0823
	Lw10	297	26	0.6782	0.0645
	M11a	296	28	0.7254	0.0652
	Pv9	297	18	0.4253	0.0712
	PvcA	297	14	0.3981	0.0592
	PvcE	297	23	0.5896	0.0657
	Overall		163		0.0635

Table 5 Summary of the frequencies of different types of genotyping error, identified using a variety of error-checking methods (see text for a description of error types). For each method, the numbers of errors detected are shown together with the percentage of the total number of errors in parentheses. Genotyping errors attributed to unknown causes or allele dropout could only be ascertained for comparisons involving individuals whose relationships were known from field records. However in one instance, a genotyping error detected by repeat-genotyping that could be attributed to allele dropout also yielded a father–offspring mismatch. Consequently, this error is shown twice in the table, although it is only counted once in the 'overall' section

Error type	Method						
	Repeat-genotyping	Deliberately resampled females	Deliberately resampled males	Unintentionally resampled males	Mother–offspring mismatches	Offspring–putative father mismatches	Overall
Adjacent allele heterozygote scoring error	2 (15.4)	2 (100)	2 (50)	7 (30.4)	5 (33.3)	4 (21.1)	22 (29.3)
Other scoring error	3 (23.1)	0 (0)	2 (50)	16 (69.6)	5 (33.3)	12 (63.2)	38 (50.7)
Data input	2 (15.4)	0 (0)	0 (0)	0 (0)	1 (6.7)	2 (10.5)	5 (6.7)
Allele dropout	4 (30.8)	0 (0)	0 (0)	—	4 (26.7)	1 (5.3)	8 (10.7)
Unknown	2 (15.4)	0 (0)	0 (0)	—	—	—	2 (2.7)
Overall	13 (100)	2 (100)	4 (100)	23 (100)	15 (100)	19 (100)	75 (100)

involved only a single mistyped allele. Eighty percent of mistyped reactions ($n = 60$) were due to scoring errors and of these, over a third ($n = 22$, 36.7%) were due to confusion between homozygote and adjacent allele heterozygote genotypes. Thirteen heterozygotes were mis-scored as homozygotes (six cases being attributed to the genotype being overexposed) and nine homozygotes were scored as heterozygotes (two cases being attributed to the genotype being underexposed). Causes for the remaining scoring errors included nonspecific amplification products being scored as alleles ($n = 4$, 6.7%); rare alleles being sized incorrectly ($n = 2$, 3.3%), preferential amplification of smaller alleles leading to fainter larger alleles being missed ($n = 5$, 8.3%), alleles appearing blurry on autoradiographs ($n = 1$, 1.7%), autoradiograph underexposure ($n = 8$, 13.3%) and autoradiograph overexposure ($n = 1$, 1.7%).

Locus-by-locus variation in genotyping errors

Because the likelihood of a multilocus genotype containing an error increases with the number of loci typed, Waits & Leberg (2003) proposed genotyping individuals for fewer but more informative loci as an error-reduction strategy. While this approach seems intuitive, it could be problematic if error rates are higher for more polymorphic loci. Therefore, to investigate this possibility, we examined the relationship between the total number of errors identified at each locus and a number of polymorphism characteristics. First, we expressed polymorphism in terms of expected heterozygosity (H_E) and the number of observed alleles. Then, because stutter bands can influence the reliability of allele size estimation (Jones et al. 1997), we also counted the mode number of stutter bands observed at each locus.

Finally, because experience from our laboratory suggests that larger PCR products can be more difficult to score, mainly because of the increased difficulty in achieving adequate separation of alleles on a polyacrylamide sequencing gel, we also correlated the number of errors found at each locus with mode product size (e.g. size of the commonest allele). Three of the above characteristics — H_E, the number of alleles and the mode number of stutter bands, were significantly correlated with one another (Table 6), suggesting that loci with greater numbers of alleles and heterozygosity tended to possess more stutter bands. These three measures correlated positively with the number of errors (Fig. 1), revealing a tendency for greater numbers of errors to be found at more polymorphic loci, although these correlations were not statistically significant (H_E $r_s = 0.412$, $n = 9$, $P = 0.271$; number of alleles $r_s = 0.460$, $n = 9$, $P = 0.213$; mode number of stutter bands $r_s = 0.581$, $n = 9$, $P = 0.101$). Mode product size was more strongly correlated with the number of errors ($r_s = 0.739$, $n = 9$, $P = 0.023$), suggesting that loci yielding larger products were more error-prone.

Influence of error rate on paternity assignment

Despite the various estimates of the genotyping error rate for our data set being very low (ranging between 0.0013 and 0.0074 per reaction), in our paternity analysis (Hoffman et al. 2003) we assigned 19 additional paternities (4.9% of the total number of paternities) following rectification of genotyping errors. This finding suggests that even a small genotyping error rate could lead to a significant rate of false paternal exclusion. To explore the relationship between genotyping error rate and paternal exclusion further, we conducted simulations using our data set. All 75 of the

Table 6 Table showing correlations among four polymorphism characteristics measured for nine microsatellite loci. Spearman rank correlation coefficients are given in the top half of the matrix, and *P*-values, following Bonferroni correction for multiple tests, are given in the bottom half

	H_E	Number of alleles	Mode number of stutter bands	Mode product size
H_E	*	0.929	0.798	−0.167
Number of alleles	0.002	*	0.864	−0.092
Mode number of stutter bands	0.058	0.016	*	0.177
Mode product size	0.999	0.999	0.998	*
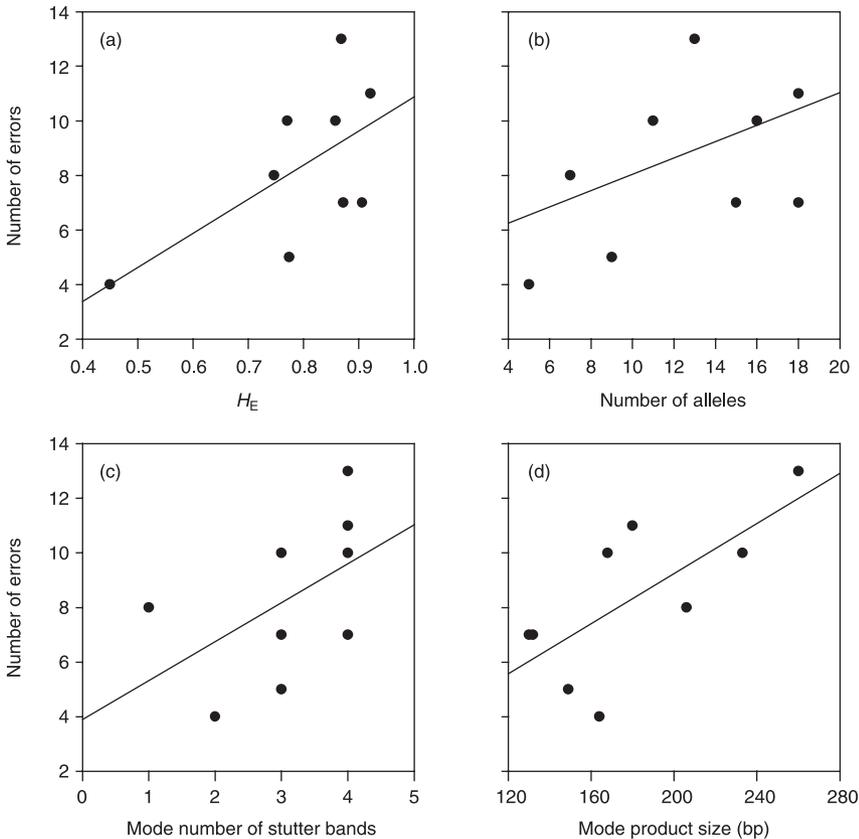
Fig. 1 Plots showing the relationship between the total number of errors identified at each locus and (a) expected hetero-zygosity; (b) the number of alleles observed; (c) the mode number of stutter bands; and (d) mode product size. Linear regression lines are presented for ease of viewing.

errors that we identified in this study were rectified, effectively reducing the error rate to a negligible level. We then re-introduced random errors at several different rates, each time rerunning the paternity analysis to determine how many of the paternities that were originally assigned remained afterwards. Very few of the errors that we intro-duced yielded mismatches between mothers and offspring (only eight out of 660 mother–pup pairs mismatched at the maximum error rate of 0.02 per allele). However, the genotyping error rate did have a strong influence on the overall rate of paternity assignment (Fig. 2). Of the 388 paternities that were originally assigned before any errors were introduced, the percentage of pups assigned paternity

fell to 66.2% with a 0.02 error rate per allele. Interestingly, genotyping errors also resulted in a lower apparent skew in male success (Fig. 2). This was because errors seldom create but often remove paternity assignments. Consequently, while males who started with only a single pup assigned are either eliminated or remain unchanged, the most successful males have their tally progressively reduced.

Discussion

Although it has long been acknowledged that micro-satellite genotyping can be prone to error, few studies have examined either the prevalence or consequences of
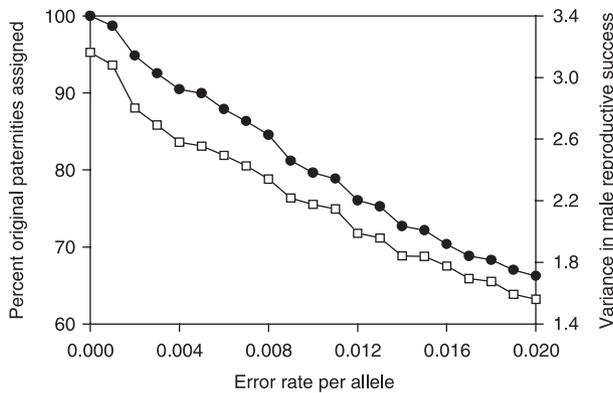
Fig. 2 Simulation showing the influence of genotyping error rate on paternity assignment for our data set (see text for details). The graph shows error rate per allele vs. the percentage of original paternities that were assigned (closed circles) and the variance in the number of paternities assigned to males (open squares).

such errors. Here, using a large microsatellite data set, we explored the utility of different methods for detecting errors, identified the most common sources of error and explored the impact of error rates on paternal exclusion. Our findings lend further support to the notion that errors are likely to be prevalent even in high quality data sets, and also enable us to propose a number of strategies for error reduction.

First, we used a number of different methods to detect genotyping errors, including repeat-genotyping a randomly selected subset of samples, comparing the genotypes of deliberately and accidentally resampled individuals and checking for mismatches between pups and their putative parents. We found generally good concordance among the resulting error rate estimates. The highest and lowest rates, 0.0074 and 0.0013, differ by a factor of nearly six, but the numbers involved are small. Testing the four comparable error rates (i.e. all except mother–offspring mismatches), we find no evidence of heterogeneity ($\chi^2 = 5.38$, d.f. = 3, ns).

Arguably, the easiest and most widely accessible form of error-checking is by inspection of mother–offspring mismatches. This approach can detect loci with serious genotyping problems arising from null alleles or artefacts, and may also help to identify cases where entire autoradiographs have been misread. However, by their nature, mother–offspring mismatches only detect a subset of errors. This expectation is reflected both in the observation that this class yields the lowest estimated error rate and in our simulation studies where an error rate of 2% per allele generates fewer than 10 discernable mother–offspring mismatches among 660 pairs (1.2% mismatching rate). Furthermore, our loci are reasonably polymorphic and, as

marker loci become individually less polymorphic, genotyping errors become less easy to eliminate through detection of Mendelian inconsistencies (Göring & Terwilliger 2000). The mother–offspring mismatch approach also depends critically on how many offspring each mother has, with more offspring increasing the chance that an error will be detected. Similarly, confidence must be reasonably high that mother–offspring pairs identified in the field are genuine. In some systems, fostering, egg-dumping and other similar behaviours can result in nonfilial mother–offspring pairs. Our analysis shows that, as expected, inclusion of genuinely mismatching pairs leads to a vastly inflated estimated error rate. Thus, we believe that mother–offspring mismatches are useful, but are not the best method available for error rate estimation, and may instead be best used in conjunction with other approaches. Indeed, because we found little overlap between the different methods in terms of the specific errors that were detected (99% of errors were detected only once), we recommend that data rectification should employ as many different methods as are available.

A large literature exists on those genotyping errors that arise when the template DNA is of low quantity and/or quality, typically either allele dropout or 'misprinting', but beyond this, few studies have sought to determine the most likely sources of error under more normal circumstances. Here using the total 75 errors that we identified, we found that scoring errors were by far the most prevalent source of genotyping error. Of these, adjacent allele heterozygotes were the most common form, accounting for 36.7% of scoring errors (22.9% of all errors). Experience from our lab suggests that adjacent allele heterozygotes are problematic to score across a variety of projects involving different species, and this finding is also consistent with concerns raised by Litt *et al.* (1993) and Johansson *et al.* (2003). Interestingly, we observed a tendency for heterozygotes to be scored as homozygotes when autoradiographs were overexposed and for homozygotes to be scored as heterozygotes when autoradiographs were underexposed. Together these findings present avenues for improved error reduction. First, adjacent allele heterozygotes can be rechecked and regenotyped whenever in doubt, particularly when the exposure is imperfect. Second, the frequencies of particular adjacent allele heterozygotes and of adjacent allele heterozygotes overall can be compared with the frequencies expected assuming random assortment of alleles, and this could be used to provide a diagnostic test. We have written an Excel macro to do this (please contact WA).

Even with low genotyping error rates, the problem becomes worse as the number of loci used increases. Consequently, Waits & Leberg (2003) have proposed using fewer, but more informative loci, as an error reduction strategy. However, this approach could be problematic

if error rates are higher for more polymorphic loci. Here, using 75 errors detected across a panel of nine loci, we found that loci with greater numbers of alleles, increased heterozygosity and greater numbers of stutter bands showed a weak tendency to exhibit more genotyping errors. In addition, the number of errors found correlated significantly with mode product size, a finding that is consistent with another observation by Sefc *et al.* (2003) that PCR failure and allele dropout rates increase with product size. Taken together, these findings suggest that more polymorphic loci should be treated with greater caution, and that it may be worthwhile either avoiding loci yielding larger product sizes or redesigning primers to amplify smaller fragments.

Simulations using our data set suggest that an error rate as low as 1% results in over 20% of paternities being no longer assigned (Fig. 2). This finding is broadly consistent with the results of a previous study in which error rates were investigated within a likelihood framework (Marshall *et al.* 1998). Interestingly, we also found that the variance in the number of paternities assigned to different males decreased with the error rate, suggesting that the relationship between genotyping error rate and false exclusion will vary with the degree of reproductive skew. Furthermore, the presence of even a small number of genotyping errors had a more subtle influence on our paternity analysis. Errors that arise in the genotypes of candidate males lead to overestimation of the number of unique candidate male genotypes, and the frequency of pups matching more than one male increases.

Finally, it is worth asking at what point does it become more efficient to accept a certain level of error and instead spend resources on typing additional loci? The answer is not simple and depends both on how sensitive downstream analyses are to the presence of errors and how much repeat-genotyping has to be carried out for other reasons. For example, genotyping errors may have little impact on the measurement of relatedness between individuals (except where duplicate samples yield different genotypes) but allele dropout and null alleles could present major problems for the estimation of heterozygosity. With paternity testing, the use of more loci increases the total number of errors, but this may be more than compensated for if the increase in resolution allows fathers with single mismatches to be accepted as genuine. Perhaps most problematic are subtle patterns like low rates of extra pair paternities in birds, where some or even most of events could in principle be ascribed to genotyping error. On the plus side, there is relatively little experimental effort involved in retyping all critical samples. Given this diversity of effect, we urge the development of analytical approaches to detect patterns suggestive of errors, and also programs to simulate the likely outcome of different error rates on the analyses being applied.

Conclusion

We explored a variety of different approaches for detecting genotyping errors and show that these tended to yield similar estimates of the underlying error rate. Our analyses also highlight systematic sources of error, such as confusion between homozygotes and adjacent allele heterozygotes and variation among loci, which suggest areas where any given effort in error reduction is best directed. Because we find very little overlap in the errors detected by different approaches, in general it is advisable to use all possible methods of detection in order to drive the overall rate as low as possible.

References

Allen PJ, Amos W, Pomeroy PP, Twiss SD (1995) Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Molecular Ecology*, **4**, 653–662.

Amos W, Hoelzel AR (1991) Long-term preservation of whale skin for DNA analysis. *Report of the International Whaling Commission Special Issue*, **13**, 99–103.

Arnould JPY, Duck CD (1997) The cost and benefits of territorial tenure, and factors affecting mating success in male Antarctic fur seals. *Journal of Zoology (London)*, **241**, 649–664.

Bradley BJ, Vigilant L (2002) False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces. *Molecular Ecology Notes*, **2**, 602–605.

Brzustowicz LM, Merette C, Xie X *et al.* (1993) Molecular and statistical approaches to the detection and correction of errors in genotype databases. *American Journal of Human Genetics*, **53**, 1137–1145.

Callen DF, Thompson AD, Shen Y *et al.* (1993) Incidence and origin of 'null' alleles in the (AC)$_n$ microsatellite markers. *American Journal of Human Genetics*, **52**, 922–927.

Coltman DW, Bowen WD, Wright JM (1996) PCR primers for harbour seal (*Phoca vitulina concolour*) microsatellites amplify polymorphic loci in other pinniped species. *Molecular Ecology*, **5**, 161–163.

Constable JL, Ashley MV, Goodall J, Pusey AE (2001) Noninvasive paternity assignment in Gombe chimpanzees. *Molecular Ecology*, **10**, 1279–1300.

Creel S, Spong G, Sands JL *et al.* (2003) Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology*, **12**, 2003–2009.

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, 1–6.

Davidson A, Chiba S (2003) Laboratory temperature variation is a previously unrecognised source of genotyping error during capillary electrophoresis. *Molecular Ecology Notes*, **3**, 321–323.

Davis CS, Gelatt TS, Siniff D, Strobeck C (2002) Dinucleotide microsatellite markers from the Antarctic seals and their use in other pinnipeds. *Molecular Ecology Notes*, **2**, 203–208.

Doidge DW, Croxall JP, Baker JR (1984) Density-dependent pup mortality in the Antarctic fur seal *Arctocephalus gazella* at south Georgia. *Journal of Zoology (London)*, **202**, 449–460.

Ernest HB, Penedo MCT, May BP, Syvanen M, Boyce WM (2000) Molecular tracking of mountain lions in the Yosemite Valley region in California: genetic analysis using microsatellites and faecal DNA. *Molecular Ecology*, **9**, 433–441.

Evett IW, Weir BS (1998) *Interpreting DNA Evidence*. Sinaur Associates, Inc, Sunderland, Massachusetts.

Ewen KR, Bahlo M, Treloar SA *et al.* (2000) Identification and analysis of error types in high-throughput genotyping. *American Journal of Human Genetics*, **67**, 727–736.

Fernando P, Evans BJ, Morales JC, Melnick DJ (2001) Electrophoresis artefacts — a previously unrecognised cause of error in microsatellite analysis. *Molecular Ecology Notes*, **1**, 325–328.

Fernando P, Vidya TNC, Rajapakse C, Dangolla A, Melnick DJ (2003) Reliable noninvasive genotyping: fantasy or reality? *Journal of Heredity*, **94**, 115–123.

Frantz AC, Pope LC, Carpenter PJ *et al.* (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, **12**, 1649–1661.

Gagneux P, Boesch C, Woodruff DS (1997a) Microsatellite scoring errors associated with noninvasive genotyping based on nuclear DNA amplified from shed hair. *Molecular Ecology*, **6**, 861–868.

Gagneux P, Woodruff DS, Boesch C (1997b) Furtive mating by female chimpanzees. *Nature (London)*, **387**, 327–328.

Gagneux P, Woodruff DS, Boesch C (2001) Retraction. Furtive mating in female chimpanzees. *Nature (London)*, **414**, 508.

Gemmell NJ, Allen PJ, Goodman SJ, Reed JZ (1997) Interspecific microsatellite markers for the study of pinniped populations. *Molecular Ecology*, **6**, 661–666.

Gemmell NJ, Burg TM, Boyd IL, Amos W (2001) Low reproductive success in territorial male Antarctic fur seals (*Arctocephalus gazella*) suggests the existence of alternative mating strategies. *Molecular Ecology*, **10**, 451–460.

Gemmell NJ, Majluf P (1997) Projectile biopsy sampling of fur seals. *Marine Mammal Science*, **13**, 512–516.

Gerloff U, Schlotterer C, Rassman K *et al.* (1995) Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living bonobos (*Pan paniscus*). *Molecular Ecology*, **4**, 515–518.

Ghosh S, Karanjawala ZE, Hauser ER *et al.* (1997) Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently-labelled dinucleotide markers. FUSION (Finland-US Investigation of NIDDM Genetics) study group. *Genome Research*, **7**, 165–178.

Ginot F, Bordelais I, Nguyen S, Gyapay G (1996) Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research*, **24**, 540–541.

Gomes I, Collins A, Lonjou C *et al.* (1999) Hardy–Weinberg quality control. *Annals of Human Genetics*, **63**, 535–538.

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.

Göring HHH, Terwilliger JD (2000) Linkage analysis in the presence of errors II: marker-locus genotyping errors modeled with hypercomplex recombination fractions. *American Journal of Human Genetics*, **66**, 1107–1118.

Harker N (2001) Collection, reporting and storage of microsatellite genotype data. In: *Plant Genotyping: the DNA Fingerprinting of Plants* (ed. Henry RJ), pp. 251–264. CAB International, Wallingford, UK & New York, USA.

Hoelzel AR, LeBoeuf BJ, Reiter J, Campagna C (1999) Alpha-male paternity in elephant seals. *Behavioral Ecology and Sociobiology*, **46**, 298–306.

Hoffman JI, Boyd IL, Amos W (2003) Male reproductive strategy and the importance of maternal status in the Antarctic fur seal *Arctocephalus gazella*. *Evolution*, **57**, 1917–1930.

Jarne P, Lagoda PJL (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Johansson A, Karlsson P, Gyllensten U (2003) A novel method for automatic genotyping of microsatellite markers based on parametric pattern recognition. *Human Genetics*, **113**, 316–324.

Jones CJ, Edwards KJ, Castaglione S *et al.* (1997) Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. *Molecular Breeding*, **3**, 381–390.

Kohn MH, York EC, Kamradt DA *et al.* (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London Series B-Biology Sciences*, **266**, 657–663.

Litt M, Hauge X, Sharma V (1993) Shadow bands seen when typing polymorphic dinucleotide repeats — some causes and cures. *Biotechniques*, **15**, 280 et seq.

Luikart G, England PR (1999) Statistical analysis of microsatellite DNA data. *Trends in Ecology and Evolution*, **14**, 253–256.

Majluf P, Goebel ME (1992) The capture and handling of female South American fur seals and their pups. *Marine Mammal Science*, **8**, 187–190.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639–655.

Miller CR, Joyce P, Waits LP (2002) Assessing allelic dropout and genotype reliability using maximum likelihood. *Genetics*, **160**, 357–366.

Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative polymerase chain reaction analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, **10**, 1835–1844.

Navidi W, Arnheim N, Waterman MS (1992) A multiple-tubes 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.

Palsbøll PJ, Allen J, Berube M *et al.* (1997) Genetic tagging of humpback whales. *Nature (London)*, **388**, 767–769.

Parsons KM (2001) Reliable microsatellite genotyping of dolphin DNA from faeces. *Molecular Ecology*, **1**, 341–344.

Pemberton JM, Slate J, Bancroft DR, Barrett JA (1995) Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, **4**, 249–252.

Piggott MP, Taylor AC (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research*, **30**, 1–13.

Queller DC, Strassmann JE, Hughes CR (1993) Microsatellites and kinship. *Trends in Ecology and Evolution*, **8**, 285–288.

Sefc KM, Payne RB, Sorenson MD (2003) Microsatellite amplification from museum feather samples: effects of fragment size and template concentration on genotyping errors. *Auk*, **120**, 982–989.

Segelbacher G (2002) Noninvasive genetic analysis in birds: testing reliability of feather samples. *Molecular Ecology Notes*, **2**, 367–369.

Sloane MA, Sunnucks P, Alpers D, Beheregaray B, Taylor AC (2000) Highly reliable genetic identification of individual northern hairy-nosed wombats from single remotely collected hairs: a feasible censusing method. *Molecular Ecology*, **9**, 123–1240.

Sobel E, Papp JC, Lange K (2002) Detection and integration of genotyping errors in statistical genetics. *American Journal of Human Genetics*, **70**, 496–508.

Taberlet P, Griffin S, Goosens B *et al.* (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–3194.

Taberlet P, Luikart G (1999) Non-invasive genetic sampling and individual identification. *Biology Journal of the Linnean Society*, **68**, 41–55.

Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.

Valiere N, Berthier P, Mouchiroud D, Pontier D (2002) GEMINI: software for testing the effects of genotyping errors and multi-tubes approach for individual identification. *Molecular Ecology Notes*, **2**, 83–86.

Vigilant L, Hofreiter M, Siedel H, Boesch C (2001) Paternity and relatedness in wild chimpanzee communities. *Proceedings of the National Academy of Sciences USA*, **98**, 12890–12895.

Waits JL, Leberg PL (2003) Biases associated with population estimation using molecular tagging. *Animal Conservation*, **3**, 191–199.

Walsh PS, Ehrlich HA, Higuchi R (1992) Preferential amplification of alleles: mechanisms and solutions. *PCR Methods and Applications*, **1**, 241–250.

Walsh PS, Metzger DA, Higuchi R (1991) Chelex100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques*, **10**, 506–513.

Wasser SK, Houston CS, Koehler GM, Cadd GG, Fain SR (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, **6**, 1091–1097.

Weeks DE, Conley YP, Ferrell RE, Mah TS, Gorin MB (2002) A tale of two genotypes: consistency between two high-throughput genotyping centres. *Genome Research*, **12**, 430–435.

Worthington Wilmer J, Allen PJ, Pomeroy PP, Twiss SD, Amos W (1999) Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, **8**, 1417–1429.

Joe Hoffman is a postdoctoral research associate working in Bill Amos' laboratory on the genetic analysis of male reproductive success in Antarctic fur seals. His interests include animal mating systems and genetic factors that influence fitness in natural populations. Bill Amos' work is currently focused most on the relationship between heterozygosity and fitness and he is increasingly interested in human population genetics.