

## TECHNICAL NOTE

# False alleles derived from microbial DNA pose a potential source of error in microsatellite genotyping of DNA from faeces

B. J. BRADLEY\*† and L. VIGILANT\*

\*Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, Leipzig D-04103, Germany, †Anthropological Sciences, State University of New York, Stony Brook, NY 11794–4364, USA

## Abstract

**Microsatellite genotyping of wild animals using DNA extracted from noninvasive samples such as faeces is a powerful means to identify individuals within a population and examine aspects of genetic social structure, such as relatedness and paternity. However, the use of the low quantities of poor quality DNA typically obtained from noninvasive samples can result in genotyping errors. Here we report the first instance of artefactual 'alleles' resulting from specific co-amplification of microorganismal DNA present in the total DNA derived from faeces.**

*Keywords:* artefacts, cross-species amplification, error, genotyping, gorilla

*Received 21 May 2002; revision received 9 July 2002; accepted 22 July 2002*

Although rare in occurrence, amplification artefacts, or 'false alleles' cannot be disregarded as a potential source of error in microsatellite genotyping (Taberlet *et al.* 1999). Many genotyping studies rely upon cross-species amplification, that is, the application of primers derived in one species for characterization of individuals in another, usually closely related species (Coote & Bruford 1996; Bradley *et al.* 2000). As these primers might not be wholly complementary to the DNA of the new species, this might be expected to lead to an increase in the occurrence of amplification artefacts by reducing amplification specificity. In addition, the now common use of faeces as a noninvasive source of DNA for such studies means that the total DNA extracted from the sample derives not only from the source individual, but can also contain DNA from dietary components, parasites, and commensal bacteria. Hence, a variety of template DNAs in addition to that from the target organism are available for amplification. These factors notwithstanding, the reported incidence of anomalous, irreproducible results in microsatellite genotyping studies are usually low, occurring in fewer than 1% of analysed reactions, and hence it has not been considered a significant source of error (Morin *et al.* 2001).

False or spurious alleles are typically not only of low frequency, but also sporadic in occurrence. In contrast, the results reported here consist of apparent alleles repeatedly amplified from several individuals. The locus D8s1106 is a human-derived microsatellite locus used in our laboratory for the genotyping of wild gorillas (*Gorilla beringei beringei*) from faecal DNA (Bradley *et al.* 2000). Fresh faecal samples approximately 3–5 g in weight were collected and stored dry in 50 mL tubes containing 20 g of silica gel beads (Sigma). A final volume of 200 µL of genomic DNA was extracted from 100 mg of dried faeces using the QIAamp DNA Stool Kit according to the manufacturer's instructions. Individual polymerase chain reaction (PCR) amplifications were performed in 20 µL reactions containing 2 µL DNA extract, 1 × PCR buffer (10 mM Tris-HCl, pH 8.3; 50 mM KCl), 3.5 mM MgCl<sub>2</sub>, 16 µg BSA, 250 µM each dNTP, 200 nM each primer and 1 U Amplitaq Gold DNA polymerase (Perkin-Elmer). The results of a gradient PCR showed that amplification was most successful at an annealing temperature of 50–52 °C, hence amplification conditions on a PTC-200 thermocycler (MJ Research) were: 95 °C for 3 min; 45 cycles of 30 s at 95 °C, 30 s at 50 °C, 30 s at 72 °C and a final extension of 30 min at 72 °C. The 5' end of the forward primer was labelled with the fluorescent dye NED. Amplification products were separated using capillary electrophoresis (ABI Prism 310) and allele size was

Correspondence: L. Vigilant. Fax: ++ 49-341-9952-119; E-mail: vigilant@eva.mpg.de

**Table 1** Results containing unusually long alleles at D8s1106

Individual	Alleles observed (no. of occurrences)		
INT*	138 (2)	154 (3)	185 (1)
JEN*	154 (4)	189 (2)	
KRD*	154 (5)	185 (2)	189 (7)
MAK*	154 (6)	185 (3)	
BEE	142 (16)	154 (1)	185 (1)
INS	154 (9)	185 (2)	
TBK	154 (5)	185 (5)	
TWH	154 (7)	185 (6)	
KUY	154 (4)	189 (2)	
NDA	154 (6)	189 (4)	
VUB	154 (2)	189 (2)	

\* = cloned and sequenced

determined relative to an internal size standard (HD400) using GENESCAN 2.0. All genotypes were considered final only after repeated PCRs resulted in the scoring of each allele of a heterozygote genotype at least twice and the observation of the single allele of a homozygous genotype at least the minimum number of times as required by the DNA content of the extract (Morin *et al.* 2001).

The reported minimum length in humans of the amplified segment of D8s1106 containing a tetranucleotide repeat series is 149 bp. Alleles of lengths between 138 and 162 bp were found in 64 mountain gorillas, while the genotypes of 11 additional gorillas also contained reproducible products measuring up to 189 bp in length (Table 1). The pattern of amplification of these putative alleles was similar to that of true alleles, as they were detected in some individuals and not in others, and were present as peaks of normal shape and with intensity equal to those of true alleles. Although these putative alleles fell outside the normal range of allele sizes observed at this locus, this pattern of discontinuous distribution of allele sizes could be the result of a mutation interrupting a simple sequence of repeats into two blocks of repeats (Kayser *et al.* 1995). However, the actual genotypes found at D8s1106 were anomalous as three alleles were detected in some individuals (Table 1). Furthermore, comparisons of genotypes at this locus for mother, offspring, and potential sires revealed that these putative alleles would likely cause paternity exclusions. Thus, in order to assess the validity of the results at D8s1106, we sequenced several of the alleles obtained. PCR products from each individual were cloned directly (without reamplification) using the TOPO TA cloning kit (Invitrogen). Colony PCR (Kilger *et al.* 1997) was performed on 7–8 clones from each amplification product with subsequent purification using the QIAquick kit (Qiagen). Cycle sequencing and purification of the sequencing products was done as described (Ebersberger *et al.* 2002), analysed on an ABI3700, and resulting sequences aligned by eye

using BIOEDIT (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>).

The results for each individual consisted of some sequences corresponding to the targeted microsatellite locus, along with other highly dissimilar sequences of lengths corresponding to the anomalous alleles (Fig. 1a). Not all of the alleles listed in Table 1 were represented in the sequences obtained, presumably as a result of chance amplification from small amounts of template. The sequences corresponding to alleles 154 and 162 were identical in all individuals and were highly similar (98%) to the sequence reported for human DNA at D8s1106. The sequences for the other apparent alleles (185, 189) as well as that of a new large product with a length of 200 bp, were quite different from that of the microsatellite locus. These sequences were subjected to a BLAST search (Altschul *et al.* 1997) in order to determine the most similar sequences present in GenBank. The sequences of the 185, 189, and 200 products most resembled DNA sequences from *Escherichia coli* K12, *Clostridium perfringens*, and *Xylella fastidiosa*, respectively, with 96, 88, and 86% identity of the segments compared. The sequence information in GenBank represents a small fraction of extant microorganisms, so these highly similar matches indicate closest known relatives but do not represent definitive identifications of the microorganisms detected. Comparison of the sequences of the primers used with the sequence of the inferred annealing site in the DNA of these microorganisms revealed that amplification occurred despite the presence of numerous primer-template mismatches (Fig. 1b). Inferred mismatches were, however, predominantly on the 5' end of the primer and hence would be expected to have less effect, while the 3' terminal mismatches were of types shown by direct experimentation to permit efficient amplification (Kwok *et al.* 1990).

The fact that evidence of such bacteria was detected in faeces from wild gorillas is not surprising. *E. coli* and *C. perfringens* are known to frequently occur in the intestines of wild and domestic animals (Brynstad & Granum 2002), and *X. fastidiosa* is a common plant pathogen (Keen *et al.* 2000). The potential problem posed by such co-amplification of nontarget DNA has been previously unrecognized. Our results, however, suggest that it does warrant wider attention, most obviously for the increasing number of studies using faeces as a source of DNA. Although the phenomenon appears rare, it nevertheless is a potential threat to the reliability of microsatellite genotyping studies. For example, parent-offspring mismatches in paternity analysis are often based on the nonsharing of alleles at a single locus, indicating that even low levels of error can potentially result in broad misinterpretations of results. While the use of a more stringent annealing temperature may reduce the occurrence of artefacts, in our experience the success rate of amplification of gorilla DNA was also decreased. Thus, we recommend that in addition to multiple

(a)

		10	20	30	40	50	60	70	
		.....	.....	.....	.....	.....	.....	.....	
Human	1	ttgtttacc	ctgcatcact	ggcccagaca	gttgctgcta	accgcgaaca	gagagataga	cagatgatag	70
INT (154)	1	.....	.....	.....	.....	.....TG...	.....	.....	70
ISA (162)	1	.....	.....	.....	.....	.....TG...	.....	.....	70
MAK (185)	1	.....	.....	CAT.A.CCGG	AACA.GA..G	GAATTAGCGC	..TTAT.G.C	GGATATCG.A	70
KRD (189)	1	.....	.....	T.GT.GCCT.	T.AAAAAA.C	GTGG..T.G.	AGT.ACA.TC	..A.A.T.T.	70
INT (200)	1	.....	.....	..CGG..AGTC	TAG.AA.GCT	TTTAAA..AT	AGAG.GCTC.	AG.TA.CCGC	70
		80	90	100	110	120	130	140	
		.....	.....	.....	.....	.....	.....	.....	
Human	71	atagat~gat	agatagatag	atagatagat	agatagatag	atagatagat	aga~~~~~	~~~~~ttcctt	127
INT (154)	71	.....A...	.....	.....	.....	.....	.....TAGA...	.....	132
ISA (162)	71	.....A...	.....	.....	.....	.....	.....TAGATAG	ATAGA.....	140
MAK (185)	71	CCT.GCAC.C	.CG.TT...C	C.GTGGTCCC	GAGGC.C..A	..GA.GC.G.	.AGAAGTGAA	GCTGCGCG.C	140
KRD (189)	71	.CCCC.ATT.	.A...TCG.T	CC..GC.C.A	T..GCCATA	CC.ACAC.G.	GA.GTTTTCG	TAACAGATGA	140
INT (200)	71	.C...AATTG	GATCCCTACA	TG.ATGT.GA	TCC.G.CACC	..GAGCCCC.	TACAACATGG	TGAAG..TA.	140
		150	160	170	180	190	200		
		.....	.....	.....	.....	.....	.....		
Human	128	gcactatgag	caattctgag	aa-----	-----	-----	-----	-----	149
INT (154)	133	.....	.....	..					154
ISA (162)	141	.....	.....	..					162
MAK (185)	141	TGGAC..CGT	.GCCGA.AC.	CTGCACTATG	AGCAATTCTG	AGAA			184
KRD (189)	141	CGG.GCA..A	GCTGA..T..	.CTTAGGGCA	CTATGAGCAA	TTCTGAGAA			189
INT (200)	141	.TCACTGAG.	ATGGAGCTGA	..CTGACCTT	GATTTGGGGC	ACTATGAGCA	ATTCTGAGAA		200

(b)

D8s1106	5'	TTGTTTACCCCTGCATCACT	3'	5'	TTCTCAGAATTGCTCATAGTGC	3'
E. coli		GAAAGTC.ATT.....T.			GATAGCA.....A.....	
C. perf.		GGAA.A..AG...T...T.			A...AT...CTT.....A	
X. fast.		CCC..GG.AAGG.....G.G			GCG.ACA..AC...G..A..G	

Fig. 1 (a) Consensus sequences obtained from amplification of D8s1106 in gorillas. Source individual and apparent allele size (in parentheses) are indicated. The human sequence corresponds to GenBank accession number G09378. Identical sequences for the 154 allele were also found in INT, MAK, JEN and ISA (an individual not showing the longer allele anomalies). An identical sequence for the 189 product was also found in KRD. The symbol ~ represents a space inserted to maintain alignment. (b) Comparison of primer sequences and inferred annealing sites used in amplification of the 185, 189 and 200 products. Corresponding GenBank accession numbers are: *E. coli* = AE000274, *C. perfringens* = AP003193, and *X. fastidiosa* = AE003962.

independent replication of genotypes at each locus (following Taberlet *et al.* 1996 or Morin *et al.* 2001), investigators conducting genotyping studies using noninvasive samples consider the possibility of false alleles and examine the sequence of any anomalous alleles.

## Acknowledgements

This research is part of an ongoing project on the genotyping of the mountain gorillas of the Karisoke Research Center, Rwanda, in collaboration with C. Boesch, M. Robbins, E. Williamson and the Dian Fossey Gorilla Fund International. We thank T. Biedermann, C. Richter, C. Schwarz, and H. Siedel for technical assistance, and the Max Planck Society for financial support.

## References

Altschul SF, Madden TL, Schaffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Bradley BJ, Boesch C, Vigilant L (2000) Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces. *Conservation Genetics*, **1**, 289–292.

Brynstad S, Granum PE (2002) *Clostridium perfringens* and food-borne infections. *International Journal of Food Microbiology*, **74**, 195–202.

Coote T, Bruford MW (1996) Human microsatellites applicable for analysis of genetic variation in apes and Old World monkeys. *Journal of Heredity*, **87**, 406–410.

Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genome wide comparison of DNA sequences between humans and chimpanzees. *American Journal of Human Genetics*, **70**, 1490–1497.

Kayser M, Nürnberg P, Bercovitch F, Nagy M, Roewer L (1995) Increased microsatellite variability in *Macaca mulatta* compared to humans due to a large scale deletion/insertion event during primate evolution. *Electrophoresis*, **16**, 1607–1611.

Keen NT, Korsi Dumenyo C, Yang C, Cooksey DA (2000) From rags to riches: insights from the first genomic sequence of a plant pathogenic bacterium. *Genome Biology*, **1**, 1019.

- Kilger C, Krings M, Poinar H, Pääbo S (1997) 'Colony sequencing': direct sequencing of plasmid DNA from bacterial colonies. *Biotechniques*, **22**, 412–414, 416, 418.
- Kwok S, Kellogg DE, McKinney N *et al.* (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, **18**, 999–1005.
- Morin PA, Chambers KE, Boesch C, Vigilant L (2001) Quantitative PCR analysis of DNA from noninvasive samples for accurate microsatellite genotyping of wild chimpanzees (*Pan troglodytes verus*). *Molecular Ecology*, **10**, 1835–1844.
- 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.
- Taberlet P, Waits LP, Luikart G (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.