*Conservation Genetics* **1:** 289–292, 2000.



# Identification and redesign of human microsatellite markers for genotyping wild chimpanzee (*Pan troglodytes verus*) and gorilla (*Gorilla gorilla gorilla*) DNA from faeces

Brenda J. Bradley[1,2], Christophe Boesch[1] & Linda Vigilant[1*]
[1]*Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, Leipzig D-04103, Germany;* [2]*Anthropological Sciences, State University of New York, Stony Brook, NY 1179-4364, USA; *author for correspondence (E-mail: vigilant@eva.mpg.de)*

Received 28 July 2000; accepted 5 October 2000

*Key words:* chimpanzee, faeces, gorilla, microsatellite, noninvasive

Although in principle a powerful approach to characterize closely related individuals within a population, genetic analysis using variation at microsatellite loci is subject to several technical challenges when applied to wild populations sampled noninvasively. Difficulties arise due to the characteristics of the template DNA, the occurrence of artifacts during PCR, and the typically low success rate. Noninvasive samples provide a low amount of often degraded DNA for amplification, greatly increasing the problem of allelic dropout and subsequent production of inaccurate genotypes (Taberlet et al. 1996). Previous studies of genetic variation in great apes (Gagneux et al. 1999 and references therein) have relied largely on dinucleotide repeat markers, but tri- or tetra-nucleotide repeat markers are now preferred because of the reduced occurrence of amplification artifacts such as "stutter" bands (Edwards et al. 1991) and the greater ease of consistent allele identification using automated technology.

We report here the identification of 9 and 10 microsatellite loci that exhibit variation in wild populations of west African chimpanzees (*Pan troglodytes verus*), and western lowland gorillas (*Gorilla gorilla gorilla*), respectively. Five of the same loci are in use in studies of both species, allowing for eventual interspecific comparisons. To mitigate the aforementioned difficulties, we have identified loci that are not only highly variable, but also produce amplification products less than 300 bp long and thus have a high rate of amplification efficiency and accuracy. New primers that produce shorter products have been designed for several of the loci.

Genomic DNA was extracted from faeces collected fresh in the field and stored dry using silica gel beads (Wasser et al. 1997). Up to 5 g of faeces was collected in a 50 ml tube containing 20 g of silica. In the lab the samples were stored at 4 °C with fresh beads. The QIAamp® DNA Stool Kit (Qiagen) was used to extract genomic DNA from 100 mg of faeces. The DNA was recovered in 200 $\mu$l of elution buffer, aliquotted and stored at –20 °C. PCR amplification was carried out in a total volume of 15 $\mu$l consisting of 2 $\mu$l template, 1x PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl), 2 mM $MgCl_2$, 6 $\mu$g BSA, 250 $\mu$M each dNTP, 200 nM each primer and 0.375 U Amplitaq® Gold (Perkin-Elmer). Amplification conditions on an PTC-200 thermocycler (MJ Research) were: initial denaturation at 95 °C for 3 minutes; 45 cycles of 30 s at 95 °C, 30 s at 55 or 60 °C (see legends, Table 1), 30 s at 72 °C, and a final extension of 30 min at 72 °C. The 5′ end of the forward primer was fluorescently labeled, and amplification products were separated using capillary electrophoresis (ABI 310 PRISM). Alleles were sized relative to an internal size standard (HD400 labeled ROX) using Gene Scan 2.0 (Perkin-Elmer Applied Biosystems). Genotypes were scored multiple times from independent PCRs (twice for heterozygotes, and 4 to 7 times for homozygotes). Each sample was extracted in duplicate. Two negative extraction controls were processed along with each set of 8 to 10 fecal extractions and each set of PCRs contained two or more negative PCR controls as well. For each extract multiple PCRs (three or more) were attempted

*Table 1.* Fourteen microsatellite loci successfully utilized in analysis of DNA from chimpanzee or gorilla feces

| Locus | Label | Primer sequences (5′–3′) | | Allele size range | | Standard DNA* | Marker ref. |
|---|---|---|---|---|---|---|---|
| | | | | Chimpanzee | Gorilla | | |
| D2s1329 | FAM | F: TTGTGGAACCGTTCTCAAAT<br>R: GAAACTTCCACCTGGGTTCT | F–R: | 171–223 | nd | 219 / 223 | G08141 |
| | HEX | F2: ACCGTTCTCAAATACCAGGAATC<br>R2: CCTGGGTTCTTAATTTAACCATAATTC | F2–R2: | 154–206 | nd | 178 / 182 | |
| D9s910 | NED | F: AAGTCAGTTAGCTGAAGGTTGC<br>R: TATATGAAGTGCTTAGAAAAAGTGC | F–R: | 101–122 | nd | 105 / 105 | G08725 |
| D11s2002 | NED | F: CATGGCCCTTCTTTTCATAG<br>R: AATGAGGTCTTACTTTGTTGCC | F–R: | 238–266 | nd | 238 / 242 | G09598 |
| | | R2: AGTGTGAGCCACCACACCAGC | F–R2: | 144–172 | nd | 144 / 148 | |
| D12s66 | NED | F: TCATTTAAGCATTTGAGGGAA<br>R: AGACTTCAAAACAGACACTT | F–R: | 146–178 | nd | 170 / 174 | GBD: 181964 |
| D2s1326 | FAM | F: AGACAGTCAAGAATAACTGCCC<br>R: CTGTGGCTCAAAAGCTGAAT | F–R: | 222–266 | 250–286 | 266 / 270 | G08136 |
| | | R2: AGGGAATTCTCTGAGCTAATAC | F–R2: | 174–218 | nd | 198 / 202 | |
| D5s1470 | HEX | F: CATGCACAGTGTGTTTACTGG<br>R: TAGGATTTTACTATATTCCCCAGG | F–R: | 170–190 | 170–202 | 182 / 186 | G08475 |
| D7s2204 | FAM | F: TCATGACAAAACAGAAATTAAGTG<br>R: AGTAAATGGAATTGCTTGTTACC | F-R: | 229–265 | 217–249 | 241 / 249 | G08635 |
| | | R2: GTTCACTGTAGAGTTCCTTTATGC | F–R2: | 156–192 | nd | nd | |
| D7s817 | FAM | F: TTGGGACCTCTTATTTTCCA<br>R: GGGTTCTGCAGAGAAACAGA | F–R: | 164–200 | 160–196 | 184 / 188 | G08595 |
| | | F2: TAAATCTCTTTATGGCTGACTG | F2–R: | 120–156 | 112–148 | 136 / 140 | |
| vWF | HEX | F: CCCTAGTGGATGATAAGAATAATC<br>R: GGACAGATGATAAATACATAGGATGGATGG | F–R: | 116–148 | 144–160 | 152 / 156 | HUMVWFA31 |
| D1s550 | FAM | F: CCTGTTGCCACCTACAAAAG<br>R: TAAGTTAGTTCAAATTCATCAGTGC | F–R: | nd | 170–194 | 186 / 186 | G07836 |
| D4s1627 | NED | F: AGCATTAGCATTTGTCCTGG<br>R: GACTAACCTGACTCCCCCTC | F–R: | nd | 230–246 | 230 / 246 | G08393 |
| D8s1106 | NED | F: TTGTTTACCCCTGCATCACT<br>R: TTCTCAGAATTGCTCATAGTGC | F–R: | nd | 123–151 | 139 / 143 | G09378 |
| D10s1432 | HEX | F: CAGTGGACACTAAACACAATCC<br>R: TAGATTATCTAAATGGTGGATTTCC | F–R: | nd | 156–176 | 160 / 168 | G08816 |
| D16s2624 | FAM | F: TGAGGCAATTTGTTACAGAGC<br>R: TAATGTACCTGGTACCAAAAACA | F–R: | nd | 128–144 | 132 / 136 | G07938 |

*Commercially available DNA from *Gorilla gorilla* cell line, European Collection of Animal Cell Cultures Ref. No: 89072703
F, original forward primer; R, original reverse primer; F2, redesigned forward primer; R2, redesigned reverse primer; nd, not determined.
The annealing temperature for D12s66, D10s1432, D2s1329 and D7s2204 was 55 °C, for all other loci it was 60 °C

*Table 2.* Variability of loci in wild populations of western chimpanzees and western lowland gorillas

| | Chimpanzee (*P. t. verus*) | | | | Gorilla (*G. g. gorilla*) | | | |
|---|---|---|---|---|---|---|---|---|
| Locus | N | k | He | Ho | N | k | He | Ho |
| D2s1329 | 43 | 7 | 0.769 | 0.860 | nd | | | |
| D9s910 | 48 | 7 | 0.799 | 0.854 | nd | | | |
| D11s2002 | 44 | 8 | 0.800 | 0.841 | nd | | | |
| D12s66 | 48 | 8 | 0.724 | 0.708 | nd | | | |
| D2s1326 | 46 | 12 | 0.887 | 0.804 | 24 | 9 | 0.917 | 0.795 |
| D5s1470 | 47 | 6 | 0.823 | 0.975 | 26 | 8 | 0.845 | 0.885 |
| D7s2204 | 38 | 8 | 0.819 | 0.763 | 14 | 7 | 0.796 | 0.857 |
| D7s817 | 47 | 9 | 0.857 | 0.894 | 14 | 9 | 0.823 | 0.857 |
| vWF | 46 | 6 | 0.629 | 0.522 | 23 | 5 | 0.703 | 0.696 |
| D1s550 | nd | | | | 19 | 7 | 0.831 | 0.895 |
| D4s1627 | nd | | | | 13 | 5 | 0.729 | 0.846 |
| D8s1106 | nd | | | | 11 | 6 | 0.779 | 0.636 |
| D10s1432 | nd | | | | 16 | 6 | 0.837 | 0.875 |
| D16s2624 | nd | | | | 15 | 4 | 0.577 | 0.400 |
| Mean | 45 | 7.9 | 0.790 | 0.802 | 17.5 | 6.6 | 0.772 | 0.774 |
| Total exclusionary power (first parent): | | | | 0.993 | 0.991 | | | |
| Total exclusionary power (second parent): | | | | 0.999 | 0.999 | | | |

N, number of individuals typed
k, number of alleles observed
He, expected heterozygosity
Ho, observed heterozygosity

at each locus. A successful PCR reaction was defined as one producing specific product in the expected size range.

The fourteen microsatellite loci (Table 1) are described in the references as tetranucleotide repeats in humans, with the exception of D9s910, a trinucleotide repeat. The first nine primers were used in chimpanzees, and the final ten were used in gorillas. Results from a readily available standard DNA are included in Table 1, along with label information, allowing laboratories using the same loci but different allele scoring systems to compare results. Initial results using chimpanzee DNA revealed that amplifications at five loci (D2s1329, D11s2002, D2s1326 D7s2204 and D7s817) were successful in only 62 to 75% of attempts. Since amplifications at the other four loci were more successful (77 to 89% of attempts) and these four loci also have smaller average allele sizes, we reasoned that the DNA from these faecal samples was likely to be degraded, and hence amplification of shorter rather than longer products would be more efficient. Using the publicly available human sequence information for these loci (see genome database ID numbers in Table 1), we designed new primers to amplify shorter products for the five loci. New primers were designed using three considerations: i) no obvious potential hairpin structures, ii) similar melting temperature to other primer in pair (calculated as in Breslauer et al.1986) and iii) primer sequence ending with a C or G for high specificity (Kwok et al. 1990). The redesign of primers resulted in significant increases in the success of amplification from chimpanzee faecal DNAs using D2s1329, D11s2002, D7s817, and D7s2204 (df = 1; $\chi^2$ = 12.3, 35.6, 6.9 and 7.3 respectively, $p < 0.01$) and a statistically insignificant increase in amplification success rate for D2s1326. The average success rate of amplification from chimpanzee faecal DNA was 82% (973 attempts) with the least successful locus (D7s2204F-R) at 62% and the most successful locus (D11s2002 F-R2) at 95%. In general, success rates from gorilla faecal DNA were lower and averaged 56% (672 attempts). This lower value may reflect differences in dietary components that have been associated with PCR inhibition (Monteiro et al. 1997). We guarded against contamination (within and between

species) using several means: i) the multiple negative controls mentioned above, ii)extensive replication of results using the multiple tubes approach (Taberlet et al. 1996), iii) genotyping of all laboratory workers, and iv) checking for consistent allele-sharing results from known mother-offspring pairs. We estimate our error rate at less than 1%, based upon the incidence of inconsistent, unreproducible results.

Table 2 summarizes the variation at the loci using results derived from ongoing genotyping studies of three habituated communities of chimpanzees from the Taï National Park, Côte d'Ivoire and multiple unhabituated gorilla groups at Mondika Research Center, Central African Republic. The high level of variation at these loci will allow for unambiguous assessment of paternity, an important component to our studies of reproductive strategies in these species. The same systems are being used in our laboratory to also analyze genetic variation and social structure in bonobos (*Pan paniscus*), two communities of east African chimpanzees (*Pan troglodytes schweinfurthii*), two populations of mountain gorillas (*Gorilla gorilla beringei*), and a population of eastern lowland gorillas (*G. g. graueri*). Thus, genetic data derived in this way will in principle allow us to investigate the genetic structure of individual populations, and potentially make broader interpopulation, intraspecific and interspecific comparisons.

**Acknowledgements**

The authors thank H. Siedel and A. Abraham for technical assistance, and D. Doran and field assistants at Mondika and Taï for sample collection. We would also like to thank P. Morin for information on human primers used in baboon studies. B. Bradley is financially supported by the Max Planck Society, NSF, Wenner-Gren, Leakey Foundation, and Sigma Xi.

## References

Breslauer KJ, Frank R, Blöcker H, Marky LA (1986) Predicting DNA duplex stability from the base sequence. *Proc. Natl. Acad Sci.*, **83**, 3746–3750.

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.*, **49**, 746–757.

Gagneux P, Boesch C, Woodruff D (1999) Female reproductive strategies, paternity, and community structure in wild West African chimpanzees. *Ani. Beh.*, **57**, 19–32.

Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, Levenson C, Sninsky JJ (1990) Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies. *Nucl. Acids Res.*, **18**, 999–1005.

Monteiro L, Bonnemaison D, Vekris A, Petry KG, Bonnet J, Vidal R, Cabrita J, Megraud F (1997) Complex polysaccharides as PCR inhibitors in feces: Helicobacter pylori model. *J. Clin. Microbiology*, **35**, 995–998.

Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, Waits LP, Bouvet J (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucl. Acids Res.*, **24**, 3189–3194.

Wasser S, Houston C, Koehler G, Cadd G, Fain S (1997) Techniques for application of fecal DNA methods to field studies of Ursids. *Mol. Ecol.*, **6**,1091–1097.