



First generation microarray-system for identification of primate species subject to bushmeat trade

A.-C. Rönn¹, O. Andrés², F. López-Giráldez^{2,3}, C. Johnsson-Glans¹, E. J. Verschoor⁴, X. Domingo-Roura^{2,†}, M. W. Bruford⁵, A.-C. Syvänen¹, M. Bosch^{2,*}

¹Molecular Medicine, Department of Medical Sciences, Uppsala University, University Hospital, Entr. 70, 3rd floor, Res. Dep. 2, 75185 Uppsala, Sweden

²Genètica de la Conservació Animal, Institut de Recerca i Tecnologia Agroalimentàries, Carretera de Cabrils Km2, 08348 Cabrils, Spain

³Department of Ecology and Evolutionary Biology, Yale University, PO Box 208106, Prospect Street 165, New Haven, Connecticut 06520, USA

⁴Department of Virology, Biomedical Primate Research Centre, Lange Kleiweg 139, 2288 GJ Rijswijk, The Netherlands

⁵Cardiff School of Biosciences, Cardiff University, PO Box 915 Cathays Park, Cardiff CF10 3TL, UK

ABSTRACT: About a quarter of non-human primate species are threatened by extinction in the near future. Loss of habitat, disease and illegal hunting, especially for the bushmeat trade, are major causes of concern. Here, we develop an identification tool for primate genera using diagnostic nucleotide positions in the epsilon globin gene, apolipoprotein B gene and mitochondrial 12S rRNA. We identified 111 diagnostic nucleotide positions suitable for genotyping by a minisequencing assay in a microarray format. To show the applicability of the microarray, we typed 70 non-human primates representing all primate infraorders. Sixty-five samples were assigned to the correct infraorder, and 32 were assigned to the correct genus (the highest level of taxonomic resolution attempted here). Our results show that it is feasible to distinguish among a high number of primate taxa if the system allows hierarchical assignment of the samples at different taxonomic levels and includes both taxon-specific and redundant positions.

KEY WORDS: Primates · Bushmeat · Minisequencing microarray · Epsilon globin · Apolipoprotein B · 12S rRNA

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Non-human primates are spread worldwide and are important taxa in conservation, evolution and biomedicine. Traditional classification (e.g. Rowe 1996) has included more than 230 species in this order, but nowadays more than 350 different species are proposed (Groves 2001). About half of all non-human primate species are considered to be of conservation concern, with conservation categories ranging from 'lower risk' to 'critically endangered,' and as many as 26% of the species are in immediate danger of extinction within 20 yr (Mittermeier et al. 2005). All non-human primate species are included in Appendix 2 of the

Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

The greatest factors in primate population declines are habitat destruction, hunting and disease (Walsh et al. 2003). Primates are an important source for food protein, particularly in parts of Africa, Asia and the Amazon region of South America. This threat involves a multitude of species, but is higher for certain primate taxa, such as the drill *Mandrillus leucophaeus*, which is under imminent threat of extinction due to hunting (Wild et al. 2005). Much of the bushmeat trade is illegal, as it involves species protected from commercial trade. In addition, primates are also hunted for skins and body parts, as ingredients in traditional medicine, as pets and for exhibition.

*Corresponding author: Email: montse.bosch.gallego@gmail.com
†Deceased

From 1990 to 2004 more than 5.5 million primate specimens were traded legally or illegally according to the CITES trade database (www.cites.org). The illegal trade represents less than 1% of that number. However, this percentage and therefore the total number of specimens traded are underestimated, since illegal trade usually is not documented. Forensic laboratories frequently encounter powdered or processed samples lacking any morphological detail for identification purposes. In these cases, molecular genetic identification of the species to which the sample belongs becomes the preferable option. These techniques may exploit nucleotide positions that are fixed within a given taxon (e.g. a species) and variable across taxa. Methods currently employed to identify taxon-specific markers are based on sequencing mitochondrial DNA (especially the cytochrome b, cytochrome oxidase I, 12S rRNA, 16S rRNA genes and the D-loop region; Prakash et al. 2000, Hsieh et al. 2001, Dalebout et al. 2004, Balitzki-Korte et al. 2005), species-specific PCR (Nekrutenko et al. 2000) and restriction fragment length polymorphisms (RFLPs; Wolf et al. 1999, Pfeiffer et al. 2004). However, none of these techniques fulfill all the requirements—such as good multiplexing capacity and high accuracy—desirable for an effective methodology with forensic applications. Moreover, reliable and accurate methods for determining species in a broad spectrum of degraded and processed substrates are required. In these cases, identification can only be based on the analysis of very short DNA fragments. Thus, the use of single nucleotide polymorphisms (SNPs) for identification of forensic species might be appropriate. Microarray-based DNA analysis provides a powerful tool for fast and cost-effective detection of a large number of SNPs simultaneously (Syvänen 2005). DNA microarrays allow the examination of complex mixtures of PCR products and, potentially, the identification of a number of species simultaneously. Similar methods have already been used in ecology for simultaneous detection of 5 marine pathogens (Gonzalez et al. 2004), but not extensively for species identification in food and forensic samples (Teletchea et al. 2008).

We developed a genotyping system based on a ‘minisequencing’ single base extension assay in a microarray format. Here, we used diagnostic nucleotide positions in 2 nuclear genes, the epsilon globin and apolipoprotein B gene sequences, to differentiate among infraorders and Catarrhine subfamilies, respectively, as nuclear sequences are more informative for distinguishing divergence events from order-family level for mammals than mitochondrial DNA (Springer et al. 2001) and there were a high number of sequences for these genes in primates available in the literature. In this study, we used the mitochondrial 12S

rRNA gene to provide diagnostic nucleotide positions to distinguish among genera due to its higher level of variation (Balitzki-Korte et al. 2005). This multiplexed format was designed to allow the determination of the genus to which an unknown primate DNA sample belongs (we did not attempt to assign samples to species level for this first attempt), and is focused on species threatened by extinction and target species for the bushmeat trade. We show proof-of-principle that this microarray system will potentially be a valuable tool in wildlife management and law enforcement, helping to prevent illegal trade of endangered primates.

MATERIALS AND METHODS

Samples. Primate samples were obtained from the INPRIMAT sample collection (www.inprimat.org). For tissue and blood samples DNA was extracted using a DNeasy Tissue kit (QIAGEN Sample & Assay Technologies) following the manufacturer's instructions. Initial amounts were 25 mg for muscle tissues and 100 µl for blood samples. DNA from the INPRIMAT DNA collection was also analyzed. Non-primate samples (2 *Ovis aries*, a dermoptera, a scandentia and a tupaia) were provided by M.W. Bruford and H. Zischler (INPRIMAT Consortium). The 70 primate samples used for testing the applicability of the microarray system are shown in Supplement 1, Table S1 available as supplementary material at www.int-res.com/articles/suppl/n009p133_app.pdf.

Identification of diagnostic nucleotide positions. We used diagnostic nucleotide positions previously described in the 341 bp fragment of the epsilon globin gene to distinguish among primate infraorders (Bosch et al. 2006). The epsilon globin gene was also used as a control to demonstrate the presence of DNA as it is a highly conserved gene and the designed primers were expected to recognize non-primate samples. Moreover, it was reasonable to assume that non-primate samples would generate a genotype that was distinct from that in any primate infraorder to indicate a non-primate individual. We identified previously unknown diagnostic nucleotide positions by sequencing a fragment of the apolipoprotein B gene (ApoB) to differentiate among Catarrhine subfamilies. For ApoB we used sequences from GenBank for designing universal primate PCR primers ApoBF (5'-GAC TGC AAA GGC AGA CAA CT-3') and ApoB1R (5'-TGA ATA TCA CAA ATT CTC AGG GTG-3') to amplify a 550 bp fragment. PCR reactions were performed at 94°C for 5 min, followed by 45 cycles at 94°C for 45 s, 58°C for 45 s and 72°C for 1 min, and 72°C for 7 min using 1.5 mM MgCl₂, 3.25 pmol of each primer, 320 µM of each dNTP, and 0.65 units of EcoTaq DNA polymerase (Eco-

gen) in 20 μ l. The PCR products were digested with 5 units of Shrimp Alkaline Phosphatase (USB) and 2 units of *ExoI* (USB) for 1 h at 37°C, followed by inactivation at 75°C for 15 min. PCR products were sequenced using a Big Dye Terminator Kit (Applied Biosystems) and run on an ABI3100 automated DNA sequencer. We compiled 34 partial ApoB sequences from Catarrhine species, of which 9 were from GenBank and 25 were determined by sequencing for this study. Of these sequences, 20 belonged to the Cercopitheciinae subfamily, representing 8 of the 11 genera, 9 sequences belonged to the Colobinae subfamily, with 6 of 8 genera represented, 4 belonged to the Hominae family, with all the genera represented, and 1 belonged to the Hylobatidae family (Table 1). We aligned sequences using ClustalW (Thompson et al. 1994) and located the positions that were fixed within a subfamily and that varied among subfamilies.

The mitochondrial 12S rRNA gene was sequenced to distinguish among primate genera. A 389 bp fragment was amplified essentially according to Kocher et al. (1989) using primers L1091 (5'-CAA ACT GGG ATT AGA TAC CCC ACT AT-3') and H1478 (5'-AGG GTG ACG GGC GGT GTG T-3'). PCRs used 2 mM MgCl₂, 50 pmol of each primer, 200 μ M of each dNTP, and 2 units AmpliTaq Gold DNA polymerase (Applied Biosystems) in 50 μ l of PCR buffer (final concentration 10 mM TrisHCl/50 mM KCl) supplemented with 0.01% bovine serum albumin (Applied Biosystems) and proceeded at 95°C for 15 min, followed by 35 cycles at 95°C for 20 s, 55°C for 20 s, 72°C for 40 s, and 72°C for 5 min.

We compiled 722 partial primate 12S rRNA sequences representing most primate genera as reference (Table 1). Of these sequences, 410 belonged to Catarrhini, 114 to Platyrrhini, and 198 to Strepsirrhini. We aligned the sequences and identified diagnostic nucleotide positions using the Diagnostic Position Search (DIAPOS) program (F. López-Giráldez et al. unpubl. data). The software creates a degenerate consensus sequence for each genus, identifies the diagnostic nucleotide positions and orders them according to different parameters, such as the number of nucleotide differences at this position, the flanking regions, or the polymorphic information content (PIC).

Minisequencing on oligonucleotide arrays. Minisequencing primers that anneal immediately adjacent to each diagnostic nucleotide position were designed using the sequence alignments created by ClustalW. Positions that were polymorphic among taxa, but monomorphic within taxa, and with a 20 bp flanking region conserved among taxa, were favored. The NetPrimer software (www.premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html) was used for primer design to avoid hairpin loop formation in the mini-

sequencing primers and to ensure an approximate melting temperature of 55 to 60°C.

The minisequencing primers used in the final genotyping system are provided in Table 2. The minisequencing primers were covalently coupled to CodeLink™ Activated Slides (GE Healthcare) via 5'-NH₂-groups according to the manufacturer's instructions. The primers were printed in an 'array-of-arrays' format onto the slide to form 80 subarrays per slide. This configuration allows simultaneous genotyping of 80 samples for all selected diagnostic nucleotide positions. The diameter of the printed spots was 130 μ m and they printed at a spot-to-spot distance of 185 μ m using a ProSys 5510A instrument (Cartesian Technologies) with 4 Stealth Micro Spotting Pins (SMP3, TeleChem International). The excess of amine-reactive groups were deactivated by immersing the arrayed slides into 50 mM ethanolamine, 100 mM Tris-HCl, pH 9.0 and 0.1% SDS for 30 min at 50°C.

Sequences from GenBank and the NetPrimer software were used to design PCR primers 12S1F (5'-GGT TTG GTC CTA GCC TTT CTA-3') and 12S1R (5'-CTT ACC ATG TTA CGA CTT GTC TC-3'), for amplifying a 900 bp fragment of mRNA that contained the 12S rRNA gene. Multiplex PCRs, including these primers for 12SrRNA, the primers ApoBF and ApoB1R for ApoB and EGF (5'-GAA GCC TTG GGC AGG TAA) and EGR (5'-TGA AGT TCT CAG GAT CCA C) for the epsilon globin gene (Bosch et al. 2006), were performed using 20 to 50 ng of DNA, 5 mM MgCl₂, 0.75 pmol of each primer, 100 μ M of each dNTP and 1.5 U of Smart Taq Thermostable DNA Polymerase (Naxo) in 15 μ l of PCR buffer containing 10 mM Tris-HCl, 50 mM KCl and 0.8% Nonidet 40 (Naxo), supplemented with 0.5 g l⁻¹ of bovine serum albumin.

The microarrays carrying minisequencing primers were placed in a custom-made reaction rack for individual array hybridization, covered by custom-made silicon rubber grid to form individual chambers as described in detail by Pastinen et al. (2000), and pre-heated to 37°C in an oven. The PCR products, including a negative PCR control and water controls, were denatured at 95°C for 2 min and placed on ice, mixed with 14 μ l annealing buffer consisting of 4.5 M NaCl, 50mM Tris-HCl pH 8.0, 5mM EDTA, and added to the silicon rubber chambers on the microarray slide for hybridization to the immobilized primers for 2 h at 37°C. Moist paper towels and cling film were placed on top of the rack to create humid chambers during hybridization. After hybridization, the rack was disassembled and the microarray rinsed in 0.1 M NaCl, 5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 0.1% Triton X-100. The rubber grid and microarray were briefly rinsed in water and air-dried. The microarray slide was placed back into the rack

Table 1. Sequences used for determining diagnostic nucleotide positions. Sequences include those available in the literature and those produced in this study

| Infraorder | Family | Subfamily | Genus | Genes/number of sequences | | | |
|---------------------|-----------------------|--------------------|-----------------------|---------------------------|-------------------|----------|----|
| | | | | Epsilon globin | Apolipoprotein B | 12S rRNA | |
| Catarrhini | Cercopithecidae | Cercopithecinae | <i>Allenopithecus</i> | | | 4 | |
| | | | <i>Cercocebus</i> | | | 15 | |
| | | | <i>Cercopithecus</i> | 3 | 2 | 49 | |
| | | | <i>Erythrocebus</i> | 2 | 2 | 3 | |
| | | | <i>Lophocebus</i> | | 1 | 10 | |
| | | | <i>Macaca</i> | 3 | 5 | 146 | |
| | | | <i>Mandrillus</i> | 1 | 3 | 8 | |
| | | | <i>Miopithecus</i> | | | 7 | |
| | | | <i>Papio</i> | 1 | 6 | 27 | |
| | | | <i>Theropithecus</i> | 1 | 1 | 3 | |
| | | | <i>Chlorocebus</i> | | | 26 | |
| | | | Colobinae | <i>Colobus</i> | 2 | 2 | 6 |
| | | | | <i>Nasalis</i> | | 1 | 1 |
| | | | | <i>Presbytis</i> | 1 | 2 | 2 |
| | <i>Ptilocolobus</i> | | | | 1 | | |
| | <i>Pygathrix</i> | | | 1 | 7 | | |
| | <i>Semnopithecus</i> | 1 | | 2 | 5 | | |
| | <i>Trachypithecus</i> | 1 | | 1 | 2 | | |
| | Hominidae | Ponginae | <i>Pongo</i> | 1 | 1 | 53 | |
| | | | Hominidae | <i>Gorilla</i> | 1 | 1 | 6 |
| | | <i>Pan</i> | | 3 | 2 | 14 | |
| <i>Homo sapiens</i> | | 1 | | | 2 | | |
| <i>Hylobates</i> | | 2 | 1 | 13 | | | |
| Platyrrhini | Cebidae | Hapalinae | <i>Callimico</i> | 2 | | 2 | |
| | | | <i>Callithrix</i> | 3 | | 15 | |
| | | | <i>Leontopithecus</i> | 1 | | 8 | |
| | | | <i>Saguinus</i> | 1 | | 18 | |
| | | Cebinae | <i>Cebus</i> | 2 | | 23 | |
| | | | Chrysotrichinae | <i>Saimiri</i> | 3 | | 7 |
| | | Nyctipithecidae | | <i>Aotus</i> | 1 | | 13 |
| | | | Pitheciidae | Callicebinae | <i>Callicebus</i> | 2 | |
| | | Pitheciinae | | | <i>Cacajao</i> | 1 | |
| | | | | <i>Chiropotes</i> | 1 | | 1 |
| | Atelidae | Mycetinae | <i>Pithecia</i> | 1 | | 2 | |
| | | | <i>Alouatta</i> | 2 | | 2 | |
| | | Atelinae | <i>Ateles</i> | 2 | | 14 | |
| | | | <i>Brachyteles</i> | 1 | | 1 | |
| | | | <i>Lagothrix</i> | 1 | | 7 | |
| | Strepsirrhini | Cheirogalidae | <i>Cheirogaleus</i> | 1 | | 2 | |
| | | | <i>Microcebus</i> | 1 | | 1 | |
| Daubentoniidae | | <i>Daubentonia</i> | 1 | | 1 | | |
| Galagonidae | | <i>Galago</i> | 1 | | 4 | | |
| | | <i>Otolemur</i> | 1 | | 4 | | |
| | | <i>Euoticus</i> | | | 1 | | |
| Indridae | | Indrinae | <i>Avahi</i> | | | 3 | |
| | | | <i>Indri</i> | | | 2 | |
| | | | <i>Propithecus</i> | 1 | | 6 | |
| Lemuridae | | <i>Eulemur</i> | 1 | | 41 | | |
| | | <i>Hapalemur</i> | | | 20 | | |
| | | <i>Lemur</i> | 1 | | 12 | | |
| | | <i>Varecia</i> | | | 7 | | |
| Loridae | | Perodictinae | <i>Perodicticus</i> | 1 | | 2 | |
| | | | <i>Arctocebus</i> | | | 3 | |
| | Lorinae | <i>Loris</i> | 1 | | 4 | | |
| | | <i>Nycticebus</i> | 1 | | 4 | | |
| Megaladapidae | <i>Lepilemur</i> | | | 82 | | | |
| Tarsiiforme | | <i>Tarsius</i> | 1 | | | | |

with the grid, and the assembly was preheated to 50°C. A minisequencing reaction mixture containing 0.35 µM of each of ddATP (labelled with Texas Red), ddCTP (labelled with TAMRA), ddGTP (labelled with R110) and ddUTP (labelled with Cy5), 3.5 mM MgCl₂, 0.067 U µl⁻¹ KlenThermas DNA polymerase in 15 µl of buffer, was added to the microarray and the primer extension reactions were allowed to proceed for 15 min at 50°C. The microarray was then washed briefly with water, denaturated with 50 mM NaOH for 2 min at room temperature and washed twice for 5 min in 0.2× SSC containing 0.1% SDS, preheated to 65°C, using a shaker and avoiding exposure to light. The microarray was then briefly rinsed with water and air-dried.

The fluorescence signals on the microarray were measured using a ScanArray® 5000 instrument (Perkin-Elmer Life Science). The signals from the 4 fluorophores were normalized to similar intensities below saturation by adjusting the photomultiplier tube gain of the ScanArray instrument. The fluorescent signals were quantified with the QuantArray® analysis 3.1 software (Perkin-Elmer Life Sciences), and the QuantArray file was exported to the SNPSnapper v4.1 software (provided by J. Saharinen, www.giu.fi) for genotype assignment by scatter plot analysis (Lindroos et al. 2002). The scatter plot for each assay was constructed by plotting the log₁₀ of the sum of the fluorescence signal intensities from the 2 possible nucleotides in the assay against the fraction of the signal intensity from one of the nucleotides. Genotypes were assigned for each individual sample based on manual inspection of the clusters. Data points from samples that clustered close to either of the vertical axes and were above

Table 2. Sequences of the minisequencing primers in the epsilon globin (EG), apolipoprotein B (ApoB) and 12s rRNA gene used on the final microarray for applicability testing. Degenerate sites: M = A+C, R = A+G, W = A+T, Y = C+T, S = G+C, K = G+T, H = A+C+T, B = G+T+C

| Gene | Primer | Primer sequence 5' - 3' | Nucleotide variation |
|-------------------------------------|--------|-----------------------------------|----------------------|
| EG | 1 | TAY CCC TGG ACC CAG AGA TT | T/C |
| | 2 | BCC TCT GCY ATC CTG GG | C/A |
| | 3 | GGC CTT GAC CTT GGG GTT | G/T |
| | 4 | TCA AGM YYA CCT TTG CTA AGC T | G/A |
| | 5 | AGG GYR CCT TTG CTA AGC T | G/A |
| ApoB | 1 | CTG AAG TTC AGC YCC AGT GA | T/A |
| | 2 | GCA GAG GCT CAG AGA TGT TAA RT | A/T |
| | 3 | AAG CAG AGG CTC AGA GAT GTT AA | A/T |
| | 4 | AGA ACT CAG CCA GAA GCT CCA | G/C |
| 12S rRNA Strepsirrhini (S) | 1 | CTC CTA GAG GAG CCT GTT CTA | T/C |
| | 2 | CCC CTA GAG GAG CCT GTT CTA | T/C |
| | 3 | CGA TAA ACC CCG ATA AAC C | T/C |
| | 4 | ATA TAC CGC CAT CCY CAG C | T/C |
| | 5 | GAC TAC ACC TTG ACC TAA CGT TTT | A/T |
| | 6 | GTT AGG TCA AGG TGT AGT CTA TGR R | G/A |
| | 7 | ATG TAG CCC ATT TCT TYC C | A/G |
| | 8 | CTA TTC TTA RTT TAC TRC TAA ATC C | T/G |
| | 9 | TAG YAG TAA AYY AAG AAT AGA GAG C | A/C |
| | 10 | TSA ACA AAG YTA YTC GCC AGA G | C/T |
| | 11 | AAC AAM RTT AYT CGC CAG AGT ACT A | C/T |
| | 12 | CCG CCA AGT CCT TTG AGT TT | T/C |
| | 13 | AGC TTA AAA CTC AAA GGA CTT GG | C/G |
| | 14 | CTA GAG GGG ATT GAA GCA CC | G/C |
| | 15 | AAG GAC TTG GCG GTG CTT | T/C |
| 12S rRNA Platyrrhini (P) | 1 | CTG GCG GTG CTT TAY ATC C | G/T |
| | 2 | CAG CCT GTA TAC CGC CAT C | T/C |
| | 3 | CGC CAY CTT CAG CAA ACT C | C/T |
| | 4 | CGC CAY CTT CAG CAA ACT CC | C/T |
| | 5 | GCC AYC TTC AGC AAA CTC CTT AA | A/T |
| | 6 | TTT TAT GAT RAT RYT TCT GCT TAC T | T/C |
| | 7 | AAA CTC CYT AAW GAT YGT GAA GTA A | G/C |
| | 8 | AAA CTC CYT AAW GAT TGC GAA GTA A | G/C |
| | 9 | ACG TTT TTA TGA TRA TAC TYC TG | C/G |
| | 10 | GCT TAG CCC TAA ACC THA ATA AY | T/C |
| | 11 | AAT GGG CTA CAT TTT CTA AM | C/T |
| | 12 | AAT GGG CTA CAT TTT CTA AMY | C/T |
| | 13 | GTG TTC TGG CGA ATR AWT TTG TT | G/A |
| | 14 | TTA GAT TTC ATA ARG GYT RTC GT | G/A |
| | 15 | CCA CCT TRG GYC TTT ARA TTT CA | T/C |
| | 16 | CAC CTT GGR CCC TTA GRT TTC A | T/C |
| | 17 | CGC CAG AAC ACT ACA AGC CA | C/T |
| | 18 | CAC CGC CAG GTC CTT TGA | G/A |
| | 19 | GGA CCT GGC GGT GCT TTA | C/T |
| 12S rRNA Cercopithecinae (Ce) | 1 | GAT CCA CCC CAC CCT CTC T | T/C |
| | 2 | ATG GCG GTA TAT AGG CTG AGC | A/G |
| | 3 | TGA AGA TGG CGG TAT ATA GGC T | G/A |
| | 4 | CCC ACC CTC TCT TGC TYA G | C/T |
| | 5 | TTG TGG CCT TCA TCA GGG TT | T/C |
| | 6 | GCC CAT TTT TTA CCR TCT CAT | G/A |
| | 7 | CAC CGT CAA GTC CTT TGA GTT T | T/C |
| 12S rRNA Colobinae (Co) | 1 | TGG GCR TTT GTG CTT ACT TY | G/A |
| | 2 | GTG GGC GTT TGT GCT TAC TT | T/C |
| | 3 | ACG TTT TTG CGT GGG C | G/A |
| | 4 | ACG TTA GST CAA GGT GTA GCY | T/C |
| | 5 | ACG TTA GGT CAA GGT GTA GCC | T/C |
| | 6 | CAC YGC CAA GTC CTT TGA | G/T |

Table 2 (continued)

| Gene | Primer | Primer sequence 5' - 3' | Nucleotide variation |
|----------------------------|-----------------------------------|-----------------------------------|----------------------|
| 12S rRNA Hominoidea (H) | 7 | CTT AAA AYT CAA AGG ACT TGG C | G/A |
| | 8 | CTC TAG GKG GRT GTG AAG CAC | C/T |
| | 9 | GGT KGA TCG GGR TTA TCG | A/G |
| | 1 | AAC CCC GWT CAA CCT CAC C | A/G |
| | 2 | TGG RTG YTT GCG CTT ACT T | T/C |
| | 3 | GAC GTT AGG TCA AGG TGT AGC C | C/T |
| | 4 | CCC ATT TCT TGC CAC CTC AT | G/A |
| | 5 | TCA TAA GGG TTA TCG TAG TTT TCT G | G/A |
| | 6 | CTT AGG TTT CAT AAG GGT TAT CGT A | G/A |
| 7 | TTA RTT TAC TGM TAA ATC CWC CTT C | G/T | |
| 8 | CCT GTT CAA CTA AGC ACT CTA YTM T | T/C | |
| 9 | CCT GTT CAA CTA AGC ACT CTA CTC T | T/C | |

(Pfinder et al. 2004). Neighbour-joining phenograms of the epsilon globin, ApoB and 12S rRNA Colobinae sequences are shown as examples in Supplement 1, Figs. S2 to S4. The topology of the trees showed an unambiguous discrimination among the different primate groups. There were only a few exceptions at the genus level with the 12S rRNA sequences. Three sequences from *Cercocebus torquatus* grouped with sequences from *Lophocebus*. The results of BLASTn searches of these sequences showed the highest simi-

the background signals from negative controls were assigned a genotype. Scatter plots of the fluorescence signals used to assign the genotypes in 2 assays are shown as an example, in Supplement 1, Fig. S1.

RESULTS

Identification of diagnostic nucleotide positions

Our system combines variable nucleotide positions in nuclear and mitochondrial genes to be genotyped for hierarchical assignment of an unknown primate sample to its genus (Fig. 1). For the first step we used the diagnostic nucleotide positions in the epsilon globin gene to distinguish among primate infraorders. Because the Catarrhini infraorder showed a particularly high diversity, a second step using variants of the ApoB gene was included to distinguish among Catarrhine subfamilies. In the final step, the genus of a sample was determined using 12S rRNA sequences. This approach is illustrated in Fig. 2 by examples for *Pan*, *Lemur* and Tarsiiformes.

To evaluate whether the selected genes were appropriate to differentiate among the primate groups at the 3 levels described above, we constructed neighbour-joining trees

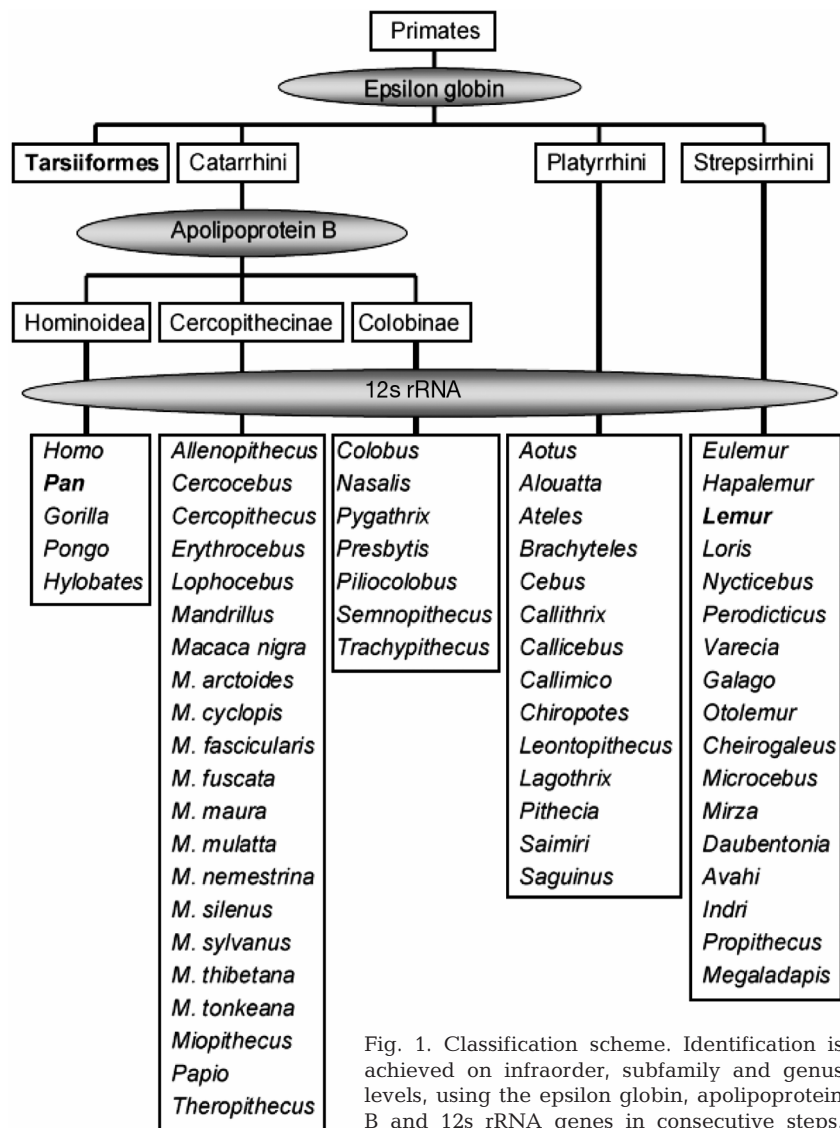


Fig. 1. Classification scheme. Identification is achieved on infraorder, subfamily and genus levels, using the epsilon globin, apolipoprotein B and 12s rRNA genes in consecutive steps. The 3 cases used as an example of sample identification in Fig. 2 are shown in bold

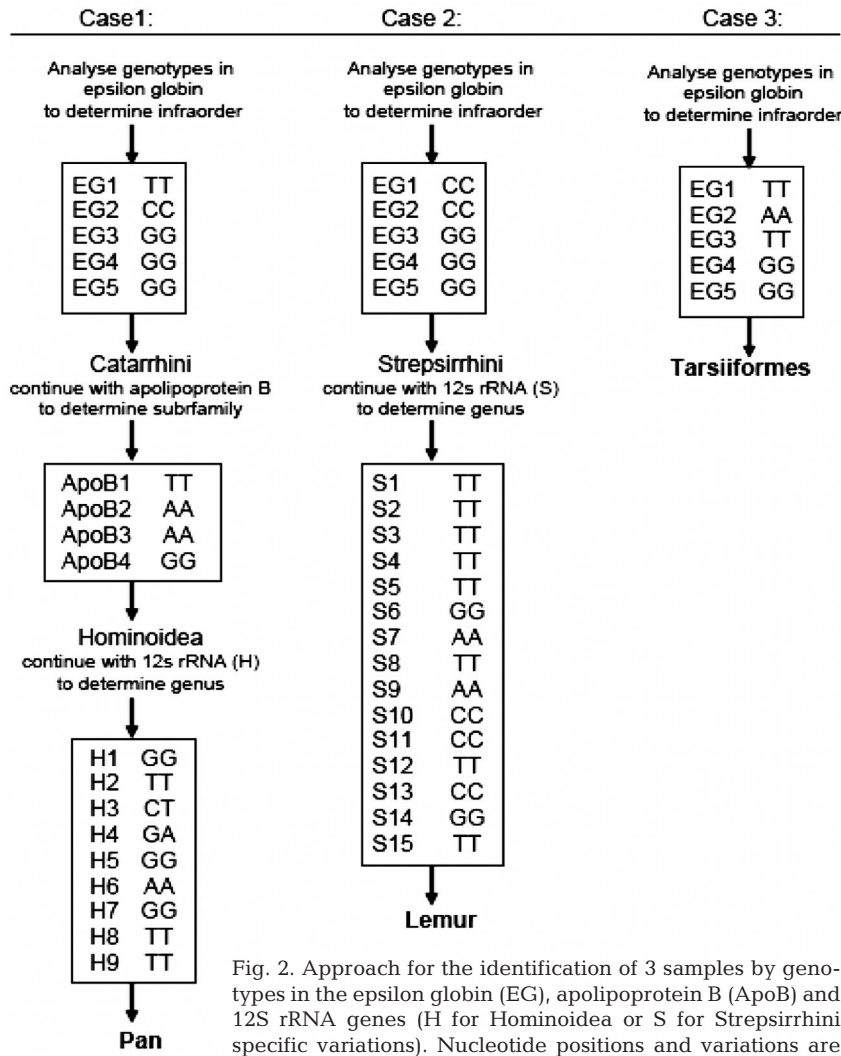


Fig. 2. Approach for the identification of 3 samples by genotypes in the epsilon globin (EG), apolipoprotein B (ApoB) and 12S rRNA genes (H for Hominoidea or S for Strepsirrhini specific variations). Nucleotide positions and variations are defined in Table 1

larity with sequences of *C. aterritus*, a species now considered as *L. aterritus* (Groves 2001). Bootstrap values were low in some cases and, at the genus level, some groups, such as *Cercopithecus*, *Mandrillus*, *Semnopithecus* or *Trachypitecus*, did not appear to be monophyletic. Both of these phenomena may be partially attributable to the limited size of the DNA fragments used. However, taken together, these results suggest that the fragments we selected seem to be appropriate for our analysis.

We identified 12 informative nucleotide positions at the infraorder level in the epsilon globin gene. Just 3 of these were sufficient to unambiguously distinguish among the 4 primate infraorders. At the Catarrhini subfamily level, we identified 8 informative positions in the ApoB gene to distinguish between Colobinae, Cercopithecinae and Hominidae subfamilies. At the genus level, we identified 111 diagnostic nucleotide positions in the 12S rRNA gene. We identified 18, 9, 11, 32 and 21 diagnostic nucleotide positions to distinguish among the

genera included in Cercopithecinae, Colobinae, Hominidae, Platyrrhini and Strepsirrhini, respectively. We included redundant positions to increase the reliability of the test. While we did not find sufficient sequence polymorphism to differentiate between *Cercocebus* and *Lophocebus* and between *Cercocebus* and *Mandrillus*, at least 3 diagnostic positions distinguished *Lophocebus* and *Mandrillus*. As expected, we were unable to identify positions to differentiate between *Semnopithecus* and *Trachypithecus*, since these groups were grouped in the same clade in the tree constructed to study the suitability of the fragments used. Despite the fact that the taxa were separated in the tree, we could not identify diagnostic nucleotide positions with a flanking region suitable for designing primers to distinguish between *Chlorocebus* and *Cercopithecus*, and between *Papio* and *Theropithecus*. In summary, the 111 sequence differences identified were able to differentiate 45 out of the 65 primate genera currently recognized.

Genotyping

We designed a microarray-based minisequencing system for the selected 111 diagnostic nucleotide positions. Based on initial experiments (data not shown), we were able to design 68 successful minisequencing assays (Table 2). The reason the assay conversion rate was below the theoretical value of 100% is probably that genetic variation in the primer binding sites remained undetected due to the limited number of sequences in the alignments used for selection of the nucleotide positions and for primer design. The successful assays for 68 nucleotide positions were selected for an experimental applicability test of the system.

Applicability of the microarray

To test the applicability of the microarray-based system, we genotyped 70 primate and 5 non-primate samples at the 68 nucleotide positions in a blind test. None of these samples were used during the design of the minisequencing primers and some samples be-

longed to species not incorporated in the alignments used to identify the diagnostic nucleotide positions. The primate test samples represented all primate infraorders, including 42 Catarrhini, 13 Platyrrhini, 14 Strepsirrhini and 1 Tarsier, and were primarily African primate species heavily traded for both medical and bushmeat purposes. We successfully genotyped 68 out of the 70 primate samples (Table 3). One of the *Ovis aries* samples was identified as non-primate; however, the other non-primate samples (another *O. aries*, a Dermoptera, a Scandentia and a Tupaia) were not positively assigned as non-primates, implying that the sequences used here are not efficient at discriminating ordinal differentiation. However, 96% of the primate samples were correctly assigned at the infraorder level, demonstrating the high reliability of the positions selected in the epsilon globin gene (Table 3). Two of the samples that were incorrectly assigned belonged to the *Cercocebus* and 1 to the *Chlorocebus* genera. We did not use sequences from these genera for identifying the diagnostic nucleotide positions; however, 3 *Chlorocebus* and 2 *Cercocebus* samples were successfully assigned to their correct infraorder (Table 3). At the Catarrhine subfamily level, 85% of the samples were assigned to the correct subfamily (Table 3). Six samples that failed to be assigned were from the Cercopithecinae subfamily (3 *Cercopithecus*, 2 *Miopithecus* and 1 *Macaca*), and were assigned as hominoids, which might indicate that these species have unknown polymorphisms at 4 of the 8 genotyped positions in this gene. Of the samples correctly assigned at the infraorder and subfamily levels, 80% were successfully genotyped at the genus level (Table 3). Most of the samples (11) not assigned to any genus were from the Cercopithecinae subfamily. These samples failed due to the too low number of nucleotide positions with successful assays for an unambiguous genus assignment, or because the determined genotypes did not

match any known genus. Of the samples genotyped at the genus level, 68% were assigned to the correct genus. Most of the falsely assigned samples were from the Strepsirrhini infraorder.

DISCUSSION

The need for accurate and reliable methods for the identification of animal species has steadily increased during the past decades and has become essential to control illegal trade of endangered species. Traditional methods based on sequencing mitochondrial sequences, such as the DNA barcoding initiative, provide efficient means for assigning unidentified individuals/samples to known species (Hebert et al. 2003, Hajibabaei et al. 2006), but are not necessarily the most efficient methods for analyzing unknown samples. In contrast, a microarray format, such as the one devised here, allows analysis of 80 samples for up to 196 SNPs simultaneously. This would be required for instance for performing a comprehensive market survey. Here we present the first outline of a microarray system that uses minisequencing primer-extension technology for diagnostic nucleotide positions and allows the identification of a relatively large number of species. Our microarray system allows identification of 45 out of the 65 primate genera currently recognized. The 20 genera that are not distinguished represent less than 14% of the total number of primate specimens traded worldwide during the last 2 decades, according to the CITES trade database.

When genotyping the novel samples for testing the applicability of the microarray, we found that surprisingly only 1 out of the 5 non-primate samples were assigned as non-primate. Based on this result, it could prove useful to include known primate-specific sequences, for example positions located within Alu elements or in primate-specific transposable element insertions (Schmitz et al. 2002).

Table 3. Correctly assigned samples. N, number of samples tested at each level; only samples correctly assigned at one level are considered for genotyping in levels shown below; n/a, not applicable

| Primate taxa | — Infraorder level — | | — Subfamily level — | | — Genus level — | |
|-----------------|----------------------|--------------------|---------------------|--------------------|-----------------|--------------------|
| | N | Correctly assigned | N | Correctly assigned | N | Correctly assigned |
| Cercopithecinae | 24 | 21 | 21 | 15 | 4 ^a | 1 |
| Colobinae | 3 | 3 | 3 | 3 | 3 | 2 |
| Hominoidea | 15 | 15 | 15 | 15 | 15 | 13 |
| Platyrrhini | 11 | 11 | | n/a | 11 | 10 |
| Strepsirrhini | 14 | 14 | | n/a | 13 ^b | 5 |
| Tarsiiformes | 1 | 1 | | n/a | 1 | 1 |
| Total | 68 | 65 (96%) | 39 | 33 (85%) | 47 | 32 (68%) |

^a11 Cercopithecinae samples assigned to the correct subfamily could not be genotyped at genus level
^b1 Strepsirrhini sample assigned to the correct infraorder could not be genotyped at genus level

The reliability of the microarray is high both at the infraorder and Catarrhini subfamily levels, with 95% and 85% correctly assigned samples, respectively. The samples incorrectly assigned belong to species from which little or no information at infraorder or Catarrhini subfamily levels was available. Adding more minisequencing primers with redundant information to the microarray-based assay would increase the reliability of the system further. It is important to note that we have identified at least 9 alternative informative positions in epsilon globin and 3 positions in ApoB that were able to differentiate between Cercopitheciinae and Hominoids.

At the genus level, our system is more accurate in hominoids (which include some important species subjected to the bushmeat trade), Colobinae, Platyrrhini and Tarsiiformes. In those groups 87, 66, 91 and 100% of the samples were correctly assigned at the genus level, respectively (Table 3). However, the microarray is less satisfactory in Cercopitheciinae and Strepsirrhini groups. For the Cercopitheciinae subfamily, 11 out of 15 samples could not be genotyped at the genus level, probably due to the low number of diagnostic nucleotide positions in 12S rRNA that were functional in the microarray system (7 out of the 18 identified). This issue could be solved by the addition of alternative positions identified in other regions of the 12S rRNA gene or in other mitochondrial genes, such as cytochrome oxidase I (coxI) or cytochrome b (cytb), which are also widely used for species identification (e.g. Pfunder et al. 2004). It is interesting to note that recently a high number of primate sequences from coxI have been generated. This information could perhaps be used in a future version of the array to improve species-level resolution. Moreover, the inclusion of diagnostic nucleotide positions would allow grouping the closest genera. Then, new specific informative positions in 12S rRNA would be used for differentiating among the genera included in the different groups. As a result, the efficiency of the microarray could be improved. This strategy would allow the design of group-specific primers to genotype the informative positions at the genus level and improve the percentage of accurately genotyped positions. A major feature of the Strepsirrhini results was that most of the samples genotyped (8 out of 13) were assigned to the incorrect genus. This is likely due to a relative paucity of published sequences from most of the genera of this infraorder and the existence of undetected polymorphisms, which seems frequent in mitochondrial regions (Belfiore et al. 2003). Clearly more data is needed to be deposited for this infraorder.

Interestingly, from the samples correctly assigned at genus level, 50% were narrowed down to between 2 and 4 potentially related genera. Most of the samples

that were unambiguously identified belonged to genera with genus-specific diagnostic nucleotide positions, which might indicate that probably the positions fixed in only one of the genera of our database are less prone to be polymorphic in other genera. By obtaining genus-specific positions for all genera, the applicability of the microarray could be further increased.

In a future improved system, it would be a major advantage to include more minisequencing primers that are informative at the species level to distinguish among primate species that are especially subjected to trade for bushmeat and/or medical purposes.

One of the main criticisms of our system, and of other initiatives such as DNA barcoding (Hebert et al. 2003), could be the reliance on a single locus to identify a genus or species a sample belongs to. However, Tavares & Baker (2008) demonstrated the reliability of a single mitochondrial gene to identify sister species in birds. It seems clear that more sequencing efforts are needed and with recent advances in sequencing technologies this could be done in the near future. One advantage of our system is that once new information is generated (new diagnostic nucleotide positions identified in other loci), this could easily be implemented in the current microarray.

In conclusion, our results show that the design of a microarray system to identify a high number of groups requires the inclusion of markers informative at different levels to allow the hierarchical differentiation of taxa. This strategy allows the design of group-specific primers for solving the problem of the presence of polymorphisms in the regions flanking the diagnostic positions. Moreover, taxon-specific diagnostic positions seem to be very efficient at unambiguously identifying multiple species and the inclusion of redundant positions is necessary for increasing the reliability of the system. The first generation microarray system developed here is unsuitable for forensic applications; however, once the suggested improvements are implemented, this microarray may be promising for efficient and sensitive species identification of primate samples, and could become a valuable tool in wildlife management and law enforcement. This technology can be extended to other species, for example a 'mammal chip' that unambiguously identifies all mammal species or a 'biodiversity chip' containing minisequencing primers to distinguish key species from all live phyla, as suggested by Pfunder et al. (2004), could be developed in the future.

Acknowledgements. This study was funded by the European Commission (INPRIMAT Consortium, contract QLRI-CT-2002-01325), the DURSI of the Generalitat de Catalunya (to O.A. with ref. 2002FI-00280) and the Swedish Research Council for Science and Technology (to A-C.S). We are grate-

ful for the help and support of H. Zischler, B. Crouau-Roy, G. Doxiadis, M. Rocchi, L. Rönn, M. Bonhomme, A. Roeder, M. Ventura and K. Jeffery, all members of the INPRIMAT Consortium. We thank H. Zischler and M. Rocchi for providing samples. This article is dedicated to the memory of Dr. Xavier Domingo-Roura.

LITERATURE CITED

- Balitzki-Korte B, Anslinger K, Bartsch C, Rolf B (2005) Species identification by means of pyrosequencing the mitochondrial 12s rRNA gene. *Int J Legal Med* 119:291–294
- Belfiore NM, Hoffman FG, Baker RJ, Dewoody JA (2003) The use of nuclear and mitochondrial single nucleotide polymorphisms to identify cryptic species. *Mol Ecol* 12: 2011–2017
- Bosch M, Andrés O, Domingo-Roura X (2006) Single-nucleotide polymorphisms in the epsilon-globin gene for differentiating primate infraorders. *Folia Primatol (Basel)* 77:387–392
- Dalebout ML, Baker CS, Mead JG, Cockcroft VG, Yamada TK (2004) A comprehensive and validated molecular taxonomy of beaked whales, family Ziphiidae. *J Hered* 95: 459–473
- Gonzalez SF, Krug MJ, Nielsen ME, Santos Y, Call DR (2004) Simultaneous detection of marine fish pathogens by using multiplex PCR and a DNA microarray. *J Clin Microbiol* 42:1414–1419
- Groves CP (2001). *Primate taxonomy*. Smithsonian Institution Press, Washington, DC
- Hajibabaei M, Singer GA, Hickey DA (2006) Benchmarking DNA barcodes: an assessment using available primate sequences. *Genome* 49:851–854
- Hebert PD, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. *Proc R Soc B Biol* 270: 313–321
- Hsieh HM, Chiang HL, Tsai LC, Lai SY, Huang NE, Linacre A, Lee JC (2001) Cytochrome b gene for species identification of the conservation animals. *Forensic Sci Int* 122: 7–18
- Kocher TD, Thomas WK, Meyer A, Edwards SV, Pääbo S, Villablanca FX, Wilson AC (1989) Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc Natl Acad Sci USA* 86: 6196–6200
- Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, Syvänen AC (2002) Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res* 30:70e
- Mittermeier RA, Valladares-Pádua C, Rylands AB and others (2005) Primates in peril: the world's most endangered primates 2004–2006. IUCN/SSC Primate Specialist Group (PSG), the International Primatology Society (IPS) and Conservation International (CI), Washington, DC
- Nekrutenko A, Makova KD, Baker RJ (2000) Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. *Gene* 249:47–51
- Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, Syvänen A-C (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res* 10:1031–1042
- Pfeiffer I, Burger J, Brenig B (2004) Diagnostic polymorphisms in the mitochondrial cytochrome b gene allow discrimination between cattle, sheep, goat, roe buck and deer by PCR-RFLP. *BMC Genet* 5:30
- Pfunder M, Holzgang O, Frey JE (2004) Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome b as genetic markers. *Mol Ecol* 13:1277–1286
- Prakash S, Patole MS, Ghumatkar SV, Nandode SV, Yogesh SK, Schouche YS (2000) Mitochondrial 12s rRNA sequence analysis in wildlife forensics. *Curr Biol* 78:1239–1241
- Rowe N (1996). *The pictorial guide to the living primates*. Pogonias Press, East Hampton, NY
- Schmitz J, Ohme M, Suryobroto B, Zischler H (2002) The colugo (*Cynocephalus variegatus*, dermoptera): The primates' gliding sister? *Mol Biol Evol* 19:2308–2312
- Springer MS, DeBry RW, Douady C, Amrine HM, Madsen O, de Jong WW, Stanhope MJ (2001) Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Mol Biol Evol* 18:132–143
- Syvänen AC (2005) Toward genome-wide SNP genotyping. *Nat Genet* 37(Suppl):S5–S10
- Tavares ES, Baker AJ (2008) Single mitochondrial gene barcodes reliably identify sister species in diverse clades of birds. *BMC Evol Biol* 8:81
- Teletchea F, Bernillon J, Duffraisse M, Laudet V, Hänni C (2008) Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples. *J Appl Ecol* 45:967–975
- Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Walsh PD, Abernethy KA, Bermejo M and others (2003) Catastrophic ape decline in western equatorial africa. *Nature* 422:611–614
- Wild C, Morgan BJ, Dixon A (2005) Conservation of drill populations in Bakossiland, Cameroon: historical trends and current status. *Int J Primatol* 26:759–773
- Wolf C, Rentsch J, Hubner P (1999) PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification. *J Agric Food Chem* 47:1350–1355

Editorial responsibility: Anna Nekaris,
Oxford, UK

Submitted: October 28, 2008; Accepted: February 28, 2009
Proofs received from author(s): May 9, 2009