4-1-2007

A SINE-based dichotomous key for primate identification

Scott W. Herke
Center for BioModular Multi-Scale Systems

Jinchuan Xing
Center for BioModular Multi-Scale Systems

David A. Ray
West Virginia University

Jacquelyn W. Zimmerman
Center for BioModular Multi-Scale Systems

Richard Cordaux
Center for BioModular Multi-Scale Systems

See next page for additional authors

Recommended Citation

Herke, S., Xing, J., Ray, D., Zimmerman, J., Cordaux, R., & Batzer, M. (2007). A SINE-based dichotomous key for primate identification. *Gene, 390* (1-2), 39-51. https://doi.org/10.1016/j.gene.2006.08.015

Available online at www.sciencedirect.com
ScienceDirect
Gene 390 (2007) 39–51
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

A SINE-based dichotomous key for primate identification

Scott W. Herke [a], Jinchuan Xing [a], David A. Ray [b], Jacquelyn W. Zimmerman [a], Richard Cordaux [a], Mark A. Batzer [a,*]

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Microsystems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, United States*
[b] *Department of Biology, 53 Campus Drive, West Virginia University, Morgantown, WV 26506, USA*

Received 28 June 2006; received in revised form 1 August 2006; accepted 2 August 2006
Available online 30 August 2006

Abstract

For DNA samples or 'divorced' tissues, identifying the organism from which they were taken generally requires some type of analytical method. The ideal approach would be robust even in the hands of a novice, requiring minimal equipment, time, and effort. Genotyping SINEs (Short INterspersed Elements) is such an approach as it requires only PCR-related equipment, and the analysis consists solely of interpreting fragment sizes in agarose gels. Modern primate genomes are known to contain lineage-specific insertions of *Alu* elements (a primate-specific SINE); thus, to demonstrate the utility of this approach, we used members of the *Alu* family to identify DNA samples from evolutionarily divergent primate species. For each node of a combined phylogenetic tree (56 species; $n=8$ [Hominids]; 11 [New World monkeys]; 21 [Old World monkeys]; 2 [Tarsiformes]; and, 14 [Strepsirrhines]), we tested loci (>400 in total) from prior phylogenetic studies as well as newly identified elements for their ability to amplify in all 56 species. Ultimately, 195 loci were selected for inclusion in this *Alu*-based key for primate identification. This dichotomous SINE-based key is best used through hierarchical amplification, with the starting point determined by the level of initial uncertainty regarding sample origin. With newly emerging genome databases, finding informative retrotransposon insertions is becoming much more rapid; thus, the general principle of using SINEs to identify organisms is broadly applicable.

Keywords: *Alu*; Mobile elements; Retrotransposons

1. Introduction

Unless definitive proof already exists, identifying the organism from which DNA samples, and often even tissues, were taken is a crucial step during biological and medical studies. For instance, researchers may receive DNA from sources that lack proof of its origin or samples taken in the field from 'divorced' tissues such as hair or feces (Kohn et al., 1995; Taberlet et al., 1997; Lathuilliere et al., 2001; Matsubara et al., 2005). Further, cross-contamination can occur within cell cultures grown in the presence of lines from multiple species and within concurrent large-scale DNA extractions from multiple species. Even previously documented samples may lose their labels or be cross-contaminated. In addition, investigations of the illegal wildlife trade require methods for documenting the identity of seized products (Hsieh et al., 2003; Malisa et al., 2005; Yan et al., 2005; Domingo-Roura et al., 2006). Unless the identities of the sample organisms are properly established and cross-contamination issues are detected, future analyses based on these types of samples can be seriously compromised.

In many cases, molecular techniques are required for a positive identification of undocumented samples. The ideal approach would require minimal equipment, time, and effort; further, it would be robust even in the hands of a novice. Genotyping mitochondrial DNA (mtDNA) is one of the

Abbreviations: numtDNA, nuclear pseudogene of mitochondrial origin; PCR, polymerase chain reaction; SINE, Short INterspersed Element; TE, tris-EDTA; H, Hominid; OWM, Old World monkey; NWM, New World monkey; IPBIR, Integrated Primate Biomaterials and Information Resource; IUPAC, International Union of Pure and Applied Chemistry; NISC, NIH Intramural Sequencing Center.

* Corresponding author. Tel.: +1 225 578 7102; fax: +1 225 578 7113.
E-mail address: mbatzer@lsu.edu (M.A. Batzer).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.08.015

commonly used procedures (Kohn et al., 1995; Palumbi and Cipriano, 1998; Malisa et al., 2005), and DNA barcoding based on mtDNA sequences has been advocated for the identification of species (Hebert et al., 2004). Even the Integrated Primate Biomaterials and Information Resource (www.IPBIR.org) uses mtDNA (mitochondrial cytochrome *c* oxidase subunit I) to verify the identity of samples and is generating a database of 'DNA barcodes' for primates (Lorenz et al., 2005). However, the use of mtDNA is complicated by taxon-specific patterns of 'universal primer' failure, as well as by the amplification or co-amplification of nuclear pseudogenes (numtDNA) of mitochondrial origins (Moritz and Cicero, 2004; Lorenz et al., 2005).

Recently, retrotransposons were used to develop characteristic 'fingerprints' for 10 individual human cell lines to ensure that each line used in an experiment was the expected cell line (Ustyugova et al., 2005). Thus, genotyping retrotransposons (*i.e.*, mobile elements) offers another approach for ensuring accurate identification of DNA and tissue samples with respect to their organismal origins. SINEs (Short INterspersed Elements) are a class of non-autonomous mobile elements that are <500 bp in length and have no open reading frames. Given that they are ubiquitous in mammalian genomes as well as in many non-mammalian genomes (Deininger and Batzer, 2002; Okada et al., 2004; Ray et al., in press), SINEs are well known as powerful tools for systematic biologists (Hamdi et al., 1999; Hillis, 1999; Shedlock and Okada, 2000; Walker et al., 2004a) and forensic scientists (Walker et al., 2003a,b; Walker et al., 2004b; Ray et al., 2005a; Walker et al., 2005). For example, Waddell et al. (2001) stated that, while DNA sequence analysis can help set up hypothesized clades of placental genomes, SINEs are essential to testing them. In that context, the close relationship of cetaceans (whales, dolphins, and porpoises) to Hippopotamidae was established through an analysis of SINE data (Nikaido et al., 1999). In a different vein, Walker et al. (2003a) demonstrated that SINEs could be used to distinguish bovine, porcine, chicken and ruminant species from as little as 0.1 pg of starting DNA, offering a more sensitive method for detecting products relevant to bovine spongiform encephalopathy. Finally, SINEs have been used for several phylogenetic studies of Mammals, including the Primate order (Schmitz et al., 2001; Salem et al., 2003b; Roos et al., 2004; Ray et al., 2005b; Schmitz et al., 2005; Xing et al., 2005; Kriegs et al., 2006; Nikaido et al., 2006).

Alu elements are primate-specific SINEs, with a full length of ~300 bp. As the most successful SINEs in primate genomes, these mobile elements have proliferated during the primate radiation to more than one million copies (~10% by mass) in the human genome (Lander et al., 2001; Batzer and Deininger, 2002). The study of *Alu* elements has produced definitive pictures regarding the phylogenies of hominids (Salem et al., 2003b), Old World monkeys (Xing et al., 2005), and New World monkeys (Ray et al., 2005b). Even the hypothesized close relationship between tarsiers and anthropoids received support from *Alu* elements (Schmitz et al., 2001; Roos et al., 2004). Thus, to demonstrate the utility of a dichotomous SINE-based key, we have focused on the use of *Alu* elements for the identification of primate DNA samples.

The diverse order of primates is currently divided into two major groups, Strepsirrhini and Haplorhini (see phylogenetic review by Disotell, 2003). The Strepsirrhini include the Lorisiformes (~8 genera) and the Lemuriformes (~14 genera). The Haplorhini include the Tarsiformes (1 genus) and the Anthropoidea, which is further subdivided into the Platyrrhini (New World monkeys [NWM], ~16 genera) and the Catarrhini. Cercopithecidae (Old World monkeys [OWM], ~21 genera) and Hominoidea (Hominids [H], ~5 genera) comprise the Catarrhini. For a combined phylogenetic tree (56 species; $n=8$ [H]; 11 [NWM]; 21 [OWM]; 2 [Tarsiformes]; and, 14 [Strepsirrhines]), we tested 443 loci for amplification in all five currently existing primate groups: (A) 362 anthropoid loci which came from four previous phylogenetic studies (Salem et al., 2003b; Han et al., 2005; Ray et al., 2005b; Xing et al., 2005), including 25 existing OWM loci for which we redesigned primers to improve amplification efficiency; and, (B) 81 new loci that we characterized to provide resolution at the deepest nodes in the combined phylogenetic tree (24 loci), to identify members of Strepsirrhines (31 loci), or to define several terminal branches in the tree (26 loci).

The benefits of a SINE-based dichotomous key are amply demonstrated by this work, which is based exclusively on *Alu* elements. For instance, the 195 loci contained in this dataset can be used to specifically identify 28 terminal branches of our combined primate phylogenetic tree; further, samples derived from the remaining 28 species (and even species not included in the analysis) can at least be partially identified by the internal node loci. As such, the key is a valuable tool for clarifying the identity of primate samples which lack documented provenance, to check for cross-contamination of primate DNA samples, and to identify primate products seized from the illegal wildlife trade.

2. Materials and methods

2.1. DNA samples

Sufficient amounts of genomic DNA were available for only nine species; very limited amounts of genomic DNA were available for the other 47 species (Appendix A). Thus, those samples of limited quantity were first subjected to whole genome pre-amplification using the Genomiphi genome amplification kit (Amersham, Sunnyvale, CA) following the manufacturer's instructions. The genome-amplified DNA was ethanol precipitated, dried, and resuspended in 1× TE. Based on A_{260} readings, all samples were diluted in 1× TE to ~20 ng/μl, and then used as templates for PCR assays.

2.2. Primer design

The vast majority of the primers tested here were designed previously (Salem et al., 2003b; Han et al., 2005; Ray et al., 2005b; Xing et al., 2005). However, additional primers were designed for: (1) new loci in *Hylobates*, *Pongo*, *Gorilla*, and *Pan*; (2) some existing OWM loci; and, (3) new loci that would group various combinations of anthropoid primates or members of Strepsirrhini. Loci specific for *Hylobates* and

Pongo were characterized as previously described (Xing et al., 2005), and flanking oligonucleotide primers for PCR amplification of each *Alu* element were designed in Primer3 (Rozen and Skaletsky, 2000). From the phylogenetic studies, unless a locus had been amplified from more than ~75% of the species in the original analyses, we did not include it in the current analysis (except for terminal branch loci). Nevertheless, PCR amplifications frequently failed when H and OWM primers were used on species not included in the original analyses; thus, we redesigned primers whenever possible based on multiple species sequence alignments. To this end, we first linked (by 20 N's) each pair of forward and reverse (reverse-complemented) primers and used them as queries in BLAST (Basic Local Alignment Search Tool, both nr and htgs databases, available at http://www.ensembl.org/multi/blastview; Altschul et al., 1990). Although only human or chimpanzee sequences were returned for the H primers, useful data were generated for about 25 OWM loci. For each locus, we generated a consensus sequence in BioEdit (ver 7.0.1; Hall, 1999) from the manually adjusted ClustalW alignment (Thompson et al., 1994) of the returned sequences. In regions deemed potentially suitable for primers, clear zones for primer design were created in each consensus sequence by manually converting IUPAC codes to the most appropriate specific bases. All consensus sequences were then processed in Primer3, which generated potential primers only in relatively conserved regions because the program treats IUPAC codes as excluded zones for primer design.

Additional new primer pairs were designed based on genomic scaffolds for 12 primates: *Cercopithecus* (*Chlorocebus*) *aethiops* (DP000029.1); *Macaca mulatta* (DP000005.1); *Papio anubis* (AE017182.1); *Callicebus moloch* (DP000019.1); *Callithrix jacchus* (DP000014.1); *Eulemur macaco macaco* (DP000024.1); *Lemur catta* (DP000004.1); *Microcebus murinus* (DP000022.1); *Otolemur garnettii* (DP000013.1); *Gorilla gorilla* (DP000025.1); *Pan troglodytes* (DP000016.1); and, *Pongo pygmaeus* (DP000026.1). All sequences were generated by the NISC Comparative Sequencing Program, and they are orthologous to an ~1.9-Mb region ('Target 1') on human chromosome 7q31.3 (http://www.nisc.nih.gov). We used MultiPipMaker (http://bio.cse.psu.edu/pipmaker; Schwartz et al., 2000) to generate several alignments of the complete sequences from all 12 non-human primates and a portion of the human chromosome 7 genomic contig (NT_007933.14|Hs7_8090:40831359–42536062); the reference RepeatMasker (http://repeatmasker.genome.washington.edu/) file in each case was derived from the species of interest in that particular alignment. Parameters were set as 'search one strand' and 'chaining', with default sensitivity and time limit settings. The MultiPipMaker 'Verbose text' alignment files were processed in MEGA3 (Kumar et al., 2004) to generate fasta files, which were imported into BioEdit. We then searched the MultiPipMaker PDF output file for *Alu* insertions that were specific to the species or group of species of interest. After locating a candidate *Alu* element, we exported ~1500 bp from each sequence in the BioEdit file (centered on the *Alu* element), excised all gaps, and checked the sequences with RepeatMasker for repeat elements. Primers for suitable loci were designed as described above.

2.3. PCR analyses

Initial tests of the existing primers showed that PCR done with relatively low DNA concentrations (*e.g.*, 25 ng) and a standard annealing time of 30 s with 32 cycles often failed to generate either empty or filled site fragments for many species (data not shown). The final parameters chosen as optimal for amplifying products from across all 56 species were: 80 ng DNA (4 μl in 1× Tris-EDTA) in 25 μl reactions with an initial denaturation at 94 °C (2 min); 37 cycles of denaturation at 94 °C (30 s), optimal annealing temperature (60 s), and extension at 72 °C (60 s); and, a final extension at 72 °C (5 min). PCR reagents included 200 nM of each oligonucleotide primer, 200 μM each dNTP in 50 mM KCl, 1.8 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.4), 2.0 units Taq DNA polymerase, and 10% RediLoad. Resulting PCR products (~25 μl) were separated for 25 min (325 V) on 2% Na-borate agarose gels (15×25 cm; 300 ml) containing ethidium bromide for visualization by UV fluorescence (Brody and Kern, 2004a,b). Optimal annealing temperatures were initially derived from temperature gradient tests done with two to four divergent species; any further refinements were based on initial primate panel results. Most of the loci developed from the Target 1 scaffold alignments were tested on all non-anthropoid species because the loci provided resolution at the deepest nodes in the combined phylogenetic tree or identified members of Strepsirrhini. In contrast, all other primers were unlikely to efficiently amplify Strepsirrhini loci (due to their high genetic divergence from anthropoid primates), so we tested them only on *L. catta*, *Nycticebus pygmaeus*, *Galago senegalensis* and *Tarsius syrichta*.

2.4. DNA sequencing

Selected PCR products were purified using the Wizard PCR gel purification kit (Promega, Madison, WI) and directly sequenced using chain termination sequencing (Sanger et al., 1977) on an ABI 3130XL Genetic Analyzer. The majority of the *Alu* elements had already been verified by sequence analysis in the prior phylogenetic studies and the newly developed loci were designed from genomic scaffold data. Thus, we sequenced only those few loci that demonstrated amplification patterns consistent with the independent insertion of *Alu* elements (GenBank accession numbers DQ822046–DQ822070 and DQ843660–DQ843663).

3. Results

3.1. Dichotomous key branch loci

For each branch with informative loci, data on primer sequences, expected fragment sizes, annealing temperatures, and branches with atypical PCR results can be found in Appendix B. An ideal primer pair amplified the locus in most species such that a clean, distinct band of the appropriate size for an 'empty' or a 'filled' site was visible in a 2% agarose gel. While the use of Na-borate for electrophoresis allowed for faster and cleaner resolution of the PCR products, the system is completely compatible with

TBE (Tris-borate-EDTA, which is more commonly used in most laboratories). With respect to loci from which primers generated substantial amounts of non-specific products, apparent independent insertions, or numerous species dropouts, we generally discarded them if changing the annealing temperature or redesigning the primers did not improve the result. With rare exception, only the primers based on Target 1 scaffold sequences gave reliable results within Tarsiformes and Strepsirrhini.

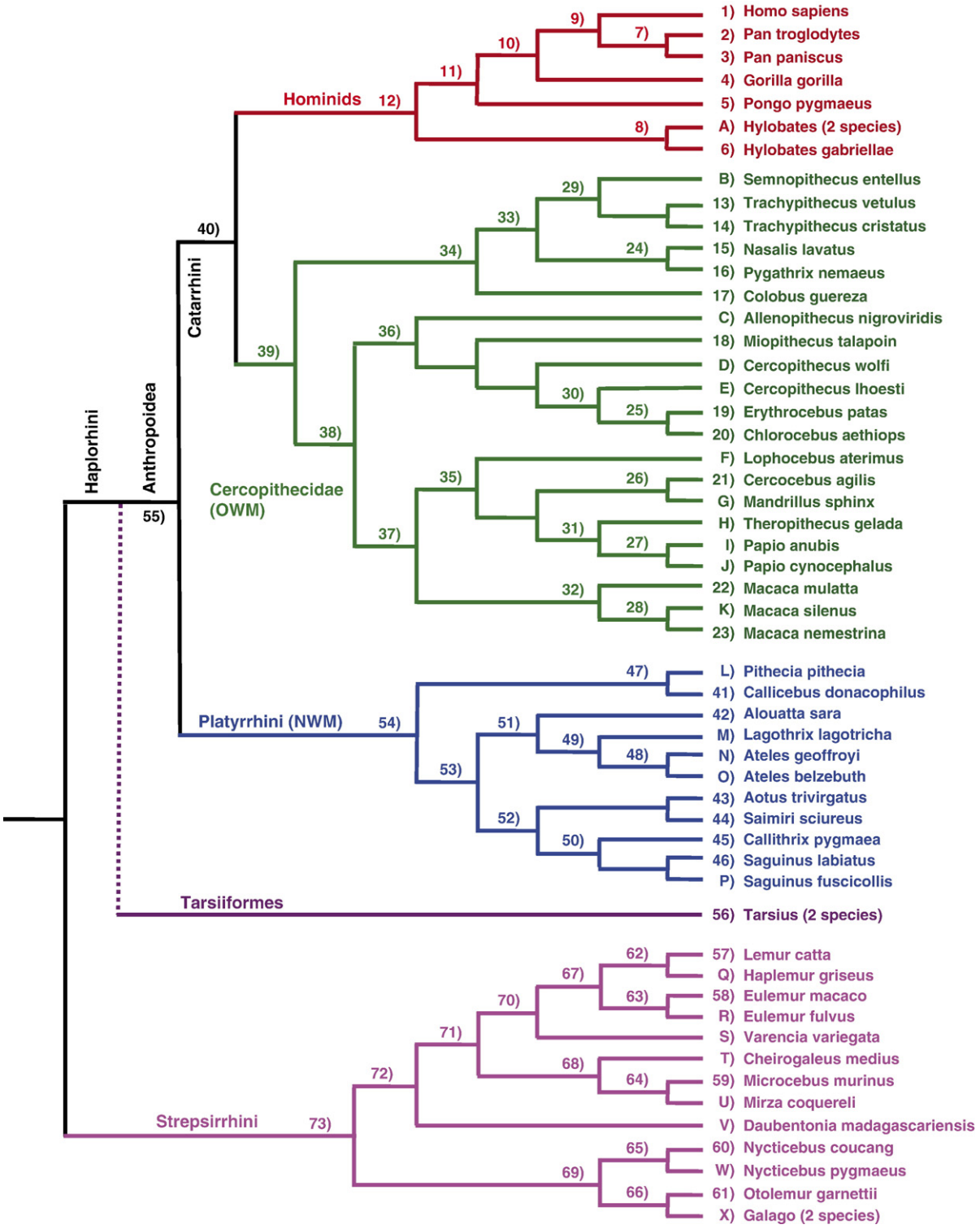

Fig. 1. Combined phylogenetic tree. Numbers indicate branches that are characterized by various *Alu* insertions; letters are used for terminal branches that are characterized only by internal branch loci. See Appendix B for information on primers that amplify branch-specific loci; typically, at least three loci are listed for each branch. Figs. 2 and 3 depict an example of an identification analysis involving an anthropoid primate DNA sample.

3.2. Sequence analysis of problematic loci

Sequence analysis revealed that the 'filled' site alleles shared by *Tarsius* and most lemuriformes at locus 'Str-71_C' were independent insertion events (*i.e.* different *Alu* elements found within the same small genomic pre-integration site). Further, although the *Nycticebus* PCR products for locus 'Str-66_A' were initially interpreted as being a small deletion in *N. pygmaeus* and a small insertion in *N. coucang* (relative to *Otolemur* and the two *Galago* spp), neither *Nycticebus* species actually contained the *Alu* element (~270 bp) found in our analysis of the *Otolemur* Target 1 alignment. Instead, there was a truncated independent insertion (~50 bp) in both *N. pygmaeus* and *N. coucang* and an additional independent insertion (~270 bp) in *N. coucang*. For the 27 Strepsirrhini loci examined, no other cases of independent insertion were detected. For the 11 'anthropoid' loci in which the 'filled' site appeared to be present in species outside the expected clade, sequence analysis showed that those loci were subject to independent insertions or paralogous amplification (two cases).

3.3. Phylogenetic context

Phylogenetic results from this analysis (Fig. 1) were entirely consistent with those described previously (Disotell, 2003; Salem et al., 2003b; Ray et al., 2005b; Xing et al., 2005). Given that most of the H, OWM, and NWM loci were already part of those phylogenetic studies, this concurrence was expected. By contrast, the loci designed for distinguishing Strepsirrhini were new and represent additional support for that phylogeny. However, we were unable to identify any *Alu* element that linked *Tarsius* to any other group, despite having characterized 16 anthropoid-specific loci and 4 Strepsirrhini-specific loci. Even so, a combination of loci (showing the pre-integration site allele) and one independent insertion (showing the same fragment size as found for the filled allele of lemuriformes) suffices to generate a tentative identification of *Tarsius*. For studies involving the Tarsiformes, researchers may wish to refer to Schmitz et al. (2001) for three additional loci which can be used to identify *Tarsius*, but are not restricted to that genus.

4. Discussion

Prior to any project, it is crucial that a researcher be certain of the identity of the organism from which the samples were taken; unless there is definitive proof of identity (*e.g.*, documented provenance, geographic origin, or morphological characteristics), molecular methods are required. This issue is becoming ever more prominent as researchers turn to non-invasive methods (*e.g.*, hair or fecal samples) for obtaining molecular data (Kohn et al., 1995; Taberlet et al., 1997; Lathuilliere et al., 2001; Matsubara et al., 2005). Further, the conclusions of any project can be negatively affected if the samples used have been contaminated or inadvertently switched with DNA from other species. Finally, the identity of products seized from the illegal wildlife trade may be ambiguous without resorting to molecular markers (Hsieh et al., 2003; Yan et al., 2005; Domingo-Roura et al., 2006). The ability to determine the authentic identity of those products, even if only to the genus level or higher, is crucial to criminal investigations or to determining the likelihood of the products containing diseases representing a threat to human populations (for a discussion of disease issues with respect to primates, see Wolfe et al., 2005). The SINEs (*i.e.*, *Alu* elements) comprising this demonstration dataset are valuable tools in these respects, and even the 'limitation' that *Alu* elements are primate-specific could be beneficial in some cases. For example, the detection of *Alu* elements in fecal samples from non-primate carnivores would identify particular primates as part of that animal's diet.

4.1. Rationale for a SINE-based dichotomous key

Several factors make a SINE-based dichotomous key valuable for identifying samples. First, genotyping SINEs requires only a PCR machine, agarose gel electrophoresis equipment, and a means of gel documentation. Second, the technique is robust even in the hands of a novice. Third, analyses are simplified by the fact that the data consist solely of the presence of either a 'filled' site (*i.e.*, SINE present) or an 'empty' site (*i.e.*, SINE absent) fragment for each locus. As a result, detection of these mobile elements at a series of loci can be used in a hierarchical fashion to rapidly narrow the possibilities regarding the identity of anonymous primate DNA samples (Fig. 2). Fourth, 'universal' primers are used in a simple PCR assay to amplify loci from most relevant taxa. By contrast, the mtDNA barcoding approach of IPBIR requires taxon-specific primers and procedures for eliminating the amplification of numtDNA (Lorenz et al., 2005). Finally, unlike 'single locus' (*e.g.*, mainly mtDNA) systems, a SINE-based dichotomous key identifies organisms on the basis of multiple loci and there is no *a priori* reason to expect such loci to be evolutionarily linked.

SINEs are also useful characters for identifying organisms because their precise removal is extremely rare (Van de Lagemaat et al., 2005), and larger deletions that also remove a particular SINE are very unlikely to happen in multiple genomes. Thus, mobile elements are generally considered to be homoplasy-free characters (Perna et al., 1992; Hamdi et al., 1999; Batzer and Deininger, 2002; Roy-Engel et al., 2002; Salem et al., 2003b; Shedlock et al., 2004; Salem et al., 2005; Ray et al., in press). In other words, SINEs detected at a particular locus, unlike the more uncertain status of DNA sequence data, are known to be identical by descent rather than just identical by state. Further, the mode of SINE evolution is essentially unidirectional, making absence of the SINE the ancestral state for any genomic locus under consideration. Occasionally, independent insertions of SINEs in similar genomic locations will cause evolutionarily distant species to generate nearly identical DNA fragments following gel electrophoresis. DNA sequence analysis ultimately resolves these types of ambiguous loci (Salem et al., 2003a,b; Ray et al., 2005b; Xing et al., 2005; Ray et al., in press). When developing a dichotomous key, the simplest solution is to exclude ambiguous loci from consideration. When such loci are essential, additional loci from deeper within the tree can be used to identify the broader clade to which the organism belongs, thereby restricting the identification to one of the two branches characterized by the insertion.

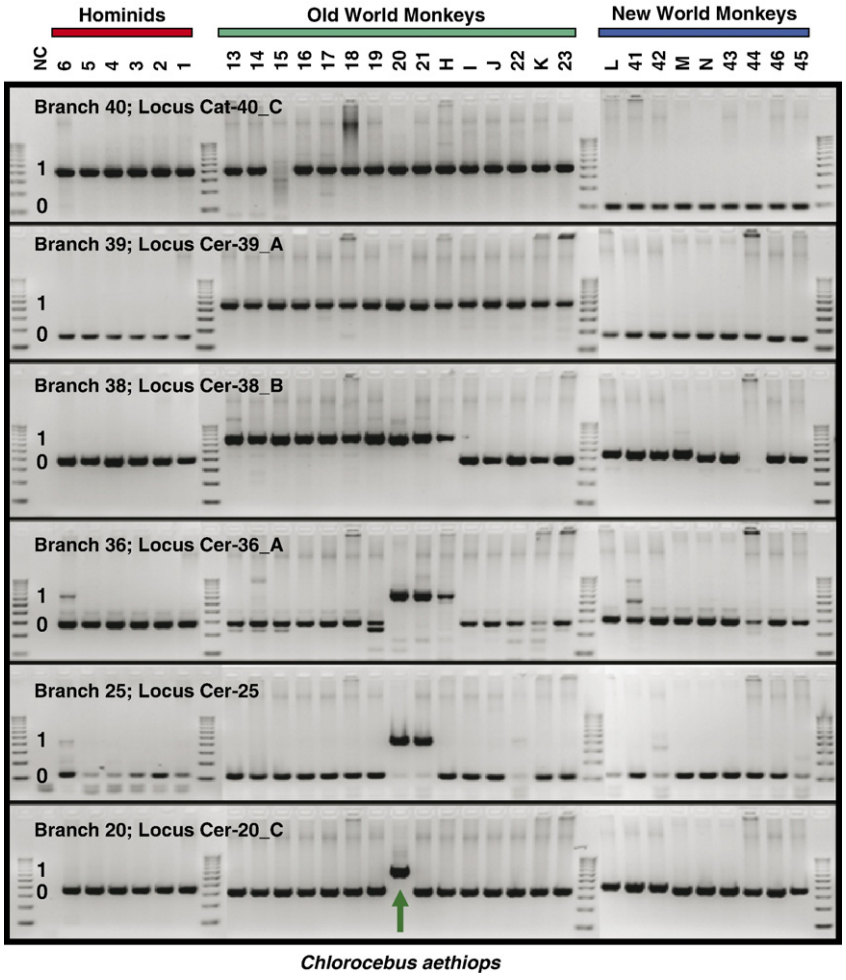

Fig. 2. Agarose gel (2%) chromatographs demonstrating the power of the SINE-based dichotomous key approach. All 30 species, including the 'target' DNA (*Chlorocebus aethiops*, Green monkey), were known to be anthropoid primates; thus, loci were successively amplified from the following branches: 40 vs. 54; 12 vs. 39; 34 vs. 38; 36 vs. 37; 18 vs. 25; and, 19 vs. 20. Results are shown only for six loci that harbored *Alu* insertions in the 'target' DNA, and demonstrate the range of band intensities and data quality of all loci included in the overall key. Species designations (below colored bars) are keyed to Fig. 1. *Alu* present ('filled' site), 1; *Alu* absent ('empty' site), 0. Molecular ladder ranges from 100 bp to 1000 bp, by 100 bp increments. NC, negative water control.

4.2. Development considerations

The development of a SINE-based dichotomous key for identifying organisms must begin with the discovery of loci which will characterize branches in the reference phylogenetic tree. The process can be simplified by restricting primer testing to those that are either (1) designed from multiple species alignments or (2) taken from projects in which the breadth of species included in those projects inherently screened out primers with more restricted specificity. As occurred in this study for 10 terminal branches, primers designed from genome databases can generate species-specific loci; however, some loci may be shared by species related to the sequenced species that were not included in the studied taxonomic panel. Further, in some cases, loci developed from genomic databases will not be useful because the SINE elements exist at low polymorphism levels within the species or are even restricted to the sequenced individual; about 5% of the new loci developed for this analysis had to be discarded for these reasons. Finally, when sequence data are not available, species-specific loci must be generated through techniques such as PCR-display (Roy et al., 1999; Sheen et al., 2000; Badge et al., 2003; Biedler et al., 2003; Okada et al., 2004; Ray et al., 2005b; Xing et al., 2005).

With the possible exception of terminal branches, each node of a robust SINE-based dichotomous key will have two primer sets, each one amplifying different SINE elements and identifying one of the bifurcations. Ideally, the key will also consist almost exclusively of loci that produce clearly empty alleles from organisms which are not supposed to harbor the SINE insert; as a result, if the analysis is started in the wrong part of the tree, the process will self-correct because all primer pairs will generate only 'empty' site fragments (Fig. 3). This allows the key to be used at multiple levels, depending on the researcher's needs. If absolutely nothing is certain about the sample's identity, the analysis should begin at the deepest node of the tree (*i.e.*, with loci from Branches 55 and 73) and work toward the terminal branches. At each bifurcation in the tree, *Alu* elements in the sample DNA will be detected within loci from one of the two branches, directing the analysis to the next bifurcation along that branch (Figs. 1–3). On the other hand,

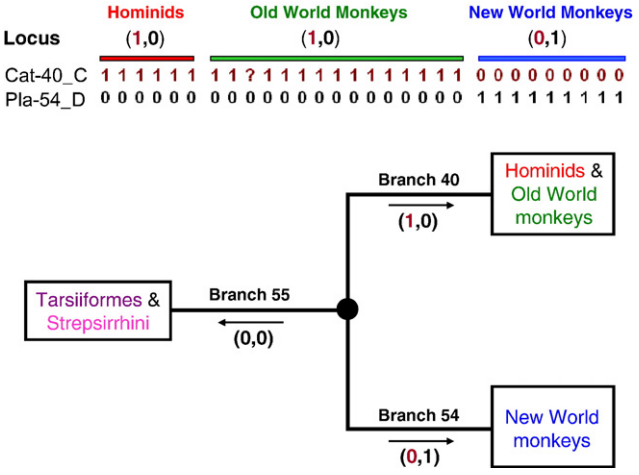

Fig. 3. Example of data and decision key. At each bifurcation in the tree, one set of primers detects *Alu* insertions in species downstream from the top branch and a second set of primers detects different *Alu* insertions in species associated with the bottom branch. Thus, if the analysis begins at an internal branch (see Fig. 2) with one locus from the top branch (*e.g.*, Cat-40_C) and one locus from the bottom branch (*e.g.*, Pla-54-D), the result can be: 1,0 (analysis proceeds down Branch 40); 0,1 (analysis proceeds down Branch 54); or 0,0 (which indicates the analysis began in the wrong part of the tree). *Alu* present, 1; *Alu* absent, 0; No amplification or multiple, weakly amplifying fragments, ?.

validated information about the sample beyond the scope of the key can justify bypassing the deepest nodes. At the most extreme, for cell cultures involving a set number of species, using only loci from close to the terminal branches is entirely appropriate. Further, to some extent, the identity of a primate can be ascertained by knowing the world region from which it came. For instance, lemuriformes are restricted to Madagascar; the NWM genera are found only in Central and South America; and (with the exception of the world-wide distribution of *Homo sapiens*) all other primates share varying degrees of overlap in Africa and Asia (Disotell, 2003). Thus, for a primate sample taken from Central or South America, it would be reasonable to begin the analysis within the Platyrrhini clade (*i.e.*, with loci from Branches 47 and 53).

4.3. Analysis considerations

When analyzing data in the context of a dichotomous SINE-based key, it is critical that the evaluator keeps several issues in mind regarding the key's characteristics (for a general discussion of interpreting SINE data, see Shedlock et al., 2004). (1) If the key is based on an organismally-restricted SINE family, using it to analyze DNA samples from organisms outside of that organismal group will be uninformative at best. (2) In most cases, the dichotomous key is likely to have been based on a single specimen for each terminal branch (as occurred in this demonstration project); therefore, terminal branch loci might be polymorphic with respect to presence or absence of the SINE element. If true, then some individuals belonging to that terminal branch will show an empty site rather than a filled site for that locus. This means that, while data for deeper nodes of the tree will apply to all relevant organisms, amplification of a single allele (either null or filled) at a terminal branch is not definitive. Nevertheless, if multiple loci are examined (from the terminal branch and preceding branches), the overall pattern of results will be definitive. (3) Closely related (but not analyzed during the key's construction) species may or may not have the SINE element. As such, for a truly anonymous DNA sample, any match to a terminal branch can only be considered as being consistent with that identity rather than as being a positive identification. Of course, the identification might be clarified by other factors which further narrow the possibilities (*e.g.*, the sample's geographic origin). (4) For any single locus, an individual specimen can generate an aberrant and misleading PCR product. (5) Some SINE elements are truncated, leading to less than the typical size difference (for that family of SINEs) between the empty and filled site fragments. Further, predicted fragment sizes may not be realized in some species (or individuals) because of deletions or insertions nearby the SINE element (but within the primer sites) or even within the SINE element itself.

4.4. Recommendations for use of the dichotomous key

To guard against the above concerns, we strongly recommend that multiple loci be used in any identification. For this dichotomous *Alu*-based key, we have typically provided at least six loci (three for each branch) at each node to facilitate this process. In addition, running known controls (for the empty and filled site fragment sizes) concurrently with anonymous DNA samples will facilitate band sizing. If certain loci will be used routinely, it may be useful to optimize primer designs or PCR protocols to generate cleaner bands and to eliminate PCR dropouts. Typical optimization procedures may include designing new primers based on current sequence databases, adjustments to annealing temperatures or times, and reductions in the number of PCR cycles. These types of adjustments were used to create the current dataset; however, the goal was to generate useful PCR products from as many species as possible, which means that the current conditions are sometimes suboptimal for particular clades within the phylogenetic tree. Thus, when electrophoresing the entire PCR product, band intensities ranged from very weak to extremely strong (sometimes also displaying extraneous, weakly amplifying PCR products). Details on these issues and additional potentially useful loci can be found in Supplemental Table 1 on our website (http://batzerlab.lsu.edu) under publications. Finally, given that our dataset included 56 species, it was impractical to determine locus-specific population insertion polymorphism levels for each species; however, such testing (at least with terminal branch loci) would be immensely beneficial for anyone working with a smaller subset of species.

5. Conclusion

Because *Alu* elements are restricted to the primate lineage, this particular dataset can only be applied to primate DNA. Nevertheless, the general principle of using SINEs and other transposable elements to identify organisms is broadly applicable, and the newly emerging genome databases will facilitate finding informative loci based on other elements. For example,

the genomic scaffolds for Target 1 are already available for at least 34 species (ranging from fish and amphibians to birds and reptiles to marsupials, rodents, primates and other mammals). In the human, Target 1 represents ~1.9 Mb, and this region is a rich source for many different families of SINEs. Further, a computer software now exists that can process multiple species alignments of those genomic scaffolds. Finally, in this demonstration project, we used a single SINE family (*Alu* elements) to characterize informative loci; the future development of dichotomous SINE-based keys for other groups of organisms will be simplified by including a mix of SINE families.

Acknowledgments

We thank Drs. L. Chemnick, T. Disotell, O. Ryder, and C.-B. Stewart for providing DNA samples. This research was supported by the National Science Foundation grants BCS-0218338 and EPS-0346411 and by the State of Louisiana Board of Regents Support Fund (to M.A.B.). Jacquelyn Zimmerman was supported by a Howard Hughes Medical Institute grant to Louisiana State University through the Undergraduate Biological Sciences Education program.

Appendix A. DNA samples of all species included in this SINE-based dichotomous key

Species names	Common names	Origin	ID number
Hominids (5 genera; 8 species)			
Gorilla gorilla	Lowland gorilla	Coriell[a]	AG05251[b]
Homo sapiens	Human	Roche[c]	1691112
Pan paniscus	Bonobo	IPBIR[d]	PR00661[b]
Pan troglodytes	Common chimpanzee ('Clint')	IPBIR	NS06006[b]
Hylobates gabriellae	Buff (or red) cheeked gibbon	IPBIR	PR00652
Hylobates lar	White-handed gibbon	IPBIR	PR00715
Hylobates syndactylus syndactylus	Siamang	IPBIR	PR00598
Pongo pygmaeus	Sumatran orangutan	Coriell	GM06213A[b]
Old World Monkeys (16 genera; 21 species)			
Allenopithecus nigroviridis	Allen's swamp monkey	IPBIR	PR00198
Cercocebus agilis	Agile mangabey	Disotell[e]	N/A
Cercopithecus lhoesti	L'Hoest's monkey	IPBIR	PR00283
Cercopithecus wolfi wolfi	Wolf's guenon	IPBIR	PR00486
Chlorocebus aethiops	African green monkey	ATCC[f]	CCL70[b]
Colobus guereza kikuyuensis	Kikuyu colobus	SDFZ[g]	OR 160
Erythrocebus patas	Patas monkey	SDFZ	KB5435
Lophocebus aterimus	Mangabey	Disotell	G139
Macaca nemestrina	Pigtailed macaque	Coriell	NG08452
Macaca silenus	Lion-tailed macaque	SDFZ	OR 1890
Macaca mulatta	Rhesus macaque	Coriell	NG07109A
Mandrillus sphinx	Mandrill	Disotell	1148
Miopithecus talapoin	Talapoin	SDFZ	OR 755
Nasalis larvatus	Proboscis monkey	Stewart[h]	N/A
Papio cynocephalus	Yellow baboon	SFBR[i]	9656
Papio anubis	Olive baboon	SFBR	8229
Pygathrix nemaeus	Douc langur	SDFZ	OR 259
Semnopithecus entellus entellus	Hanuman langur	IPBIR	PR00739
Theropithecus gelada	Gelada baboon	SDFZ	KB 10538
Trachypithecus cristatus	Silvered leaf langur	SDFZ	B 4381
Trachypithecus vetulus nestor	Western purple-faced langur	SDFZ	OR 219
New World Monkeys (9 genera; 11 species)			
Alouatta sara	Bolivian red howler monkey	SDFZ	OR 749
Aotus trivirgatus	Three-striped owl monkey	ATCC	CRL1556[b]
Ateles belzebuth	Spider monkey	SDFZ	KB 6701
Ateles geoffroyi	Spider monkey	Coriell	AG05352[b]
Callicebus d. donacophilus	Bolivian gray titi	SDFZ	OR 1522
Callithrix pygmaea	Pygmy marmoset	SDFZ	OR 690
Lagothrix lagotricha	Common woolly monkey	Coriell	AG05356[b]
Pithecia p. pithecia	Northern white-faced saki	SDFZ	OR 842
Saguinus fuscicollis nigrifrons	Geoffroy's saddle-back tarmarin	SDFZ	OR 621
Saguinus labiatus	Red-chested mustached tamarin	Coriell	NG05308
Saimiri s. sciureus	Squirrel monkey	SDFZ	KB 4544
Strepsirrhini and Tarsier (12 genera; 16 species)			
Cheirogaleus medius	Fat-tailed dwarf lemur	IPBIR	PR00794
Daubentonia madagascariensis	Aye-aye	IPBIR	PR01017
Eulemur fulvus albifrons	Brown (white-fronted) lemur	IPBIR	PR00245
Eulemur macaco	Black lemur	IPBIR	PR00266
Galago moholi	Southern lesser bushbaby	IPBIR	PR01035
Galago senegalensis	Lesser (Senegal) bushbaby	Batzer[j]	N/A
Hapalemur griseus griseus	Bamboo (gray gentle) lemur	IPBIR	PR00257
Microcebus murinus	Gray mouse lemur	SDFZ	KB6993
Lemur catta	Ring-tailed lemur	Coriell	NG07099
Mirza coquereli	Coquerel's mouse lemur	IPBIR	PR00871
Nycticebus coucang bengalensis	Bengal slow loris	SDFZ	KB 7302
Nycticebus pygmaeus	Pygmy loris	SDFZ	KB 6028
Otolemur garnettii	Garnett's galago	SDFZ	KB 9235
Tarsius bancanus	Western tarsier	SDFZ	OR2724
Tarsius syrichta	Philippine tarsier	Disotell	N/A
Varecia variegata ruber	Red ruffed lemur	SDFZ	OR 495

[a] Coriell: Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103.

[b] Genomic DNA derived from cell lines. All others (except Human) were derived from whole genome pre-amplification using the Genomiphi genome amplification kit (Amersham, Sunnyvale, CA).

[c] Roche: www.roche-applied-science.com; high molecular weight genomic DNA (blood).

[d] IPBIR: Integrated Primate Biomaterials and Information Resource (Coriell).

[e] Disotell: Blood sample from Dr. Todd Disotell.

[f] ATCC: Cell lines, American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108.

[g] SDFZ: Frozen Zoo, San Diego Zoo, conservationandscience.org.

[h] Stewart: DNA sample from Dr. Caro-Beth Stewart.

[i] SFBR: Southwest Foundation for Biomedical Research.

[j] Batzer: Adenovirus 12 SV40-transformed fibroblasts maintained in lab of Dr. Mark Batzer.

Appendix B. Branch-specific loci for a dichotomous SINE-based (*i.e.*, *Alu*) primate identification key

Primer[a]	AT[b]	Fill[c]	Null	Forward primer (5′–3′)	Reverse primer (5′–3′)	Atypical amplification[d]
Hom-1_A	60	670	370	ggttttctcaagcctttcaac	tctacaccaaggccaagctc	M·42·56·64·65^{i}·66
Hom-1_B	60	920	620	gtagaacatttcaagtctgtagtgaag	aaaagcctgaagtttggaatc	D·41·45·56·73
Hom-1_C	63	570	270	aactaccttgtgtgtsaatgttttg	cagccagagagcatcagtttc	P·17·29·73
Hom-2_A	55	560	240	gcatgtggaacatttaccaaga	cagctcaaaatagcttctaatttcc	F·14·18·54
Hom-2_B	55	830	510	cccagattgattcttcccttta	atgccagttccattatttccac	C^{d}·L·P^{d}·45
Hom-3	55	590	280	gatgtccattgcttttgcagt	gaaagaggcagattcaaagcat	H·J·K·18·19·34·42·43·47
Hom-4_A	55	530	210	aattctttggggtaggtggaat	cacaacgtacaccctaaaatgg	42
Hom-4_B	60	780	480	ttccttgctaaagagagaagtgg	tctccaggaagatgccctac	73
Hom-4_C	63	550	190	atcacactgctgcctcaaac	gcaagagaggttgggaaatg	43·44·45
Hom-5_A	60	1000	700	cagaaacaayggtaagggaggag	tccacaagtacaactasattgagagg	P·21^{Alu}
Hom-5_B	60	750	450	tcgactgatgcagattytcttag	agttgtctttctacttcctcataaytg	49
Hom-5_C	60	1200	900	tcacttaagcccrggaggtc	tttccccttgtaagattttatgc	A·6^{i}·18·27·54
Hom-6	60	600	250	cctagatgcaaatcctgaatatc	ccaaacyaagtagccactgga	B·J·P·42·49^{i}
Hom-7_A	59	620	300	cacaatttcattttccaatcca	ttgcccaaatgtagaatgtgac	D·16·18·29·54
Hom-7_B	58	720	420	gtctcttctgtgccaaaataatg	ttctaagttaattctgtctaggacctc	X^{Alu}·P·20·61·63·64
Hom-7_C	60	1030	740	tgtgagctttcttgggatgaag	cacatatatttctttgctctgtcagc	45·51·73
Hom-7_D	60	500	180	cagaaatgagtacagagcctcac	ttgctccttctattattccatcag	41·45·73
Hom-8_A	55	650	350	ctgaagartggcaggaaacc	cacrtaaatttggagtagtgaatc	H^{Alu}·44
Hom-8_B	55	600	250	tcctcacttgcctaattttctttg	gtgtccaagaagagcaagaaatc	P·46·73
Hom-8_C	58	350	180	cttaggaaagagggaaaagi	cagtaggaaaatgaagcaatc	14·73
Hom-8_D	60	500	180	tctccctcagctaggtcatc	tttgacatctatttggaagc	J·K·22·41
Hom-8_E	60	700	390	acttattttagcatcccctatgc	tttggcasacttagakgaaagac	42·48
Hom-9_A	55	450	150	tacctgtacccgtatttcattcct	tgttctagtggtattgtgggttttt	C^{Alu}·P·4·42·46^{i}·47
Hom-9_B	55	680	380	cctcaggattctactactggtt	gttcaactctttgccacatc	13·14·17·24·44·46
Hom-9_C	60	810	510	atcgactaatggcatcctcct	ctctcccagaaagactgaaaggt	P·17·43^{i}·44^{d}·45·46^{i}·47
Hom-10_A	55	430	120	tgccagacacatcatctaag	cattcctggaaaatttagtg	G^{Alu}·32^{Alu}
Hom-10_B	55	510	180	gtgaatgcagatcattatggaaag	gggaaaagatgcagtgtagagttt	44^{i}
Hom-10_C	60	430	120	atgccattttatccacaact	cacccagaaatttggtaact	P·13·17
Hom-10_D	60	470	150	taaatgtttccctggataacccta	gaaaacgaagtgctcattcagtaa	C·M·18·21·27·45·47
Hom-11_A	55	490	180	agagatatggcatccccttta	ggaagatattattgggcaaa	L·M·P·42·44·46
Hom-11_B	55	810	550	ctccctcccctcccttg	cacctcagattcaaaccaaaac	73
Hom-11_C	60	410	300	ccccagtcccctatagaagataat	tcaaacaaactgggatacttctga	M·N·P·42
Hom-12_A	55	430	130	gttattttgaaggcttgctg	tacgggcttacagtgaattt	L·16·44
Hom-12_B	55	460	130	tagatagcagccccattcca	cctttcgacacgggaaagta	45
Hom-12_C	55	500	190	acatttcgcttgtccctttc	caaagaaacggacgttgcac	H·15^{i}
Hom-12_D	55	470	160	ccagaggctattgatgtagc	cattttgtctactcccttgc	42
Hom-12_E	55	500	200	tgtctcttcaaagttgtttctctg	tttatcatagccagccctttc	W·56·63·66
Hom-12_F	55	760	430	ggtggctggartttgcac	gcaagaatgccatagcctagc	L·42·44·73
Hom-12_G	55	960	660	cctactacaaaggccaagcaag	ttctgcttttatgcttcagctc	18·44·73
Hom-12_H	55	570	270	cctaccttctgttttaatcactgg	gccaaatttttatccaaaygaa	42·73
Hom-12_I	55	500	180	gtttaataacctttctctctttgg	aatacagctgacccwggaatg	None
Cer-13	60	600	250	tgcacatcaggcactttctg	tgtttttgaaaagagcacaacc	O·P·45
Cer-14_A	50	680	370	tctgctgtagatcgctcacaat	ctccccacccctgatttattac	14^{EF}·22
Cer-14_B	51	570	260	tgttttgctctgaccaattcc	caaccaacggcctatcaagt	P·15·44·47·51
Cer-14_C	53	670	360	cgttctcatcacactgtggttc	attgctaccgagttccaaagaa	M·O·P·43
Cer-15_A	55	650	340	ttatgggcccaatttaagtttt	ataaaatggacttgccagatgc	A$^{H.\ syn}$·P
Cer-15_B	55	610	350	tgctcccatttatgaggatttt	atggggttgtctgtctgtctct	None
Cer-15_C	55	630	320	cagttaggtggcttagggaaaa	tcctattggcattaaagcatga	None
Cer-16_A	50	450	140	caatcaatgcgataacaatgg	agtgggcagtaggaacaaggt	F·K·13·14·16·17·18·47·49·50
Cer-16_B	55	450	140	aagggctgtatgccgttttat	aggactgacaacctggtattaagaa	E^{EF}·G·4·54
Cer-16_C	55	490	180	gtcagtatttgccaagttgtgg	ttggttatcctccaaaaagtgc	2
Cer-17_A	50	460	170	tctcccaacatcctcaaacac	ccacttcccaaaccatgataa	A$^{H.\ lar}$·P·17^{EF}·49
Cer-17_B	52	520	210	tatgtcttagggctgcaaatcc	tgttaaaagtgcggttctcaaa	C·P·25·49·46
Cer-17_C	55	490	180	tccttaaggctaatgcaaggac	cctccatacacagtggtaagaaaa	42·50
Cer-18_A	55	570	260	gcctggttcacagtttggtaa	ttgtcaaggagtggggaaata	F·H·K·16·54
Cer-18_B	55	600	280	gaatccaatgaaacccagtg	aagccataccaccagaccac	44
Cer-19_A	50	600	300	tcctctgagctggcacatatt	cctgctagccattcacaagtt	None
Cer-19_B	55	660	350	tcaggatagtgcttgcctcat	tgatgccatagcaataaaaagg	O·P·13·14·42·44·45·47
Cer-20_A	53	520	210	ctggacactggtggaagaaaa	cagagaacccgtttttgtgag	P·8·14
Cer-20_B	56	900	700	acaaaagttktgggtgcttc	aaagttgttttctccctaraagg	None
Cer-20_C	60	750	400	cacaaaatactaaaggactgttaaagg	ggaaatcttatgagagaggagagc	None
Cer-21_A	50	600	290	ttccaagtgtggaaattggtag	tcatcttgaaacaggaatacgg	P·6·42

(continued on next page)

Appendix B (*continued*)

Primer[a]	AT[b]	Fill[c]	Null	Forward primer (5′–3′)	Reverse primer (5′–3′)	Atypical amplification[d]
Cer-21_B	60	560	250	gcttcaaattggtcaacaacct	taaacaaggggaaaggctatga	P
Cer-21_C	60	620	260	agagggacttggacatgcttt	tggggttaaatgattaaatggtg	O
Cer-21_D	55	950	700	tgttgccaaaattgtcttacctc	aatatgttgctgccaatcacc	H^{EF}·54
Cer-22_A	50	720	410	gcagggtaaggcatgactaaaa	tgagggagagtgagtgtggat	8·33·54
Cer-22_B	60	700	390	tccagtgaacagtttgcttcaa	tcagccccatgggatatagtag	G·5·54
Cer-22_C	60	500	210	tctgtctctggtacgtcaggaa	taagattactgtgcagccctga	54
Cer-23_A	55	520	210	taaatcaggactgttcccaagg	cactaggctccacatgacagag	42
Cer-23_B	57	490	180	gattaaatcacgagggcagaac	tggatacccagaagtccaagat	None
Cer-24_A	56	520	210	tggtgctcactatcagcaattt	tgcttaaagtccatcaacatgc	P·5·51
Cer-24_B	56	560	230	ctcaagcttctcccctcctta	aaggcacaggattctgctttt	C·P·42·46
Cer-24_C	60	550	240	tgccaggagaaaacttaaatga	ttaaggggaggaagtagggaag	P·24^{EF}·42·44·45·46·47·48
Cer-25	60	460	110	cttcctagggctgctatgacaa	gagagaacatggccttgctaat	C^{Alu}·D^{EF}
Cer-26	55	520	210	tttttgaacaagagaccccata	agccagataatatccctgttgc	41
Cer-27_A	60	700	380	tctaaggcagccattgagtg	ccaggtttgcctctgactcc	46
Cer-27_B	60	500	180	agctaatatgtcaaggctsctg	cttggcctgatgtgtgtttayag	$A^{H.\ lars}$·41
Cer-28_A	57	470	160	ggcatggctcaaaacttctaaa	catacagcaatgggaagagttg	L·P·49^{i}
Cer-28_B	60	520	210	ttctctttaggtcccaattctttg	tgaattggttaaaagtcagatagca	$A^{H.\ lars}$·M·O·P·6·42
Cer-29_A	57	660	350	catgctcaccttgtattccttg	tgaggtactgctctggtgagtt	O·P·41·45
Cer-29_B	53	690	380	tgatccatccctcttaggagtc	agatctcggtgccacaaatagt	L·P·42·43·46
Cer-29_C	57	560	250	tgccaaaaactcaggttaagaga	attttgggggaaaactgctatc	42·45·47
Cer-30_A	55	640	320	atttggtcatctgtcactgcat	agattaggataaaccccccaagg	$A^{H.\ lars}$·E^{EF}·P·19^{EF}
Cer-30_B	57	630	410	tggcaggatccactcagtagta	tacctttgttgtgccttcct	P·17·41^{d}
Cer-31_A	57	520	210	caaccaacagaataggggaaaa	ttcttcatgatattgcctttgc	C^{Alu}·O·P
Cer-31_B	55	700	400	aaggcactggggaatacaaag	ttaaatcattgctgtccttttgc	D^{d}·F·18^{d}·54
Cer-32_A	55	540	230	ggctccctctcccttagactac	taccaggaatgtttggaaatgg	P·42
Cer-32_B	57	580	270	ctctcagctccctgtttctgtt	catggacatcagactagccact	P
Cer-32_C	58	530	200	tatcaagagcaatctttcaattatg	accatatgggctcagtctgc	5^{i}·44
Cer-33_A	55	530	220	acagggatttcagagcacaagt	cagctcattattcccaaacaca	None
Cer-33_B	55	620	310	ttttgtaaacagccaaagctca	ttgttgaaaatatggcacaagc	P
Cer-33_C	55	560	250	ttctaagcaggacctaaaaagca	tcatttcagagattttgctgatg	P·45·51
Cer-33_D	60	440	130	atgtgaagacctctgcccagta	tcctttgttcaaactgcttctt	M·O·42·52
Cer-34_A	50	510	200	ccacccctttcctaaatttcca	gtcttggctttcccttctacg	P·43·47
Cer-34_B	55	620	370	cctgaggctgctcagagaaa	tgcttatacgaggcaaccttta	41
Cer-34_C	60	540	230	cagattcaaagaggtgatgtcg	taggtgtctcaggatggtgcta	6·44·45·47
Cer-34_D	60	500	260	accacttttgcttggttctcat	tgtccagaccaagagacatctg	O·43
Cer-34_E	60	690	380	caaatgttccgttgagtcca	gagtcttggaagatgcagtga	D·17·47·52
Cer-35_A	55	550	190	cgtaactcttgttttaactgcttg	tcaagggtaggaaaatcattgc	M·44·45
Cer-35_B	58	600	250	ttcaccatccaccagacg	ggggatggtcaaatggaac	$A^{H.\ lars\text{-}i}$·15^{EF}·19·29·54
Cer-35_C	60	650	350	acaggtgccaaggaacaaag	tcaggggagaagacagcaag	50
Cer-36_A	55	640	300	caacgctatcatccttttgaca	taatagattgccaacccctgaa	None
Cer-36_B	54	700	400	cttgttgagcatcttytgtgc	tctagttc**c**gttgatgtaataatgc	M·P·46
Cer-36_C	63	900	700	tggtgggctaaggagaagg	cccagaggaaagtctgaattg	18·54
Cer-37_A	55	550	230	gcaacatttacaagatgctcaag	acttgggagcaaacatctcct	$A^{H.\ lars}$·L·O·P
Cer-37_B	57	750	440	aagaagtcacaaagagggaagg	ttttgctggtggttccttaaac	P·47
Cer-37_C	60	480	170	ctctacttttaggggcctgctt	ccagacctggaaactcaagaat	None
Cer-38_A	55	490	170	agcaaatcttgaagcaccagtt	gcagcaataataggtttggttca	O·P·20
Cer-38_B	60	720	410	gtagaatgaccaaagcgaggtt	cacatcagtgggtgcctattta	42
Cer-38_C	60	680	380	agtttggaagcagtgagtcagg	ccagtttgcaattgatttctga	G·P·45
Cer-38_D	53	550	250	tcatcacagaaagtcagtgtcag	aaaaggtggaaatagtcacataatc	None
Cer-38_E	56	550	200	tcaaagccagtgcttattaaas	taagtggacaaggccaaagc	54
Cer-38_F	56	600	280	aagatgtttgctgacagartatg	tctttgacagatagtgcttttattc	L^{i}
Cer-39_A	50	470	160	aaacatcagtgaaaatttgcttaca	tcacatgtatttgtgcaacacttc	P
Cer-39_B	55	850	520	tgccataaactgaccctgacta	ggcaccaaagagaaacacatct	None
Cer-39_C	60	580	250	gtgcaagtgaacatgtggctat	ccacaatttctaagagggcaag	L
Cer-39_D	60	790	450	cagccaaaacactaaatgttgg	tcattttcagtgtcagtttgcat	L·42·44
Cer-39_E	60	580	260	aggactgactgagaaccagagtc	acagctccaatttcaccctaaa	F·G·P·14·18·44
Cer-39_F	54	500	170	tcatactttatgtgcaattttatatcc	tgctcaacaacagkaaaatg	None
Cat-40_A	60	890	590	gctcaccatgatctaatttcctg	ctggttctgcaatcacatcttc	54
Cat-40_B	60	840	480	tcatgtgaggaagagcttgag	gttgtcacagcatagcagaatc	42·56·73
Cat-40_C	60	410	110	tgaatctctgagaattascaccac	aagatcaagatgctggagaatg	15·73
Cat-40_D	60	800	500	cacttatactctgtggagaacyctaac	gaaaagrcccaatttgtagggaag	$A^{H.syn.}$·T·42·43·56·59·65
Cat-40_E	60	680	130	tctgtcatggaagctagaagagc	ctgcccagagasggtatttactatg	T·44·56·60·66
Pla-41_A	55	520	200	gcacctacccacgaaaagaa	acattgcctctgcaaattga	C·P·51
Pla-41_B	55	360	220	aarcttaggccagctggaaa	ggytgaggaagtggtttgga	18·20[i]·23[i]

Appendix B (*continued*)

Primer[a]	AT[b]	Fill[c]	Null	Forward primer (5′–3′)	Reverse primer (5′–3′)	Atypical amplification[d]
Pla-42_A	53	530	200	ctcagcccaacatcacagaa	cacatcatgaaatcttccttctct	G·6
Pla-42_B	59	490	200	gtgcttggtcccaccctact	gcagcatatggagccctaga	13·38
Pla-43_A	53	490	250	cccaaagcacagcatagtctg	caccctaacctggtagaagag	C^{i}·41
Pla-43_B	56	510	210	gcaaaagcttatggtttctgag	tagagcctcccaagtgccta	17
Pla-43_C	56	590	290	agcaggattcaaggaagcag	gaacagtaagagagctgtggagg	A^{EF}·6^{EF} (with ^{E}∼400 bp<^{F})
Pla-43_D	57	580	260	cgatactgcgatgtagagagcc	gagttcctcttctcttcacagc	None
Pla-44_A	50	510	210	gcttccagtagtcgttgttcc	cagcttttgctrtgttaangc	A$^{H.\ syn.}$·45
Pla-44_B	55	520	210	tccaaagagnaaacaaaaggaa	gcctaagatgtttggggaaa	45
Pla-44_C	55	480	150	tttggtcaaagatgggtcttc	acatcaggcatctcacttgc	None
Pla-45_A	55	580	280	aacgtttggccttcatccta	ggagcagaatgnnatatggaag	G·I·17·36
Pla-45_B	55	740	430	gccaggtagccaacatcagt	ccaacarggaaatgtggtct	O·43^{Alu}
Pla-45_C	59	500	190	gagggaggtgctgaggagta	agcagctgtgtaaacccaaaa	None
Pla-46	55	1000	420	gcacayaactttgttttgnag	ggccytatttrgcaggttgg	H·J·K·P·17·18·19·22·42·45
Pla-47_A	55	540	240	gcgagatgaatcagtgctca	aaggtggcrgcaagtagata	16
Pla-47_B	55	570	230	gcaaagrrtgcataaaggttg	tccctgttattcytyataggttctg	None
Pla-48_A	56	660	320	acaggacctttttgggtttca	cytgagattgtggagaatgattt	J^{EF}·L·P·17·48^{EF}
Pla-48_B	56	420	290	atgcagcaagagacagcttg	tcccaccaaantaggactgc	E·M^{i}·P·43·46
Pla-49_A	52	590	260	ttgaacaaaaccatcagagca	cagtngaatrccttttgcatca	None
Pla-49_B	54	650	340	ccaagacagacccaatatcca	ccagccagcactttgnatca	15^{Alu}
Pla-49_C	57	490	210	tgaaacctcattgcagtcca	tgctccanttgcaagtatgc	A$^{H.\ syn.}$·49^{EF}
Pla-50_A	55	530	230	carnttaggyagtgtattgtgg	ctgcccacccctagactttt	A$^{H.\ syn.}$·K·17·32
Pla-50_B	56	620	310	gagtctcacagccctccaag	tgaatctaacacctctgaatgtaggc	A$^{H.\ syn}$·E^{d}·G^{d}·6^{d}·37
Pla-50_C	58	500	200	tgtccagaaaatggatcctga	gagttagccaaaratgyatgg	17^{Alu}·43
Pla-51_A	48	610	310	ctgggtgcacataaacctctc	tagctcagtgacttctagacac	14
Pla-51_B	56	510	240	aggaacaaagaggaaaggtactc	tgaattttatttggargattttga	C·H·42
Pla-51_C	59	630	300	gtgaggtgggctgtgagaac	caaatggcattgtgrcaagag	P
Pla-52_A	55	310	170	gtayataaagaraggcaaaagacycag	ttgagcacagagttgggaaa	A$^{H.\ syn.}$·E·11·13·14
Pla-52_B	55	500	350	cagcagtgcctgaaaatctc	ggcnatccttgacccttagt	G·H·K·P·18·19
Pla-53_A	50	500	190	ccttgaacgatgaaggcaag	tccancatagttacrtgcaagg	G·P·5·13·14
Pla-53_B	54	570	290	ggatgcttatggtatggtaagtg	gaaagagagctgaaactcaggg	45
Pla-53_C	58	610	300	ggttcatgaagctctgattgc	atcrtggagccacctctatc	C·17·35·43·44
Pla-54_A	55	460	150	ctgcatcyctgggtctggat	tcacctgrggatcatttcca	None
Pla-54_B	55	440	140	aaagaagttagcagtcatcattcaa	tgttaaccaaagcctctattaaattc	P
Pla-54_C	55	480	160	gggtctgacctaagagaactgg	gattgttgaagggagggtca	None
Pla-54_D	55	480	190	aaaaagatgtcccatgttcaaga	rggcttcacacagaaagcac	None
Ant-55_A	50	650	350	gctagggagaatgatgatgaag	aattaggtattccctttaaytcaca	P·41·56·61
Ant-55_B	60	880	580	gcagcaatgtaaaggcaattagttc	gcattttgattccccaggag	C·19^{i}·29·44
Ant-55_C	50	1060	730	ggaaaattgcaaaagcatatctac	atgcatttttgtcgttccag	P·61
Ant-55_D	60	500	280	gaagaatggaaagaaattgtgtagc	aagctggggattcattcataaaatac	P
Ant-55_E	60	1100	700	gcattttattagctccaagttcaag	agttaatacccaggatgagaatgg	None
Tar-56		650	350	Same primer as Str-71_C	same primer as Str-71_C	71^{Alu}
Str-57_A	55	840	500	tgtttcaaagatcaaatgtcatgg	cccaaggggagacagaagtt	P$^{i-(\sim 300\ bp)}$·Q·46$^{i-(\sim 300\ bp)}$·56^{i}·65
Str-57_B	58	640	300	agratcctggaatactgrtcct	gctgagatgyatccctccyg	L·56$^{T.\ syr.}$·59
Str-57_C	54	880	580	caccacaactaccctggaagr	agaactctccttbacttaacctgtttc	L·T·X·56·65
Str-58_A	52	520	160	atagccthtgcaggtcraag	gagcttaatatataavtwttctcaac	H^{i}·56·61
Str-58_B	53	570	220	ccttgtgacctctgaagtaaaractc	tgactctgtgtybttgtctaaactag	None
Str-59	60	1060	750	tacccagtgctcatttggatc	gcaatcagcaaatgcttgttttag	None
Str-60		1000	650	Same primer as Str-66_A	same primer as Str-66_A	W$^{only\ Str\text{-}65\ Alu}$·66^{Alu}
Str-61	58	1070	760	gcccagtttgtaattcttgatttc	gcaacccagctwacttcagagg	56
Str-62	55	970	620	ctcacttfcraaaatagcctactgc	aaggtcagtgttccytctctkc	46
Str-63_A	50	550	250	tcctctrtgactctyytgagttag	ggcacawttaagctatattrthcaa	X·56·65
Str-63_B	53	480	220	aggaagaggctgrtgraagg	gaatttcamatttttaggctcag	F·S^{i}·56
Str-64	55	650	300	tgaggcaartaaatatcttgakagc	tggagcccttgcwgtyg	F·H·T^{EF}·V^{d}·27^{d}·43·56$^{T.\ buc}$·69
Str-65		720	650	Same primer as Str-66_A	same primer as Str-66_A	60^{Alu}·66^{Alu}
Str-66_A	58	950	650	ctgcctcttaccatatttgacttc	gcctattcacatccatcactc	W$^{Str\text{-}65\ Alu}$·60$^{Alu\ +\ Str\text{-}65\ Alu}$
Str-66_B	60	980	680	gatgctaatgtggcttaatctcag	agccagctcttgaaatgtgc	H·65
Str-67	53	910	550	gctttctgcatrggaggtagaataat	ccaatatcdtggcattctgtwaactc	L·S·34·43·56$^{T.\ syr.}$
Str-68	55	1060	740	aggctacagattttcaagatttacc	ctttagagccctctgctgga	42
Str-69_A	55	990	690	gctgtgccaaagaaatcagg	ggcacatttccttttgtttggag	None
Str-69_B	53	610	400	tgggraaaatgtttggcagat	tgtgattcattcttttggagca	L^{i}
Str-70_A	54	600	300	gtgyatgtrtgttggggcaar	cacttcytagcytgrtggcttt	V^{EF}·T^{i}·47·56·60·61
Str-70_B	56	610	310	matacagtgaatacttgaatgkgc	taatcactaatgaaaatgtatctgarag	U·7·42·54^{i}·56·61·65
Str-71_A	52	500	190	cctagcctttgtgaggraagtr	gttcagaagaayccattaycaytg	P·T·44·46·56·62·65
Str-71_B	48	470	170	gtcttattggatargaagaanctagg	gcctaaggtdcaaahttgtctc	P·56·65

(continued on next page)

Appendix B (*continued*)

Primer[a]	AT[b]	Fill[c]	Null	Forward primer (5′–3′)	Reverse primer (5′–3′)	Atypical amplification[d]
Str-71_C	55	650	350	tgtgctatccctttttaggrmtg	caacacaaaayvttgtcttaatrgg	S·42·52·56^{Alu}
Str-72	50	1020	700	ttcartttcactggraatcctttag	agagatcwtcagaactatgcccttr	16·43·56
Str-73_A	54	580	200	ccatttaggaagatgtgycatg	tgcttcactgagatamaagavga	P·X·17^{Alu}·42$^{i-(2\ bands)\ +\ Alu}$·45
Str-73_B	61	600	280	tggagaggacctaaattcaaagyttg	gaaagtgagytgggyatgaac	19^{d}·42·61
Str-73_C	60	550	280	gaagggaagasagtttcctaarc	aacaccacctkataaccarcc	S·44^{i}·50·51·65
Str-73_D	55	790	410	ggcccacaaaaattaaatcag	atagggtttcccaggaaattayagc	56

[a] Primer designations: Hom, Hominids; Cat, Catarrhini; Cer, Cercopithecidae; Pla, Platyrrhini; Tar, Tarsiformes; and Str, Strepsirrhini. Numbers are keyed to branches in Fig. 1, which are defined by the insertion of an *Alu* element at the associated loci.

[b] AT, Annealing temperature (°C).

[c] Fragment sizes (bp) of filled and empty sites are approximate and should be used only as a guide for gel interpretation.

[d] Branches in Fig. 1 for which PCR of a locus gave atypical results: (I) no amplification or multiple fragments ($^{no\ superscript\ label}$); (II) fragments >50 bp larger (^{i}) or smaller (^{d}) than expected; (III) independent insertion (^{Alu}); or (IV) two fragments having the approximate sizes of empty and filled alleles $^{(EF)}$. Further PCR optimization might improve or eliminate these atypical results. With the exception of some loci on Branches 40 and 55 as well as the clade defined by Branch 73, PCR with DNA from Strepsirrhines and Tarsiformes usually resulted in no amplification or multiple fragments.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Badge, R.M., Alisch, R.S., Moran, J.V., 2003. ATLAS: a system to selectively identify human-specific L1 insertions. Am. J. Hum. Genet. 72, 823–838.

Batzer, M.A., Deininger, P.L., 2002. *Alu* repeats and human genomic diversity. Nat. Rev., Genet. 3, 370–379.

Biedler, J., Qi, Y., Holligan, D., della Torre, A., Wessler, S., Tu, Z., 2003. Transposable element (TE) display and rapid detection of TE insertion polymorphism in the *Anopheles gambiae* species complex. Insect Mol. Biol. 12, 211–216.

Brody, J.R., Kern, S.E., 2004a. History and principles of conductive media for standard DNA electrophoresis. Anal. Biochem. 333, 1–13.

Brody, J.R., Kern, S.E., 2004b. Sodium boric acid: a Tris-free, cooler conductive medium for DNA electrophoresis. Biotechniques 36, 214–216.

Deininger, P.L., Batzer, M.A., 2002. Mammalian retroelements. Genome Res. 12, 1455–1465.

Disotell, T., 2003. Primates: Phylogenetics. Nature Publishing Group.

Domingo-Roura, X., Marmi, J., Ferrando, A., Lopez-Giraldez, F., Macdonald, D.W., Jansman, H.A.H., 2006. Badger hair in shaving brushes comes from protected Eurasian badgers. Biol. Conserv. 128, 425–430.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. 95–98.

Hamdi, H., Nishio, H., Zielinski, R., Dugaiczyk, A., 1999. Origin and phylogenetic distribution of *Alu* DNA repeats: irreversible events in the evolution of primates. J. Mol. Biol. 289, 861–871.

Han, K., et al., 2005. Under the genomic radar: the stealth model of Alu amplification. Genome Res. 15, 655–664.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. PLoS Biol. 2, e312.

Hillis, D.M., 1999. SINEs of the perfect character. Proc. Natl. Acad. Sci. U. S. A. 96, 9979–9981.

Hsieh, H.M., et al., 2003. Species identification of rhinoceros horns using the cytochrome *b* gene. Forensic Sci. Int. 136, 1–11.

Kohn, M., Knauer, F., Stoffella, A., Schroder, W., Paabo, S., 1995. Conservation genetics of the European brown bear—a study using excremental PCR of nuclear and mitochondrial sequences. Mol. Ecol. 4, 95–103.

Kriegs, J.O., Churakov, G., Kiefmann, M., Jordan, U., Brosius, J., Schmitz, J., 2006. Retroposed elements as archives for the evolutionary history of placental mammals. PLoS Biol. 4.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Lander, E.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Lathuilliere, M., Menard, N., Gautier-Hion, A., Crouau-Roy, B., 2001. Testing the reliability of noninvasive genetic sampling by comparing analyses of blood and fecal samples in Barbary macaques (*Macaca sylvanus*). Am. J. Primatol. 55, 151–158.

Lorenz, J.G., Jackson, W.E., Beck, J.C., Hanner, R., 2005. The problems and promise of DNA barcodes for species diagnosis of primate biomaterials. Philos. Trans. R. Soc. Lond., B Biol. Sci. 360, 1869–1877.

Malisa, A., Gwakisa, P., Balthazary, S., Wasser, S., Mutayoba, B., 2005. Species and gender differentiation between and among domestic and wild animals using mitochondrial and sex-linked DNA markers. Afr. J. Biotechnol. 4, 1269–1274.

Matsubara, M., et al., 2005. Species and sex identification of western lowland gorillas (*Gorilla gorilla gorilla*), eastern lowland gorillas (*Gorilla beringei graueri*) and humans. Primates 46, 199–202.

Moritz, C., Cicero, C., 2004. DNA barcoding: promise and pitfalls. PLoS Biol. 2 e354.

Nikaido, M., Rooney, A.P., Okada, N., 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. Proc. Natl. Acad. Sci. U. S. A. 96, 10261–10266.

Nikaido, M., et al., 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. Mol. Biol. Evol. 23, 866–873.

Okada, N., Shedlock, A.M., Nikaido, M., 2004. Retroposon mapping in molecular systematics. Methods Mol. Biol. 260, 189–226.

Palumbi, S.R., Cipriano, F., 1998. Species identification using genetic tools: the value of nuclear and mitochondrial gene sequences in whale conservation. J. Heredity 89, 459–464.

Perna, N.T., Batzer, M.A., Deininger, P.L., Stoneking, M., 1992. *Alu* insertion polymorphism: a new type of marker for human population studies. Hum. Biol. 64, 641–648.

Ray, D.A., et al., 2005a. Inference of human geographic origins using *Alu* insertion polymorphisms. Forensic Sci. Int. 153, 117–124.

Ray, D.A., et al., 2005b. *Alu* insertion loci and platyrrhine primate phylogeny. Mol. Phylogenet. Evol. 35, 117–126.

Ray, D.A., Xing, J., Salem, A.-H, Batzer, M.A., in press. SINEs of a nearly perfect character. Systematic Biology.

Roos, C., Schmitz, J., Zischler, H., 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc. Natl. Acad. Sci. U. S. A. 101, 10650–10654.

Roy, A.M., et al., 1999. Recently integrated human *Alu* repeats: finding needles in the haystack. Genetica 107, 149–161.

Roy-Engel, A.M., et al., 2002. Non-traditional *Alu* evolution and primate genomic diversity. J. Mol. Biol. 316, 1033–1040.

Rozen, S., Skaletsky, H., 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

Salem, A.H., Kilroy, G.E., Watkins, W.S., Jorde, L.B., Batzer, M.A., 2003a. Recently integrated *Alu* elements and human genomic diversity. Mol. Biol. Evol. 20, 1349–1361.

Salem, A.H., et al., 2003b. *Alu* elements and hominid phylogenetics. Proc. Natl. Acad. Sci. U. S. A. 100, 12787–12791.

Salem, A.H., Ray, D.A., Batzer, M.A., 2005. Identity by descent and DNA sequence variation of human SINE and LINE elements. Cytogenet. Genome Res. 108, 63–72.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467.

Schmitz, J., Ohme, M., Zischler, H., 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157, 777–784.

Schmitz, J., Roos, C., Zischler, H., 2005. Primate phylogeny: molecular evidence from retroposons. Cytogenet. Genome Res. 108, 26–37.

Schwartz, S., et al., 2000. PipMaker — a Web server for aligning two genomic DNA sequences. Genome Res. 10, 577–586.

Shedlock, A.M., Okada, N., 2000. SINE insertions: powerful tools for molecular systematics. BioEssays 22, 148–160.

Shedlock, A.M., Takahashi, K., Okada, N., 2004. SINEs of speciation: tracking lineages with retroposons. Trends Ecol. Evol. 19, 545–553.

Sheen, F.M., et al., 2000. Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res. 10, 1496–1508.

Taberlet, P., et al., 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Mol. Ecol. 6, 869–876.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 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.

Ustyugova, S.V., Amosova, A.L., Lebedev, Y.B., Sverdlov, E.D., 2005. Cell line fingerprinting using retroelement insertion polymorphism. Biotechniques 38, 561–565.

Van de Lagemaat, L.N., Gagnier, L., Medstrand, P., Mager, D.L., 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15, 1243–1249.

Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inf. Ser. Workshop Genome Inf. 12, 141–154.

Walker, J.A., Hughes, D.A., Anders, B.A., Shewale, J., Sinha, S.K., Batzer, M.A., 2003a. Quantitative intra-short interspersed element PCR for species-specific DNA identification. Anal. Biochem. 316, 259–269.

Walker, J.A., Kilroy, G.E., Xing, J., Shewale, J., Sinha, S.K., Batzer, M.A., 2003b. Human DNA quantitation using *Alu* element-based polymerase chain reaction. Anal. Biochem. 315, 122–128.

Walker, J.A., Garber, R.K., Hedges, D.J., Kilroy, G.E., Xing, J., Batzer, M.A., 2004a. Resolution of mixed human DNA samples using mitochondrial DNA sequence variants. Anal. Biochem. 325, 171–173.

Walker, J.A., et al., 2004b. Quantitative PCR for DNA identification based on genome-specific interspersed repetitive elements. Genomics 83, 518–527.

Walker, J.A., et al., 2005. Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: application in human identification. Anal. Biochem. 337, 89–97.

Wolfe, N.D., et al., 2005. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. Proc. Natl. Acad. Sci. U. S. A. 102, 7994–7999.

Xing, J., et al., 2005. A mobile element based phylogeny of Old World monkeys. Mol. Phylogenet. Evol. 37, 872–880.

Yan, P., Wu, X.B., Shi, Y., Gu, C.M., Wang, R.P., Wang, C.L., 2005. Identification of Chinese alligators (*Alligator sinensis*) meat by diagnostic PCR of the mitochondrial cytochrome *b* gene. Biol. Conserv. 121, 45–51.