Molecular Ecology Notes (2005) **5**, 680–683 doi: 10.1111/j.1471-8286.2005.01002.x

TECHNICAL NOTE

Locus effects and sources of error in noninvasive genotyping

JASON C. BUCHAN,* ELIZABETH A. ARCHIE,* RUSSELL C. VAN HORN,* CYNTHIA J. MOSS† and SUSAN C. ALBERTS*

**Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA, †Amboseli Elephant Research Project, P.O. Box 15135, Nairobi, Kenya*

Abstract

In spite of more than a decade of research on noninvasive genetic sampling, the low quality and quantity of DNA in noninvasive studies continue to plague researchers. Effects of locus size on error have been documented but are still poorly understood. Further, sources of error other than allelic dropout have been described but are often not well quantified. Here we analyse the effects of locus size on allelic dropout, amplification success and error rates in noninvasive genotyping studies of three species, and quantify error other than allelic dropout.

Keywords: allelic dropout, faecal DNA, noninvasive genotyping, polymerase slippage, contamination

Received 17 December 2004; revision received 02 February 2005; accepted 18 March 2005

DNA obtained noninvasively is increasingly important in studying natural animal populations, but the DNA is typically of low quality and quantity. Poor amplification success, allelic dropout, false alleles and contamination make it difficult to obtain reliable genotypes using such DNA. These errors are affected by a variety of factors, including the range of allele sizes for the loci amplified (e.g. Sefc *et al*. 2003). Various methods address these problems, including the 'multitubes approach' (Navidi *et al*. 1992; Taberlet *et al*. 1996), a maximum likelihood method that estimates genotype reliability and then directs replication towards error-prone loci (Miller *et al*. 2002), and a quantitative polymerase chain reaction (qPCR) method that modifies the number of replicates in response to information about template concentration and dropout rates (Morin *et al*. 2001).

This note has two goals: (i) to measure the effect of locus size on allelic dropout, amplification success and other sources of error, and (ii) to quantify sources of error in faecal genotyping other than allelic dropout. The maximum likelihood model of Miller *et al*. (2002) assumes that the final data set contains no errors and that allelic dropout rates are constant across loci. These assumptions are largely untested in real data sets of noninvasive genotypes, as thorough accounts of error in noninvasive genotyping studies are rare. To address both goals, we draw on large-scale studies conducted in our laboratory on savannah baboons (Buchan *et al*. 2003) and African elephants (Archie *et al*. 2003). For the first goal, we compared our results on baboons and elephants with those reported for chimpanzees by Morin *et al*. (2001).

For both baboons and elephants, fresh faeces were collected as soon as possible (typically a matter of minutes) after animals of known identity were observed to defecate. Wherever possible, multiple independent faecal samples (i.e. from separate defecations) were collected for each individual. For baboons, approximately 2 g of faeces was collected from the leading end of the faecal bolus and placed in a vial containing 10 mL of 95% ethanol. For elephants, approximately 10 g of faeces was collected from the surface of the faecal bolus and placed into a 15-mL vial that was then filled with 95% ethanol. Samples were stored for up to 6 months (baboon samples) or 1 year (elephant samples) at ambient temperature in the field before being stored at –80 °C in the laboratory. Most samples were a few months to several years old at the time of DNA extraction.

The QIAamp DNA Stool Mini Kit (QIAGEN) was used to extract DNA from both species, modified slightly as described in Buchan *et al*. (2003) and Archie *et al*. (2003). Baboons were genotyped at 12 tetranucleotide and two dinucleotide microsatellite loci (Table 1). Elephants were typed at 10 tetranucleotide microsatellite loci and one dinucleotide locus (Table 1; also Nyakaana & Arctander 1998;

Correspondence: S. C. Alberts, Fax: +919 6607293

Table 1 Locus names, size ranges (bp), proportion of positive PCRs and proportion of allelic dropout for all loci used in the baboon and elephant studies. All loci were tetranucleotides except for baboon loci D7s503 and D13s159B and elephant locus LaFMs02, which were dinucleotides. Note that proportion of positive PCRs is expressed relative to the number of PCRs, whereas allelic dropout is expressed relative to the number of reactions for heterozygous genotypes

Locus	Product size range	% successful amplification	% allelic dropout
Baboons			
D1s1656	130–167	93% (489/523)	9% (31/363)
D2s1326	237–273	73% (640/875)	40% (184/455)
D3s1768	178–218	96% (517/539)	17% (64/382)
D4s243	155–179	94% (522/556)	13% (44/343)
D5s1457	110–138	94% (510/543)	9% (32/369)
D6s501	171–227	93% (571/617)	27% (107/391)
D7s503	133–169	93% (520/562)	14% (53/369)
D8s1106	128–161	99% (594/602)	12% (39/334)
D10s611	150–195	88% (612/698)	23% (100/439)
D11s 2002	252–280	90% (562/626)	30% (138/455)
D13s159B	165–181	99% (284/288)	6% (11/180)
D14s306	146–190	87% (527/607)	11% (38/333)
D18s851	225–249	73% (593/812)	34% (144/429)
AGAT006	131–181	97% (469/483)	8% (30/371)
Total		**89% (7410/8331)**	**19% (1015/5213)**
Elephants			
LaFMs02	138–150	82% (1058/1297)	19% (99/526)
LaT05	255–321	66% (1434/2178)	17% (163/954)
LaT07	336–406	53% (1215/2287)	26% (275/1074)
LaT08	166–234	65% (1151/1772)	20% (170/850)
LaT13	226–262	69% (1203/1734)	18% (120/652)
LaT16	295–327	60% (1000/1656)	21% (151/730)
LaT17	283–359	68% (1146/1697)	20% (159/793)
LaT18	286–326	67% (1079/1617)	22% (159/726)
LaT24	204–244	71% (1070/1508)	15% (114/751)
LaT25	294–322	52% (1085/2079)	23% (156/677)
LaT26	352–400	40% (1006/2506)	25% (205/797)
Total		**61% (12447/20313)**	**21% (1771/8530)**

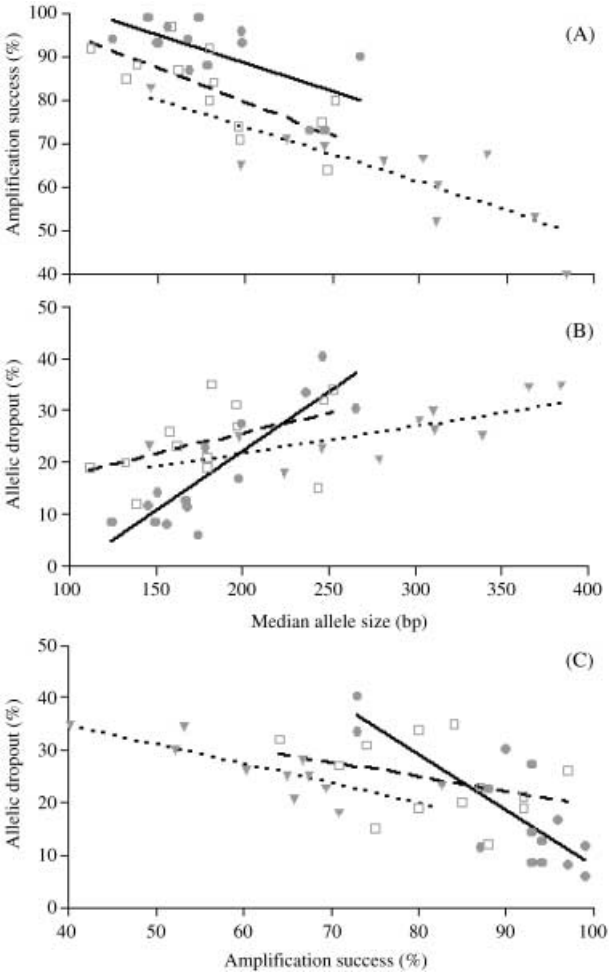

Fig. 1 Percent amplification success (A) and percent allelic dropout (B) as a function of median allele size, and percent allelic dropout (C) as a function of amplification success for loci of baboons (solid lines, filled circles), chimpanzees (dashed lines, open squares) and elephants (dotted lines, filled triangles). See text for regression equations and parameters.

Archie *et al*. 2003). For reaction details, see Buchan *et al*. (2003) and Archie *et al*. (2003).

Median allele size had a significant effect on amplification success in the baboon and elephant samples analysed in our laboratory, and in the chimpanzee samples analysed by Morin *et al*. (2001). Smaller loci had higher amplification success (Fig. 1(A); see also Sefc *et al*. 2003 for a similar result with DNA from feathers). This effect was almost identical in all three species (baboons: $y = 0.1278x + 113.86$, $r^2 = 0.4236$, $P = 0.013$; elephants: $y = 0.1275x + 99.08$, $r^2 = 0.6713$, $P = 0.002$; chimpanzees: $y = 0.1578x + 110.1$, $r^2 = 0.5178$, $P = 0.006$). Thus, researchers hoping to genotype from non-invasive samples can expect a 12–15% decline in amplification success for each 100 bp increase in allele size.

Median allele size and amplification success also influenced rates of allelic dropout (see also Sefc *et al*. 2003). For both baboons and elephants, larger loci had significantly higher rates of allelic dropout (Fig. 1(B); baboons: $y = 0.228x - 23.595$, $r^2 = 0.7608$, $P < 0.0001$; elephants: $y = 0.0522x + 11.392$, $r^2 = 0.5133$, $P = 0.0131$). For chimpanzees (with fewer loci across a narrower range of sizes), there was a nonsignificant trend for dropout to occur at higher rates for larger loci (Fig. 1(B); $y = 0.0078x + 9.7409$, $r^2 = 0.2275$, $P = 0.093$). Dropout rate declined significantly with amplification success among baboon and elephant loci, but there was not a significant effect of amplification success upon dropout rate in chimpanzee loci (Fig. 1(C); baboons: $y = 1.0817x + 115.97$, $r^2 = 0.6304$, $P = 0.004$; elephants: $y = 0.3734x + 49.69$, $r^2 = 0.6362$, $P = 0.033$; chimpanzees: $y = -0.2795x + 47.14$, $r^2 = 0.1284$, $P = 0.229$).

Other than allelic dropout, we experienced four types of error: (i) errors involving probable contamination with human DNA (in baboon samples only), (ii) errors involving probable polymerase slippage during amplification (i.e. alleles that differed by one repeat from the true allele), (iii) errors involving probable contamination with DNA from the study population and (iv) errors involving a 'false allele', a spurious allele that appeared once and was never or rarely replicated in that or any other sample.

1 Probable contamination with human DNA was identified by reference to the genotypes of laboratory and field personnel. Such contamination occurred in 1.3% of the PCRs using baboon DNA and in 1.2% of negative controls, accounting for approximately 42% of the total error other than allelic dropout. Not surprisingly, human contamination did not affect genotyping of the elephant samples.
2 Probable polymerase slippage occurred in 0.6% of baboon and 0.8% of elephant PCRs accounting for 18% and 47% of the total error in each study, respectively.
3 Rarely, PCR revealed probable contamination from the study population: 0.8% of baboon and 0.5% of elephant PCRs produced alleles that did not match the animal being typed (i.e. appeared only once in multiple replicates for that individual) but were known to exist in the study population. These accounted for 26% and 32% of the total error in each study, respectively.
4 Spurious alleles were also rare: 0.5% of baboon and 0.3% of elephant PCRs produced 'alleles' that were not replicable and were not found within the study population, accounting for 14% and 20% of total errors, respectively.

Overall, except for human contamination, the rates of various types of errors were similar between baboon and elephant samples. Rates of these errors varied across different loci in baboon (0.21–5.55%) and elephant samples (0.63–4.23%), but median allele size did not have an affect on the rate of these types of error in either baboon samples ($y = 0.0069x + 0.4697$, $N = 14$, $r^2 = 0.039$, $P = 0.50$) or elephant samples ($y = 0.0026x + 0.9375$, $N = 11$, $r^2 = 0.037$, $P = 0.57$).

The effects of locus size varied among species and may reflect the size of template DNA, which is a function of the level of degradation in samples. It does seem that, when possible, smaller loci would be better choices than larger loci for faecal genotyping studies. However, size could not explain all the variation in amplification success and dropout among our samples; some larger loci (e.g. D3s1768; see Table 1) performed as well as considerably smaller loci, and some loci had high rates of allelic dropout even as DNA concentrations increased. This suggests that factors difficult to quantify, such as the efficiency with which particular primers anneal, or secondary structures in the template DNA, may contribute to amplification success.

This interlocus variability is not accounted for in current models of noninvasive genotyping, which assume that allelic dropout is solely a function of stochastic sampling error (Navidi *et al*. 1992; Taberlet *et al*. 1996). In particular, interlocus variability in dropout has implications for the assessment of genotypic validity via the maximum likelihood method of Miller *et al*. (2002), which assumes that rates of allelic dropout are constant across loci. Although this method is somewhat robust to interlocus variation in allelic dropout, future researchers should ensure that they are not sacrificing accuracy for efficiency, especially when it is important to assign particular genotypes to specific individuals (e.g. in studies of paternity or relatedness). In such studies, it may be particularly tempting to utilize microsatellites containing more repeat units, under the expectation of higher allelic diversity (reviewed in Petit *et al*. 2005), but our results indicate that longer markers may entail significant disadvantages. Thus, because researchers cannot predict a priori the locus-specific rate of allelic dropout, we recommend that future noninvasive geneticists use pilot studies to explore locus- and species-specific properties of their genotypes, and then make informed decisions about which genotyping method will optimize accuracy and efficiency in a way that is appropriate to the research questions.

Acknowledgements

We thank the Office of the President, Republic of Kenya, the Kenya Wildlife Service, its Amboseli staff and wardens, the Institute of Primate Research, the National Museums of Kenya, and the members of the Amboseli-Longido pastoralist communities. We thank the Amboseli fieldworkers of the Amboseli Baboon Research Project and the Amboseli Elephant Research Project. Thanks to Kerri Smith, William Beggs and Jeanne Beck of the Coriell Institute and to Jeanne Altmann and Lisa Bukovnik for invaluable assistance and support. The research was funded by NSF IBN 0091612 and NSF IBN 9985910 to S.C.A., Duke University, the African Elephant Conservation Trust and the Amboseli Elephant Research Project.

References

Archie EA, Moss CJ, Alberts SC (2003) Characterization of tetranucleotide microsatellite loci in the African savannah elephant (*Loxodonta africana africana*). *Molecular Ecological Notes*, **3**, 244–246.

Buchan JC, Alberts SC, Silk JB, Altmann J (2003) True paternal care in a multi-male primate society. *Nature*, **425**, 179–181.

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.

Nyakaana S, Arctander P (1998) Isolation and characterization of microsatellite loci in the African elephant, *Loxodonta africana*. *Molecular Ecology*, **7**, 1436–1437.

Petit RJ, Deguilloux M-F, Chat J, Grivet D, Garnier-Géré P, Vendramin GG (2005) Standardizing for microsatellite length in comparisons of genetic diversity. *Molecular Ecology*, **14**, 885–890.

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

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