YEARBOOK OF PHYSICAL ANTHROPOLOGY 50:2–19 (2007)

Mobile DNA Elements in Primate and Human Evolution

Jinchuan Xing,[1] David J. Witherspoon,[1] David A. Ray,[2] Mark A. Batzer,[3] and Lynn B. Jorde[1]*

[1]*Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112*
[2]*Department of Biology, West Virginia University, PO Box 6057, Morgantown, WV 26506*
[3]*Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multi-scale Systems, Louisiana State University, Baton Rouge, LA 70803*

KEY WORDS *Alu*; retrotransposon; primate phylogeny; human population genetics; forensics

ABSTRACT Roughly 50% of the primate genome consists of mobile, repetitive DNA sequences such as *Alu* and LINE1 elements. The causes and evolutionary consequences of mobile element insertion, which have received considerable attention during the past decade, are reviewed in this article. Because of their unique mutational mechanisms, these elements are highly useful for answering phylogenetic questions. We demonstrate how they have been used to help resolve a number of questions in primate phylogeny, including the human–chimpanzee–gorilla trichotomy and New World primate phylogeny. *Alu* and LINE1 element insertion polymorphisms have also been analyzed in human populations to test hypotheses about human evolution and population affinities and to address forensic issues. Finally, these elements have had impacts on the genome itself. We review how they have influenced fundamental ongoing processes like nonhomologous recombination, genomic deletion, and X chromosome inactivation. Yrbk Phys Anthropol 50:2–19, 2007.

Mobile DNA elements are discrete DNA sequences that have the remarkable ability to transport or duplicate themselves to other regions of the genome. This process has been occurring in virtually all organisms for many millions of years. As a result, mobile elements are a major component of human and nonhuman primate genomes, accounting for 40–50% of their contents (Lander et al., 2001; CSAC, 2005; Gibbs et al., 2007). The properties and applications of these mobile elements have been the subject of increasing interest during the past two decades. In this review, we will focus on several areas in which the study of mobile elements has made significant contributions to physical anthropology.

Because of their unique manner of propagating themselves in the genome, mobile elements are highly useful in tracing relationships of individuals, populations, and species. We will demonstrate how these elements have been used to resolve questions about primate systematics. We will then show how they have been useful in addressing questions about human origins, population affinities, and population history. The potential to use mobile elements as forensic tools, both in human and in nonhuman primates, will be discussed next. Finally, the presence of more than one million mobile elements in the primate genome has had a significant impact on processes such as nonhomologous recombination, genomic deletion and duplication, and gene conversion. The nature and consequences of this impact will be the final topic of the review. Before reviewing these applications, we begin with a brief discussion of the basic biology that underlies mobile elements and their behavior in the genome.

BASIC MOBILE ELEMENT BIOLOGY

Mobile elements can be divided into two different classes based on how they duplicate themselves within the genome. DNA transposons (Fig. 1A) mobilize through DNA intermediates, typically using a so called "cut and paste" mechanism (Smit and Riggs, 1996; Pace and Feschotte, 2007). During this process, transposase, an enzyme encoded within some transposons, first excises a DNA transposon from its original genomic location. It then generates a break in another genomic location, where it reinserts the transposon. A short stretch of identical DNA sequence ("target site duplications" or TSDs) is created on either end of the newly integrated transposon during this enzymatic process. In the primate lineage, DNA transposons have been shown to be inactive for the past 40 million years or so (Pace and Feschotte, 2007). Therefore, this review will focus on the other major class of mobile elements: retrotransposons.

Retrotransposons mobilize through RNA intermediates using a "copy and paste" mechanism (Ostertag and Kazazian, 2001; Deininger and Batzer, 2002). In this process, an RNA copy is first generated from the original retrotransposon and is subsequently reverse-transcribed back into DNA using an enzyme called reverse transcriptase. It is then inserted into a new location in the genome. As in the transposition of DNA transposons, target site duplications are usually generated during this process. Retrotransposons can be further subdivided into those elements that are autonomous, meaning that they encode their own replication machinery [e.g. Long Inter-

Grant sponsor: National Science Foundation; Grant numbers: BCS-0218370, BCS-0218338, EPS-0346411; Grant sponsor: National Institutes of Health; Grant number: GM-59290; Grant sponsor: Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF; Grant numbers: (2000-05)-05, (2000-05)-01, (2001-06)-02.

*Correspondence to: Lynn B. Jorde, Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA. E-mail: lbj@genetics.utah.edu

DOI 10.1002/ajpa.20722
Published online in Wiley InterScience (www.interscience.wiley.com).

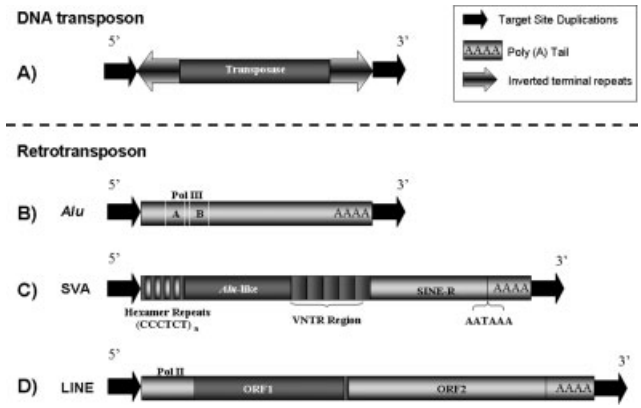

Fig. 1. General structures for the major types of primate mobile elements. (**A**) DNA transposon. Active DNA transposons encode an enzyme called transposase that is used for their transposition. On both ends of the DNA transposon are stretches of sequences that are identical when reading in opposite directions ("inverted terminal repeats" or ITRs) and stretches of identical sequences generated during the integration process ("target-site duplications" or TSDs). (**B**) *Alu* element. Each full-length *Alu* element is about 300 base pair (bp) in length. *Alu* elements do not encode any proteins. They are transcribed by RNA polymerase III, which recognizes the A and B promoter sequences near the 5′-end of the elements. (**C**) SVA element. SVA elements can be divided into five components: (1) a $(CCCTCT)_n$ hexamer simple repeat region, which is located at the 5′-end; (2) an *Alu* homologous region; (3) a VNTR region, composed of a variable number of copies of a 35–50 bp sequence; (4) a SINR-R region, which is derived from the 3′-end of the endogenous retrovirus HERV-K10; and (5) a poly(A) tail after a putative polyadenylation signal (AATAAA); (**D**) LINE. Full length LINE contains an RNA polymerase II promoter region and encode an RNA-binding protein as well as a second protein with endonuclease and reverse transcriptase activity. These proteins provide enzymatic machineries that are necessary for the LINE mobilization in the genome.

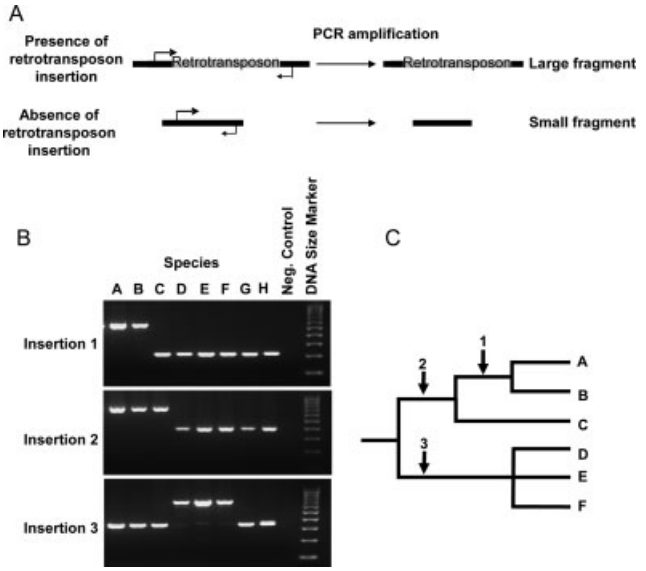

Fig. 2. Schematic of the mobile element insertion-based phylogenetic analysis. (**A**) PCR Assay for mobile element insertions. Mobile element insertions are represented by gray lines, and the flanking chromosomal sequences are represented by black lines. PCR primers are denoted by bent arrows. (**B**) Gel electrophoresis results of three mobile element insertions. The presence of mobile element generated a higher band, whereas the absence of mobile element generated a lower band. (**C**) The phylogenetic relationships among species A to F inferred from the three insertions.

spersed Element 1 (LINE1 or L1)] (Ostertag and Kazazian, 2001), and those that are nonautonomous, such as the primate-specific *Alu* family (Deininger and Batzer, 1993; Batzer and Deininger, 2002).

Alu elements are a family of retrotransposons found in all primates (Fig. 1B). They belong to a larger nonautonomous element category known as short interspersed elements (SINEs) that have been identified in a variety of taxa (Schmid and Jelinek, 1982; Singer, 1982; Weiner et al., 1986; Okada, 1991; Smit, 1996). *Alu* elements are thought to have been derived from the 7SL RNA gene early in the evolution of the primate order and are ~65 million years old. In the human, chimpanzee, and macaque genomes, these elements have reached copy numbers in excess of one million. As nonautonomous retrotransposons, they borrow the enzymatic machinery required for their propagation from L1 elements (Kajikawa and Okada, 2002; Dewannieux et al., 2003). A second family of nonautonomous retrotransposons was recently discovered in hominids and termed "SVA" for each of its three components [SINE-R, VNTR (Variable Number of Tandem Repeat), and *Alu*] (Shen et al., 1994; Wang et al., 2005) (Fig. 1C).

L1 elements (Fig. 1D), the most successful class of autonomous retrotransposons in mammalian genomes (Smit et al., 1995; Ostertag and Kazazian, 2001), arose as a family of repeated DNA sequences about 150 million years ago. In primate genomes, these elements have reached a copy number in excess of 500,000 elements (Lander et al., 2001; CSAC, 2005; Gibbs et al., 2007). The majority of these elements are truncated or defective copies that were originally produced by a relatively small number of full-length, retrotransposition-competent copies (Brouha et al., 2003; Han et al., 2007). These full-length L1 elements are about 6 kb in length and encode an RNA-binding protein as well as a second protein with endonuclease and reverse transcriptase activity. Because they can make copies of themselves and are likely the source of all other L1 elements in the genome, these L1 elements are termed "master" mobile elements. The human genome contains 80–100 of these retrotransposition-competent elements (Brouha et al., 2003). Curiously, only a few retrotransposition-competent L1 elements were recovered from the draft sequence of the rhesus macaque genome (Han et al., 2007).

Retrotransposon-based genetic systems were first applied as a research tool in the early 1990s (see Fig. 2) (Batzer and Deininger, 1991; Batzer et al., 1991; Perna et al., 1992; Minghetti and Dugaiczyk, 1993; Murata et al., 1993; Batzer et al., 1994). Their use was highlighted by phylogenetic studies of salmon (Murata et al., 1993; Murata et al., 1996), whale (Shimamura et al., 1997; Nikaido et al., 1999), and cichlid fish (Takahashi et al., 1998, 2001). Early applications of these systems also addressed questions about human origins and population affinities (Batzer et al., 1994, 1996; Stoneking et al., 1997). Since then, retrotransposons have become widely recognized as powerful tools for phylogenetic and population genetic studies (Shedlock and Okada, 2000; Shedlock et al., 2004; Schmitz et al., 2005; Ray et al.,

2006). Controversial phylogenetic relationships that could not be resolved using traditional molecular data have been successfully resolved in a variety of taxa (Zampicinini et al., 2004; Churakov et al., 2005; Nishihara et al., 2006; Xing et al., 2007) [also see (Shedlock et al., 2004; Schmitz et al., 2005; Ray et al., 2006) for reviews].

Retrotransposon-based systems have three important advantages compared to other types of genetic loci [e.g., single nucleotide polymorphisms (SNPs), microsatellites, and restriction fragment length polymorphisms (RFLPs)]. First, the probability that two retrotransposons will insert independently in the same position is essentially zero. Even if two retrotransposons did insert at the same location, they can usually be distinguished by the type of insertions and length of the flanking TSDs generated during their integration (Conley et al., 2005). In addition, the precise removal of an entire retrotransposon after its fixation in the genome is extremely rare (van de Lagemaat et al., 2005) and is very unlikely to happen in multiple genomes. These two characteristics suggest that retrotransposons are essentially free of homoplasy and that two individuals sharing the same retrotransposon insertion most likely acquired it from a common ancestor (identity by descent) (Salem et al., 2005; Ray et al., 2006). This allows retrotransposon-based studies to infer relationships more accurately than other systems, in which different individuals may acquire the same genotype independently (i.e., identity by state/homoplasy). Second, the ancestral state of any given locus can be determined unambiguously as the absence of the retrotransposon insertion. This simplifies phylogenetic inference and presents another major advantage over other marker systems. Third, unlike most other molecular systems, the use of retrotransposons in phylogenetic and population genetic inference does not directly rely on comparisons of DNA sequence variation. This feature allows retrotransposons to complement traditional DNA sequence-based molecular studies. Together, these features make retrotransposon-based systems especially useful for groups that have high levels of sequence homoplasy and for closely-related species for which there is little phylogenetic information in DNA sequences.

Nevertheless, like any other genetic system, retrotransposon-based systems are not perfect. Several factors, including lineage sorting, parallel or near-parallel insertion, and precise deletion of the element, may potentially create problems and confound the results (Hillis, 1999; Ray et al., 2006). Several studies have shown that most confounding events happen at very low rates, and basic cautions, including careful analysis and interpretation of the data, can limit their potential influence (Shedlock and Okada, 2000; Shedlock et al., 2004; Salem et al., 2005; Ray et al., 2006). Basic knowledge of the taxa under investigation and collection of a sufficient number of informative insertions can also increase confidence in results (Waddell et al., 2001).

Although in principle all retrotransposons can be used as phylogenetic or population genetic markers, SINEs have been the choice of markers in most studies for two reasons. First, the amplification of SINEs has been well studied in many taxa, and specific methods have been developed to retrieve SINE insertions from species for which little DNA sequence information is available (Okada et al., 2004). Second, the short length of SINEs makes them easier to manipulate and genotype than longer mobile elements such as LINEs, which typically require multiple PCR amplifications to genotype a single locus (Sheen et al., 2000; Myers et al., 2002; Vincent et al., 2003).

MOBILE ELEMENTS IN PRIMATE PHYLOGENETIC STUDIES

Primate phylogeny and mobile element-based primate phylogenetic studies

The order Primates is one of the most diverse mammalian orders, exceeded in number of species only by Rodentia and Chiroptera. Phylogenetic studies of primates can be traced back to the beginning of the modern systematics, when Linnaeus recognized the order in 1766 and divided it into four divisions with 35 species. Over the last two and a half centuries, primate systematics has been studied intensively by a variety of means, including the fossil record, morphology, behavioral and ecological studies, and molecular and genetic analysis (Rasmussen and Nekaris, 1998; Ross et al., 1998; Goodman, 1999; Schneider, 2000; Yoder, 2003; Kay et al., 2004; Schmitz et al., 2005). Today, more than 300 primate species are recognized (Groves, 2001), with new species and even genera still being identified (Jones et al., 2005; Sinha et al., 2005; Davenport et al., 2006).

The current consensus view of primate phylogeny (Disotell, 2003) divides order Primates into two suborders, Strepsirrhini and Haplorhini. Strepsirrhini include the Lorisiformes (loris) and the Lemuriformes (lemurs). Haplorhini include the Tarsiformes (tarsiers) and Anthropoidea, which is further subdivided into the Platyrrhini (New World monkeys, NWMs) and the Catarrhini, composed of Cercopithecidae (Old World monkeys, OWMs) and Hominoidea (apes and humans). Despite great progress in primate systematics, a number of questions remain unanswered or debated. Examples include the monophyly of the order Primates, the aye-aye affiliation, the tarsier affiliation, NWM family relationships, and the human–chimpanzee–gorilla trichotomy, just to name a few. The development of mobile element-based genetic systems has provided opportunities to resolve many of these questions.

Alu elements, which are primate-specific SINEs, have been used extensively in primate phylogenetic studies and have been investigated in nearly the entire order (see Fig. 3). The human genome draft sequence has been invaluable in determining locations and flanking sequences of *Alu* insertions (Schmitz et al., 2001; Salem et al., 2003c; Ray et al., 2005a). In addition, the wealth of knowledge of *Alu* biology (e.g., subfamily phylogenetic distributions) (Batzer and Deininger, 2002) has greatly facilitated marker-mining in different primate genomes. In the following text, we briefly review recent progress in retrotransposon-based primate phylogenetic studies.

Primate monophyly

It is generally accepted that order Primates forms a monophyletic group, which belongs to cohort Archonta along with orders Scandentia (tree shrews), Dermoptera (flying lemurs or colugos), and Chiroptera (bats). However, several recent molecular studies based on mitochondrial sequence suggested the clustering of Dermoptera with anthropoid primates (Murphy et al., 2001; Arnason et al., 2002; Schmitz et al., 2002a). A clade comprising flying lemurs and anthropoids has therefore been

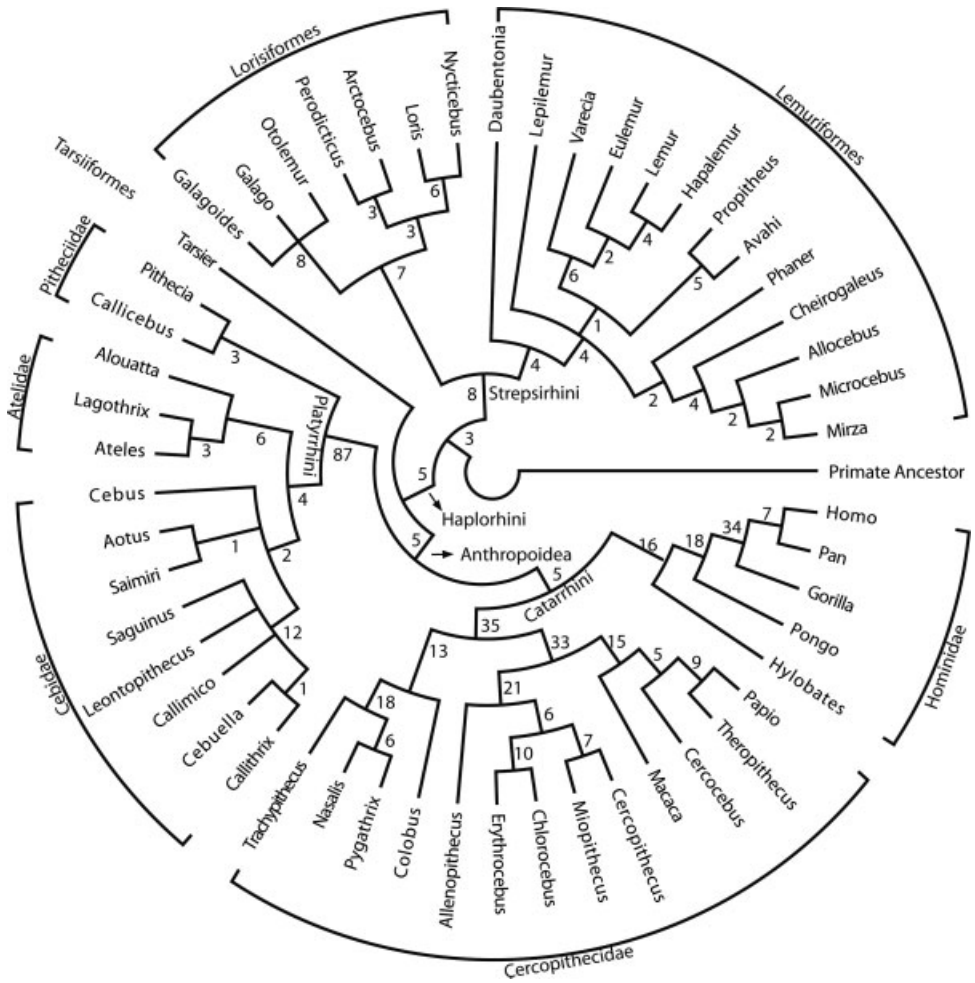

Fig. 3. A genus-level primate phylogenetic cladogram inferred from mobile element-based studies. We followed the nomenclature of Goodman et al. (1998). Numbers beside the branches indicate the number of insertions supporting each node.

proposed with the name Dermosimii (Arnason et al., 2002).

Mobile element insertions have been used to resolve the relationship between flying lemurs and primates (Schmitz et al., 2002a; Schmitz and Zischler, 2003). In these two studies, Schmitz et al. recovered three insertions present in the genomes of all tested primates but absent in the flying lemur genome. In addition, in a genome-wide analysis using *Alu* elements (present in all primates) as probes, no signal could be detected from the flying lemur genome, arguing against the inclusion of flying lemurs in order Primates. Therefore, it appears that a mitochondrial-specific nucleotide and amino acid composition bias may have caused confounding results in the earlier analysis (Schmitz et al., 2002a). These results demonstrate some of the advantages of retrotransposon-based systems in phylogenetic analysis.

Tarsier affiliation

Separated from other primates some 45 million years ago, tarsiers (genus *Tarsius*) have evolved a series of unique characters (Yoder, 2003). The affiliation of tarsiers is one of the most controversial issues in primate phylogeny and is fundamental for constructing primate infraorder relationships. Two different systems have been proposed depending on the tarsier affiliation: one divides order Primates into Prosimii (tarsiers, lemurs, and lorises) and Anthropoidea (monkeys and apes). Alternatively, if tarsiers are considered a sister clade to NWMs, OWMs, and Hominoids, then the Primate order can be divided into Strepsirrhini (lemurs and lorises) and Haplorhini (tarsiers, monkeys, and apes). Each of these two classifications is supported by a number of morphological studies. Analyses based on mitochondrial (Andrews et al., 1998; Schmitz et al., 2002b) and nuclear DNA (Goodman et al., 1998) were unable to resolve this question unambiguously [see (Yoder, 2003) for a review].

Four different SINE-based studies have yielded evidence supporting the Strepsirrhini and Haplorhini division. Zietkiewicz et al. (1999) determined the sequences of 37 *Alu* elements in the tarsier genome and compared them with the human and strepsirrhine *Alu* consensus sequence. This comparison indicated that tarsier *Alu* elements cluster with *Alu* subfamilies from the human lineage and support the sister relationship of tarsiers and anthropoids (monkeys and apes). In a second study, Schmitz et al. (2001) examined 118 human intronic *Alu* insertion loci. Three insertions were present in tarsier and anthropoids but absent in the strepsirrhines. In a recent review, Schmitz et al. (2005) provide evidence for another *Alu* insertion in the human serine palmitoyl transferase gene, which is shared between tarsier and anthropoids, further supporting the Strepsirrhini and Haplorhini division. In a fourth study, Kuryshev et al. (2001) discovered at least one unambiguous insertion

(FLAM-A) shared between tarsier and human but not strepsirrhines. Combined together, SINE insertion data provide some of the strongest evidence for the Strepsirrhini *versus* Haplorhini division.

Strepsirrhine phylogeny

Strepsirrhine primates are divided into two infraorders: Lemuriformes (lemurs) and Lorisiformes (loriformes) (Disotell, 2003), although some authors consider aye-aye (*Daubentonia madagascariensis*) as a separate infraorder, Chiromyiformes (Groves, 2001). The aye-aye is one of the most morphologically unique primates, and its relationship with other strepsirrhines is rather mysterious (Yoder, 1997). Besides the aye-aye's affiliation, a number of lower-level relationships among strepsirrhines are still under dispute.

In an attempt to resolve strepsirrhine phylogeny, Roos et al. (2004) examined SINE insertions in 20 strepsirrhine species. In a multiloci approach, two types of SINE insertions with known sequence were analyzed with Southern blots. Another 42 individual loci obtained from genomic sequence were genotyped with PCR, and a total of 61 SINE insertions were detected from these loci. With these SINE insertions, Roos et al. were able to infer a single phylogenetic tree with strong support. In addition, this tree was further supported by another 18 informative SINE insertion loci identified by Herke et al. (2007). Combined with the mitochondrial genome sequence, these results suggest a sister relationship of aye-aye and Lemuriformes and the monophyly of family Lorisidae. The inferred phylogeny supports the hypothesis that strepsirrhines originated in Africa and that Madagascar and Asia were colonized by single immigration events from Africa.

New World monkey phylogeny

NWMs (infraorder Platyrrhini) are generally considered a monophyletic group consisting of 15 to 16 genera. These genera have been divided into two (Cebidae and Atelidae) (Disotell, 2003), three (Pitheciidae, Cebidae, and Atelidae) (Goodman et al., 1998), or four families (Pitheciidae, Cebidae, Aotidae, and Atelidae) (Groves, 2001) by different authors, and the branching order of these families has not been determined with confidence (Schneider, 2000). Different topologies can be produced from the same dataset depending on the analytic method (Steiper and Ruvolo, 2003).

Two mobile element-based studies were performed to resolve family relationships.

In an earlier study (Singer et al., 2003), six informative elements were identified among 74 intronic *Alu* insertions. Besides three *Alu* insertions supporting the monophyly of Platyrrhini, one insertion grouped genus *Callithrix* and *Cebuella*, one insertion supported the monophyly of the subfamily callitrichinae, and one insertion suggested a close affiliation for *Aotus*, *Saimiri*, and *Cebus* to the callitrichine monkeys. Because of the relatively small number of SINE insertions examined in this study, the inferred relationships are suggestive and have limited statistical support.

A second study by Ray et al. (2005b) is more comprehensive: a total of 183 *Alu* elements that integrated into various NWM genomes were used in the analysis. Among them, 124 *Alu* elements are present in at least two NWM species but not in all species, making them informative in parsimony analysis (parsimony-informative markers). This analysis produced a single most parsimonious tree with a consistency index of 1.0. (The consistency index, which varies from 0 to 1, is a measure of how well the observed data fit the phylogenetic tree.)

This tree supports a sister relationship between family Atelidae (spider, woolly, and howler monkeys) and Cebidae (marmosets, tamarins, squirrel, capuchin, and owl monkeys). Family Pitheciidae (titi and saki monkeys) then joined this two-family clade to form the infraorder. This study represents the strongest evidence to date regarding NWM family relationships.

Old World monkey phylogeny

OWMs (family Cercopithecidae) can be divided into two distinct subfamilies: Cercopithecinae (cheek-pouched monkeys) and Colobinae (leaf-eating monkeys) (Delson, 1992; Groves, 2001; Disotell, 2003). The higher-level relationships (e.g., subfamily, tribe) inferred from molecular studies (Page et al., 1999; Page and Goodman, 2001) are generally congruent with classifications based on morphological data (Delson, 1992; Goodman et al., 1998). However, considerable disagreement still exists for some relationships at the genus level, especially for subfamily Colobinae (Disotell, 2000).

The OWM phylogeny has been the subject of two *Alu*-based analyses (Xing et al., 2005, 2007). In a family-wide survey, 285 phylogenetically informative *Alu* insertions were collected from sixteen OWM taxa. A phylogenetic hypothesis with strong statistical support was inferred from these insertion events. The resulting relationships among all major clades are in general agreement with other molecular and morphological data sets. In the Asian colobine clade, a sister relationship between *Nasalis* and *Pygrathrix* is strongly supported. This clade is subsequently joined by *Trachypithecus*, providing strong molecular evidence for the Asian colobine relationships.

The second study of OWMs focused on tribe Cercopithecini (guenons) (Xing et al., 2007). Two main phylogenetic questions concerning the Cercopithecini are: (1) whether terrestrial guenons (African green monkey, patas monkey, and L'hoesti's monkey) form a monophyletic group and therefore the current genus *Cercopithecus* is paraphyletic; (2) what relationships exist among the major groups within Cercopithecini. A total of 179 *Alu* insertions were identified and genotyped in eleven cercopithecine genomes. These insertions were used to construct a robust phylogeny, which supports the following relationships: (1) all terrestrial guenons cluster in a single clade, supporting a single transition to a terrestrial lifestyle and the paraphyly of current genus *Cercopithecus*; (2) *Allenopithecus* is the basal lineage of the tribe, and the split between the arboreal *Cercopithecus* group and *M. talapoin* happened after the divergence of the terrestrial group. Another interesting observation is that a number of loci supported different clustering patterns within the arboreal guenons (genus *Cercopithecus*), suggesting a rapid basal radiation of these species and possible interspecific hybridizations after their speciation.

Human, chimpanzee, and gorilla trichotomy

The relationship among human, chimpanzee, and gorilla (i.e., the trichotomy problem) has been a particularly difficult, long-standing problem (Patterson et al., 2006), and a number of studies have attempted to

address this question. Although the consensus approach (Satta et al., 2000; O'HUigin et al., 2002) and mtDNA sequences (Horai et al., 1995) support the chimpanzee as the nearest living relative of humans, the evidence is not overwhelming and some results supported alternative scenarios (Hasagawa and Kishino, 1991; Ruvolo et al., 1991; Chen and Li, 2001). For example, Satta et al. (2000) analyzed sequences from 45 nuclear loci and found that 60% of the loci support the human–chimpanzee clade, whereas the remaining 40% support the two alternatives equally.

Salem et al. (2003c) characterized 117 *Alu*Ye subfamily members along with 16 loci from other *Alu*Y subfamilies based on the human genome sequence. Using a total of 101 informative insertion loci, Salem et al. were able to infer a single most parsimonious tree with high levels of support (consistency index = 0.99; bootstrap support = 100% except one node). The resulting tree clearly clusters human and chimpanzee as a sister clade with gorilla as an out-group.

POPULATION GENETIC STUDIES

Just as retrotransposition of *Alu* and LINE-1 (L1) elements created an illuminating set of genetic markers for studies of primate evolution, their continued retrotransposition has generated a large set of more recent insertions that have not yet become fixed for their presence in the human population. These polymorphic *Alu* and L1 insertion loci share the same strengths that make loci with older, fixed insertions useful in phylogenetic analyses: they are essentially free of homoplasy caused by convergence or reversion, the ancestral state is known, and genotypes can be determined using simple procedures and widely available equipment.

As markers for studying human population history, polymorphic *Alu* insertion loci have additional desirable properties. Recent *Alu* insertions are distributed widely and almost randomly throughout the genome, the majority appear to be selectively neutral or nearly so (Cordaux et al., 2006a), and they do not appear to strongly influence local recombination rates (IHMC, 2005; Myers et al., 2006). Thus *Alu* polymorphisms should accurately reflect demographic forces that have affected the entire genome. L1 insertions have not been studied in as much detail as *Alu* insertions, but they also appear to exhibit these properties (Myers et al., 2002; Boissinot et al., 2004; Song and Boissinot, 2007). *Alu* and L1 retrotransposition has been frequent enough to generate a steady stream of insertions (Deininger and Batzer, 2002), so human population history has been marked by new insertions. Taken together, these accumulated insertion polymorphisms constitute a large data set of independent genetic markers widely dispersed throughout the genome and evolving under a common and simple evolutionary model. This is exactly the kind of data needed to make detailed and reliable statistical inferences about human demographic history.

Worldwide population genetic diversity and structure

The frequency distributions of *Alu* and L1 insertion alleles at polymorphic insertion loci are typical of neutral genetic polymorphisms in humans (IHMC, 2005; Witherspoon et al., 2006). However, it must be noted that the allele frequency distribution for any type of genetic marker is sensitive to the methods used to identify loci that are polymorphic.

The earliest *Alu* and L1 insertion polymorphisms were identified incidental to surveys of loci of medical interest (Economou-Pachnis and Tsichlis, 1985; Dombroski et al., 1991; Wallace et al., 1991; Holmes et al., 1994). With the advent of the Human Genome Project, it became possible to quickly and efficiently find *Alu* and L1 insertions and identify the most recent ones by virtue of their high sequence similarity to the reconstructed consensus sequences of the presumed "master" progenitor elements from which all these elements are generated. This is the first step in ascertaining these elements. Insertions whose sequences identify them as members of inactive *Alu* or L1 families are almost certainly so old that they have become fixed in the human population, so these—the vast majority of insertions—can be ignored. The much more manageable subset of loci that are likely to be recent insertions (and thus polymorphic for presence or absence) can then, in the second ascertainment step, be screened for polymorphism in panels of varying size and population diversity (Carroll et al., 2001; Myers et al., 2002; Salem et al., 2003b; Xing et al., 2003; Carter et al., 2004).

The initial identification step biases the data set toward insertions that have reached a high frequency in the population sample from which the public human genome sequence was obtained, and the second step introduces a modest bias toward insertions of intermediate frequency in the populations from which the screening panels were sampled. To eliminate these biases, molecular techniques have been devised specifically to find previously unknown low-frequency *Alu* and L1 polymorphisms (Boissinot et al., 2004). These techniques rely on the known sequences of *Alu* and L1 elements. The frequency distribution of the combined set of insertion polymorphisms ascertained by these different methods should approximate the distribution that would be observed with complete ascertainment of polymorphic loci in the largest ascertainment panels used.

Ascertainment biases can limit the power of genetic data sets to make inferences regarding demographic history. However, polymorphic loci with common minor alleles (>5%; "common polymorphisms" hereafter) are ideally suited for analyses of intraspecific population structure. For such questions, common polymorphisms are more informative than rare polymorphisms (Bamshad et al., 2003; Witherspoon et al., 2006). Furthermore, under the assumption that most *Alu* and L1 polymorphisms are selectively neutral, intermediate-frequency insertion-present and insertion-absent (empty) alleles are treated identically by genetic drift and demographic forces. The alleles are therefore interchangeable for purposes of population structure analyses that rely on the signal of genetic drift, so a slight bias toward loci with higher frequencies of the insertion allele does not affect the result.

Alu and L1 insertion polymorphisms have been used to illuminate the patterns and distribution of human genetic diversity in many studies. In such studies, a set of common insertion polymorphisms is chosen, and genotypes at those loci are obtained for individuals sampled from diverse populations around the world. Within a population, these loci seldom deviate from Hardy–Weinberg equilibrium, underscoring their utility for population genetic analysis (Watkins et al., 2001, 2003).

The proportion of variance in allele frequencies that is due to differences between populations, F_{ST}, depends on the choice of populations sampled and how they are grouped. F_{ST} estimates obtained from a set of *Alu* insertion genotypes range from 8 to 14%, for ~30 regional populations and three major continental groups (African, East Asian, and European), respectively (Watkins et al., 2001, 2003). F_{ST} estimates from a set of L1 insertion genotypes are similar (Sheen et al., 2000; Witherspoon et al., 2006). While relatively small, these F_{ST} values differ significantly from zero. Thus, although most genetic variation is due to differences between individuals within populations, the global human population has been structured rather than panmictic. Both *Alu* and L1 polymorphisms show higher average gene diversity (the proportion of individuals that are heterozygous at a locus, averaged across loci) in African populations (Watkins et al., 2001). This difference implies that African populations have maintained larger effective population sizes than non-African populations during recent evolutionary history (Stoneking et al., 1997).

These observations of nonzero F_{ST} and differences in genetic diversity between African and non-African populations imply that our genomes bear traces of past population migrations, geographic isolation, population fusion and fission, and changing population sizes. As early as 1994, Batzer et al. (1994) genotyped a set of four polymorphic *Alu* insertion loci in 664 individuals from 16 diverse populations. The four insertion alleles are found at different frequencies in each population, and this information can be used to compute the genetic distance (a measure of genetic dissimilarity) between the populations. Despite the small number of loci, a clear pattern emerged: populations that were geographically nearer to each other were genetically more similar. This positive correlation between genetic and geographic distance was confirmed and explored in greater detail with larger sets of *Alu* polymorphisms (Watkins et al., 2001, 2003) and corroborated with L1 polymorphisms (Witherspoon et al., 2006). These analyses showed that most of the correlation between genetic and geographic distance is due to the large genetic and geographic distances between African and non-African populations. While geographic distances between European, East Asian, and Indian populations vary over a wide range, their genetic distances are relatively more similar.

The pattern of genetic relationships between populations can be seen in a tree of populations constructed from the allele frequencies (see Fig. 4). The non-African populations cluster together and are separated from the African populations by a long branch. While a correlation between geographic and genetic location is apparent among non-African populations (European, Indian, and East Asian populations do form clusters), the genetic distances among all of these populations are relatively small. For example, the Indian populations are roughly the same geographic distance from the European and African populations, but the Indian populations are genetically much closer to the European and East Asian populations than to the African populations [see also (Watkins et al., 2003)]. Principal components analyses of *Alu* and L1 data in these populations yield qualitatively similar results. In particular, the first principal component, which typically accounts for most of the variance in the data, separates sub-Saharan African populations from the others (Watkins et al., 2003; DJW, unpublished results). The second principal component typically distinguishes East Asian from European populations.

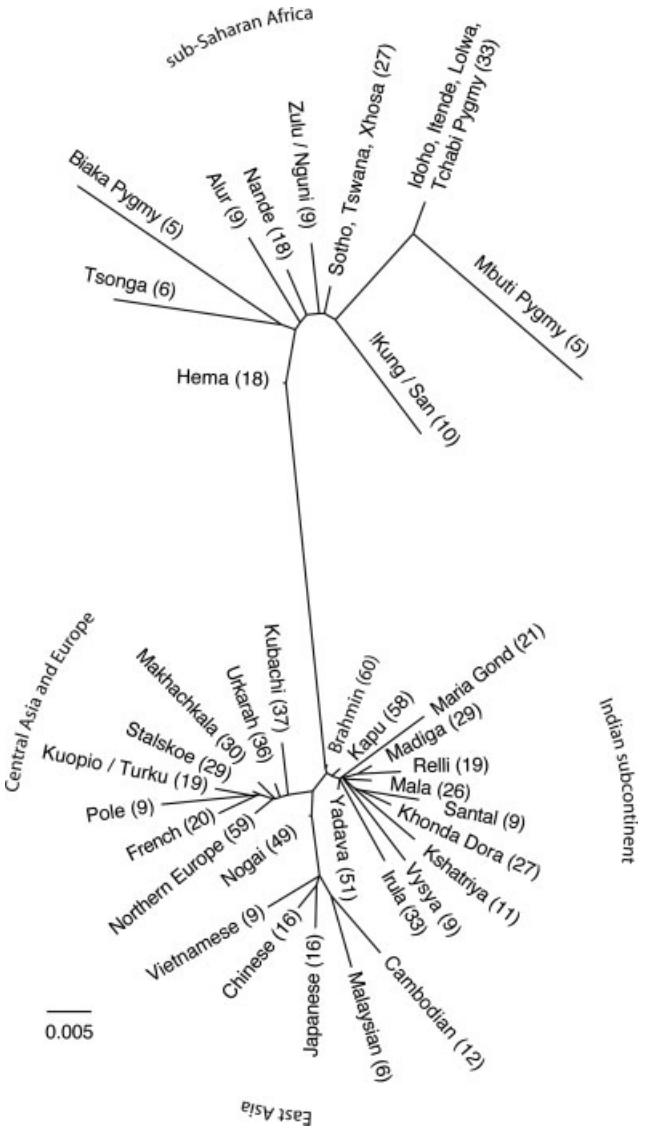

Fig. 4. Tree of 36 populations from Europe, sub-Saharan Africa, central Eurasia (Daghestan), southern Asia (India), and southeastern and eastern Asia. A maximum likelihood model of genetic drift (CONTML, PHYLIP 3.65, best tree found using global and local rearrangements, starting from 100 random taxon addition orders) was used to infer the tree, based on the frequencies of 100 *Alu* insertions genotyped in 840 individuals (data from Watkins et al., 2003; Bulyaeva et al., 2003).

Recent African origins hypothesis

These findings support the recent African origins (RAO) hypothesis. This hypothesis states that the most recent population that was ancestral to all modern humans arose in Africa during the late Paleolithic. Migrants from that population then colonized the rest of the world less than 100,000 years ago, and, in the strict version of the RAO, replaced premodern humans without significant genetic mixing (Stringer and Andrews, 1988; Tishkoff and Verrelli, 2003).

Under the RAO hypothesis, anatomically modern human populations in Africa evolved in place, accumulating a substantial level of genetic diversity. At some later time, the rest of the world was colonized by a small

number of migrants drawn from the northeastern region of the African population. These migrants represented only a subset of the genetic diversity of the parent population, and still more of that diversity was lost due to genetic drift in the small populations that emigrated and spread rapidly around the world. The RAO hypothesis explains the higher genetic diversity in African *versus* non-African populations as the consequence of this "bottleneck" in population size, in contrast with the larger population size maintained by the African population. The long branch separating African from non-African populations likewise reflects the genetic drift that would have occurred in the small emigrating population. The relatively high genetic similarity among non-African populations (despite their wide geographic distribution) is attributed to their relatively recent founding. The older African subpopulations show greater genetic differentiation (see Fig. 4, longer branches in African cluster) because they have been drifting apart for much longer.

The frequency of *Alu* and L1 insertions is lower, on average, in African populations than in non-African ones (Watkins et al., 2003; Witherspoon et al., 2006). Since the ancestral state at these loci is known to be the absence of the insertion, an ancestral population must have existed that lacked these insertions entirely. That population would be more similar to African populations than to non-African groups, which is consistent with the RAO scenario above. However, at least part of this frequency difference may be due to the ascertainment bias for *Alu* and L1 insertions that have reached moderate to high frequency in non-African populations (Watkins et al., 2001; Witherspoon et al., 2006). Moreover, the relationship between the frequencies of ancestral alleles and ancestral populations is not simple and depends in part on mutation rates (Rogers et al., 2007). Therefore, this observation must be considered with caution.

Classification

Human population history can be inferred from the patterns of human genetic diversity only because the human gene pool is not thoroughly mixed. It is subtly structured, and that structure is a result of our complex demographic history, from our ancestral African roots through the most recent population fissions, fusions, and migrations. Since cultural, linguistic, and ethnic traits are also transmitted across generations, they are often correlated with the subtle genetic differences between groups. In particular, various categorizations of humans into "races" have been based on cultural and phenotypic traits; so they are correlated to some extent with detectable population genetic structure. This has raised the question whether genetic data can be used to reliably classify individuals into groups that correspond with self-identified ancestry or ethnicity. This question is of particular importance in biomedical research, since undetected population structure within a research population can lead to spurious results (Marchini et al., 2004).

Genetic structure within a population will manifest itself as departures from the patterns expected in a homogeneous population—e.g., violations of Hardy–Weinberg equilibrium and unexpected associations between alleles at independent or loosely-linked loci. The method implemented in the program *Structure* (Pritchard et al., 2000) assigns individuals to subpopulations in a way that minimizes those violations within subpopulations. The method thus identifies and classifies genetically related groups of individuals within a sample population. Romualdi et al. (2002) applied this method to polymorphic *Alu* insertion loci in humans. They concluded that the classification of individuals into populations was error-prone and that the continental population groups defined *a priori* (Africa, Asia, Europe, Australia and the Americas) were genetically indistinct. However, only 21 polymorphic *Alu* loci were analyzed. Subsequent analyses using the same methods and 100 polymorphic *Alu* loci (Bamshad et al., 2003) or 75 polymorphic L1 loci (Witherspoon et al., 2006) found clear evidence of continental population structure, but only when at least 50 loci were used. Data sets consisting of fewer loci (unless selected initially for large inter-population differences) typically lack the statistical power to detect the level of population structure that is present in continental human populations (Rosenberg et al., 2005).

The reliability with which individuals can be assigned into groups that correspond to their populations of origin depends not only on the number of loci, but also on the populations sampled. If relatively distinct population groups are sampled (e.g., sub-Saharan Africans, East Asians, and Northern Europeans), 100 polymorphic *Alu* loci provide enough information to classify individuals into those three categories with nearly 100% accuracy (Bamshad et al., 2003). However, if more geographically intermediate populations are added (e.g., from the Indian subcontinent), the accuracy of classification into these four *a priori* groups suffers, even with 175 *Alu* and L1 loci combined (Witherspoon et al., 2006). This is due in part to a failure of the groupings to reflect the actual genetic structure in the data, and partly due to the challenge inherent in making finer distinctions between more similar individuals.

Figure 5 shows the results of a *Structure* analysis based on 175 *Alu* and L1 loci reported earlier (Watkins et al., 2003; Witherspoon et al., 2006). The optimal number of populations required to explain the population structure in the data is five, instead of the four that might have been presumed *a priori* (i.e., European, East Asian, sub-Saharan African, and Indian subcontinental). To adequately model the unusual differentiation of the Irula noncaste population of India requires a fifth population. However, other subpopulations that might seem equally unique (e.g., the Khonda Dora, another noncaste Indian population; or the group of Pygmy populations) are not modeled as separate populations. This result is similar to that of Rosenberg et al. (2002), in which a small isolated Pakistani population, the Kalash, formed a distinct group in a *Structure* analysis.

The block-like structure pattern in Figure 5 might suggest that human populations are discrete units, like islands: homogeneous within their boundaries and clearly distinguishable from their neighbors. To some extent, this appearance is due to the lack of geographically intermediate populations in this data set. An exhaustive sampling of world populations would almost certainly show a smoother gradation between neighboring populations (Witherspoon et al., 2006). However, even the apparently distinct genetic differences between the groups in Figure 5 do not imply that there are reliable categorical differences between individuals from different groups. Using the same set of 175 *Alu* and L1 insertions genotyped in the same individuals, Witherspoon et al. (2007) showed that two individuals sampled from different continental regions (European, sub-Saharan

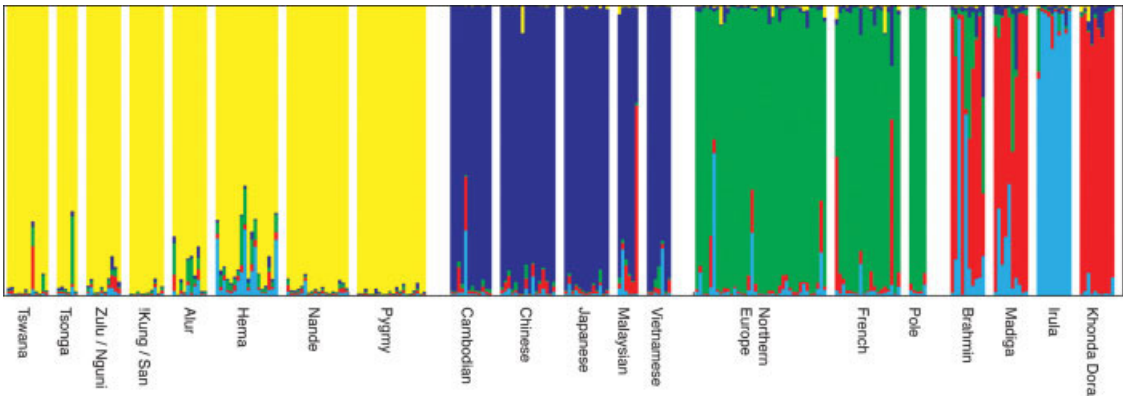

Fig. 5. Structure plot. Estimates of the proportion of each individual's ancestry that is derived from each of five inferred ancestral populations, using *Structure* (Pritchard et al. 2000; correlated allele frequencies and admixture between populations allowed). Each vertical stripe represents a single individual, and the different colors within each strip represent fractions of that individual's ancestry in five color-coded ancestral populations. Individuals are grouped according to the populations and continental regions from which they were sampled (populations set apart by narrow blank columns; continents are set off by larger gaps). The analysis is based on genotypes at 100 *Alu* and 75 L1 loci in 260 individuals from 20 population groups (see text).

African, East Asian, and Indian subcontinental) were often (25% frequency) genetically more similar than a pair drawn from within the same region. The number of loci used to compute genetic similarity (equivalently, relatedness) is critical: as more polymorphic loci are used, it becomes less likely that an individual will resemble a member of another population more than a member of his or her own population (see Fig. 6). Even with thousands of loci, however, that probability does not decrease to zero. When dealing with phenotypes whose genetic variations are determined by fewer than 20 loci, group labels are likely to be poor predictors of individual phenotypes (Witherspoon et al., 2007).

Analyzing individual human populations and their histories

As more genetic data are collected, ever-finer details of human population history can be resolved. The genetic effects of great distances and geographic barriers were the first to be detected, but more subtle barriers to gene flow exist, such as linguistic and social boundaries. In theory, a strong enough social barrier between populations, maintained long enough, could leave a genetic trace. This appears to be the case in some populations of India, where the caste system restricted intermarriage between groups for thousands of years. Although alleles and haplotypes are widely shared among castes, small genetic differences do exist, and these can be detected with sufficiently large data sets (Majumder et al., 1999; Bamshad et al., 2001). These differences probably reflect a combination of social strictures and migration of West Eurasian peoples into India around the time that the caste system arose.

Alu insertion polymorphisms have been used in many other samples and populations to test hypotheses concerning migrations and the sources of modern populations (Batzer et al., 1996; Sherry et al., 1997; Stoneking et al., 1997; Novick et al., 1998; Majumder et al., 1999; Comas et al., 2000; Bamshad et al., 2001; de Pancorbo et al., 2001; Nasidze et al., 2001; Roy-Engel et al., 2001; Antunez-de-Mayolo et al., 2002; Romualdi et al., 2002; Xiao et al., 2002; Bamshad et al., 2003; Bulayeva et al., 2003; Gonzalez-Perez et al., 2003; Mastana et al., 2003; Vishwanathan et al., 2003; Watkins et al., 2003; Barbalic et al., 2004; Comas et al., 2004; Cotrim et al., 2004; Dornelles et al., 2004; Maca-Meyer et al., 2004; Mansoor et al., 2004; Dunn et al., 2005; Mateus Pereira et al., 2005; Terreros et al., 2005; Herrera et al., 2007). In general, the results of studies of mobile element variation are in agreement with those based on other genetic systems, including mitochondrial (Budowle et al., 1999; Jorde et al., 2000; Bamshad et al., 2001), Y-chromosome (Jorde et al., 2000; Bamshad et al., 2001), microsatellite (Jorde et al., 1997; Bamshad et al., 2003), short tandem repeats (STR) (Shriver et al., 1997; Jorde et al., 2000; Budowle et al., 2001; Butler et al., 2003), and single nucleotide polymorphisms (SNPs) (Sachidanandam et al., 2001; Matise et al., 2003; Reich et al., 2003; Salisbury et al., 2003).

FORENSIC APPLICATIONS

The application of DNA to problems in forensics is an active field of research. For example, the Federal Bureau of Investigation's Combined DNA Index System (CODIS) is a standard set of 13 STR loci that is routinely employed to match a suspect's genetic material to the DNA found at a crime scene. Early in an investigation, before any definite suspects have been confirmed, however, tools that narrow the potential pool would be desirable. One method to reduce the list of possible suspects is to infer the ancestral origin of a DNA specimen found at a crime scene.

Inferring human geographic origins

A specific application to the forensic problem of determining ancestral origin of an unknown DNA sample was recently the subject of a study by Ray et al. (2005a). As background for this study, 710 individuals representing 31 different populations across four continents were evaluated for each of the 100 *Alu* insertion polymorphisms (Watkins et al., 2003). This database of genetic variation was then applied to determine the ancestry of 18 anonymous DNA samples culled from two forensic laboratories. The 18 individuals were genotyped at the same 100 *Alu* insertion loci, and the inferred geographic affiliation of the samples was calculated using the

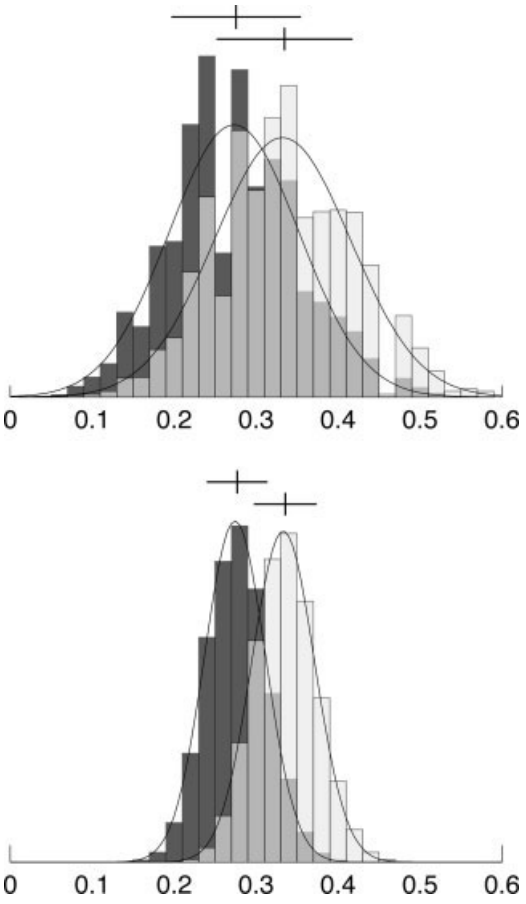

Fig. 6. Distributions of genetic distances between pairs of individuals from the same population (dark gray histograms on the left) or from different populations (white histograms on the right); overlap regions shown in gray. For simplicity, individuals are grouped into just two populations: 104 sub-Saharan African individuals and 156 European, east Asian, or southeast Asian individuals. The genetic distances are computed by the "allele-sharing" method, equivalent to one minus the average number of alleles that the two individuals share over the 175 *Alu* and L1 loci analyzed. In the upper panel, only 20 loci are used (the histograms are averaged over many resamplings of 20 loci at a time). With that number of loci, the probability that a two individuals from different populations are genetically more similar than two individuals from the same population is ~0.29. That probability decreases to 0.13 in the lower panel, where all 175 loci are used. The means and standard deviations of the distributions are shown as vertical ticks and horizontal lines above them, and the corresponding normal probability functions are superimposed (black curves). Data and methods from Witherspoon et al. (2007).

Structure software package (Pritchard et al., 2000; Falush et al., 2003).

After comparing the genotypes of the 18 anonymous individuals to the database of known geographic affiliations, Ray and coworkers were able to correctly infer the likely ancestry of all 18 individuals with regard to four major human population groups (sub-Saharan African, western European, Indian, and east Asian). Restricting the reference database to such well-defined groups allowed them to avoid the difficulties inherent in analyzing intermediate populations, as described above. This data set represents a powerful tool to ascertain the inferred geographic ancestry of unknown human DNA samples in a forensic analysis. The methodology is also technically simple, requiring only standard thermal cyclers for PCR and the ability to run and visualize agarose gel electrophoresis.

Because many recent *Alu* insertions will be polymorphic for their presence or absence and have variable insertion frequencies among world populations (Watkins et al., 2003), the presence of these population-indicative insertions allows for hierarchical testing of subgroup affiliations. For example, once the initial *Structure* analysis assigns the DNA specimen to a continental affiliation such as Europe, Africa, Asia, or India, subsequent analyses, using only insertion loci that are useful within the subcontinental population groups, can be used to further isolate the subcontinental region. Identification of these subpopulation-specific loci is a difficult undertaking. Fortunately, a recent study (Cordaux et al., 2007) demonstrated that population-indicative *Alu* insertions can be identified preferentially. We believe that the application of this targeted identification will result in the rapid identification of additional markers that are likely to be more informative for specific population groups. Incorporation of these loci will reduce the number of loci needed for accurate identification of unknown samples.

Mobile element-based human sex identification

Another potential forensic application of *Alu* elements is human sex identification (Hedges et al., 2003). Determination of sex from human DNA samples is a common procedure in forensic laboratories. The most widely used approach is based on the Amelogenin locus (one of the standard CODIS loci), which yields different-sized PCR amplicons for the X and Y chromosome versions of the Amelogenin gene (Sullivan et al., 1993). However, this method can misidentify males as females in some cases due to a deletion in the AMEL Y region that reaches an allele frequency as high as 8% in some populations (Santos et al., 1998; Steinlechner et al., 2002; Thangaraj et al., 2002). Any source of error is a legitimate cause for concern. In fact, several researchers recommend that Amelogenin should not be relied upon as the sole indicator of sex (Santos et al., 1998; Brinkmann, 2002; Steinlechner et al., 2002; Thangaraj et al., 2002).

Fixed *Alu* insertions on either the X or the Y chromosome provide an alternative method of identifying the sex chromosomes present in a sample. Hedges et al. (2003) discovered two loci, *AluSTXa* and *AluSTYa*, that consistently differ with regard to insertion presence/absence on the X and Y chromosomes. They examined 778 human DNA samples from diverse populations and achieved 100% accuracy. The power of this test over the standard Amelogenin test derives from the statistical improbability of two independent mutations occurring in two separate genomic locations. Thus, the combination of these two markers (with or without the standard Amelogenin test) provides added assurance that sex test results are accurate.

Human DNA identification and quantitation

Another forensic task that must be considered is the identification and quantitation of human DNA in a complex mix (Chakraborty et al., 1999). Because of their high copy number, *Alu* elements represent a naturally amplified marker for the identification of primate and

human-specific DNA. Several systems have been developed to identify and quantify human DNA using PCR primers designed to amplify the core *Alu* sequence (Sifis et al., 2002; Nicklas and Buel, 2003a,b,c, 2005). During intra-*Alu* PCR, primers are designed within the core body of the element to amplify multiple target copies of the *Alu* family. Most of these systems are specific to the primate lineage and are not necessarily human-specific.

Intra-*Alu*-based PCR assays have been developed for human-specific DNA identification and quantitation based on the Yb8 and Yd6 subfamilies of *Alu* (Walker et al., 2003b). These techniques can detect and quantify human DNA ranging from 10 to 0.001 ng, even in samples that involve mixtures with nonhuman contaminants (Walker et al., 2003a).

Primate identification

The application of forensic techniques is not limited to human-specific problems. The identification of illegally taken wildlife and determination of management units in conservation efforts are important tasks that rely on similar methodologies. Just as humans have their own specific subfamilies of *Alu* elements, other primate lineages do as well. New World primates, for example, harbor *Alu* subfamilies that are distinct from all other primates (Ray and Batzer, 2005). Thus, any primate DNA could potentially be identified and quantified using the same techniques as long as one knows the *Alu* subfamilies in question.

Techniques that lead in this direction have already been developed. For example, in laboratories dealing with primate genetics, it is critical that researchers ascertain that they are handling DNA from the appropriate taxa. Very often researchers collect or receive DNA that was collected in a noninvasive manner (i.e., "divorced" tissues such as hair or feces) (Kohn et al., 1995; Taberlet et al., 1997; Matsubara et al., 2005). This is especially common in investigations of the illegal wildlife trade and identification of seized products (Matsubara et al., 2005; Yan et al., 2005; Domingo-Roura et al., 2006). Even when laboratories deal in their own "homegrown" DNA via cell culture, cross-contamination can occur among cell cultures and within concurrent large-scale DNA extractions from multiple species. Furthermore, simple mishandling of well-documented samples may lead to mislabeling or cross-contamination.

Just as taxonomists and museum curators have long relied on detailed analysis of morphological characters in the form of dichotomous keys to identify unknown samples, molecular biologists can now develop their own molecular keys to do the same. Previously, mitochondrial haplotypes had been used as a method for species identification (Kohn et al., 1995; Palumbi and Cipriano, 1998; Hebert et al., 2004). Unfortunately, complications involving maternal inheritance, primer failure, and nuclear translocation of mitochondrial genes have limited the applicability of this approach (Moritz and Cicero, 2004; Lorenz et al., 2005).

Under extreme situations (e.g., highly degraded, extremely small quantities of DNA samples), high copy-number mitochondrial DNA might be the only option for forensic studies. However, several factors make a SINE-based dichotomous key for identifying samples attractive for most primate forensic studies. First, genotyping SINEs require only simple equipment to which even modest molecular biology laboratories have access. Second, the technique is reliable even in the hands of a novice because no special technical expertise is required. Third, unlike plastic morphological characters or potentially confusing sequence data, this type of SINE-based data consists solely of presence–absence states. Such data are easy to interpret in a hierarchical fashion to rapidly narrow the possibilities regarding the identity of anonymous DNA samples. Fourth, in contrast to mitochondrial DNA methods, "universal" primers are used in a simple PCR assay to amplify loci from most relevant taxa. Finally, unlike "single locus" (e.g., mainly mitochondrial DNA) systems, SINE-based dichotomous keys identify terminal branches based on multiple unlinked loci instead of the single linkage group exemplified by the maternally inherited mitochondrial genome. One such SINE-based dichotomous key was recently developed for a variety of primates (Herke et al., 2007).

THE IMPACT OF MOBILE ELEMENTS ON PRIMATE GENOME EVOLUTION

Although initially considered "junk DNA," mobile elements have proven to have profoundly influenced primate genomic evolution. In this section, we will review several aspects of mobile elements' impact.

Mobile elements and human diseases

Mobile elements can cause human diseases both during and after their integration (Deininger and Batzer, 1999; Batzer and Deininger, 2002; Callinan and Batzer, 2006; Chen et al., 2006). The insertion of mobile elements can directly disrupt genes or alter gene expression and thus lead to disease. After mobile-element integration, nonhomologous recombination between similar elements can generate disease-causing deletions. The amplification rate of mobile elements has varied dramatically during the primate evolution leading to the human lineage (Hedges et al., 2004, 2005). It is estimated that the current rate of *Alu* amplification in the human genome is of the order of one *Alu* insertion in every 200 births (Deininger and Batzer, 1999). To date, alterations caused by mobile elements have been identified in more than 100 human genetic disorders, including cases of hemophilia type A and B, familial breast cancer, and neurofibromatosis type 1 [see (Deininger and Batzer, 1999; Batzer and Deininger, 2002; Callinan and Batzer, 2006; Chen et al., 2006) for reviews].

Mobile element exonization and mobile element-mediated exon shuffling

One of the most direct impacts of mobile elements is their contribution to the formation of new genes and gene families. Some mobile elements contribute new genetic material that eventually becomes functional in itself or becomes part of a larger gene via a process termed "exonization" (Lev-Maor et al., 2003). Examples include: (1) the human *SETMAR* gene, formed from the fusion of a domain that exhibits histone methyltransferase activity (the SET domain) and the entire transposase-coding region of a Hsmar1 transposon (the MAR domain) (Cordaux et al., 2006b) and (2) the human BC200 gene, which is derived from an *Alu* element and encodes a brain-specific nonmessenger RNA (Kuryshev et al., 2001). Many additional examples have also been reported (Sorek et al., 2002; Lorenc and Makalowski, 2003; Krull et al., 2005).

Another mechanism by which mobile elements contribute to gene creation is termed sequence transduction or 3′ transduction (Holmes et al., 1994). During this process, genomic materials downstream from a retrotransposon are transcribed and mobilized along with the retrotransposon. This may result in the duplication of exons to the new location or the alteration of gene expression at a new location through the movement of promoter and enhancer sequences. This process, which was initially reported in cell culture retrotransposition assays (Moran et al., 1999) and subsequently in the human genome, leads to duplication of exons and even whole genes (Rozmahel et al., 1997; Ejima and Yang, 2003; Xing et al., 2006).

Mobile elements and primate genomic structural variation

In addition to their direct impact on gene expression, mobile elements also represent one of the important sources of structural genomic variation. Retrotransposon insertions have contributed to a 15–20% expansion of the human genome compared to strepsirrhine genomes (Liu et al., 2003). Several other mechanisms by which mobile elements shape primate genomes have been reported recently, including retrotransposon insertion-mediated deletion in the genome (Callinan et al., 2005; Han et al., 2005), mobile element-mediated nonhomologous recombination (Sen et al., 2006), and mobile element-mediated gene conversion (Kass et al., 1995; Roy et al., 2000; Salem et al., 2003a; Vincent et al., 2003; Otieno et al., 2004; Zhi, 2007).

Functional role of "junk DNA"

Whether seemingly "selfish" mobile elements possess any function has been a fundamental question since mobile elements were first recognized as a major component of mammalian genomes. Their prevalence in eukaryotic genomes suggests that mobile elements may be, on the whole, selectively advantageous, and several studies suggest that they play important roles in genome structure and gene expression (Brosius and Gould, 1992; Vidal et al., 1993; Schmid, 1998; Hamdi et al., 2000; Deininger and Roy-Engel, 2002; Labrador and Corces, 2002; Nouaud et al., 2003). One of the more compelling examples is the involvement of L1 elements in the human X chromosome inactivation process [Lyon's "repeat hypothesis" (Lyon, 1998)]. L1 elements compose ~30% of the human X chromosome, and recent studies suggest that L1 elements may function as "way stations" for X inactivation signals (Bailey et al., 2000; Hansen, 2003; Ross et al., 2005). Other roles that have been proposed for mobile elements include double-strand DNA break repair (Morrish et al., 2002), the creation of new CpG islands (short stretches of DNA, which have a high frequency of C-G dinucleotide sequences and are usually found near the promoter regions of genes) (Schmid, 1998), and the stimulation of translation during cell stress and viral infection through the regulation of double-stranded RNA-activated protein kinase (Chu et al., 1998). However, the putative functional role of mobile elements as a whole is still a topic of intense debate.

CONCLUSIONS

As this review has shown, mobile elements are proving highly useful in answering questions about primate phylogeny, human population affinities, and forensics. In addition, because of their unique mechanistic properties and ubiquity in the genome, mobile elements exert important effects on recombination, deletion and duplication, DNA sequence variation, and gene expression.

Many taxa, including a number of nonhuman primates, have not yet been characterized for mobile element variation. Such studies will improve the potential to resolve thorny phylogenetic questions, and they can be used to address issues about species conservation and identification. Another question that can be addressed by mobile elements is the relationships of primates with other mammals (i.e., the superordinal taxonomic position of primates).

In humans, a reference set of 100 *Alu* insertion polymorphisms has been established and typed in more than 700 individuals (Watkins et al., 2003). An additional reference set of 75 L1 insertion polymorphisms has also been compiled and analyzed (Witherspoon et al., 2006). Additional human populations can now be genotyped for these polymorphisms in comparative studies of human population origins and history. All of these efforts will be enhanced by new technical developments, such as microarray-based analysis of mobile elements. Clearly, much remains to be learned about the evolution and consequences of this fascinating component of our genome.

ACKNOWLEDGMENTS

The authors thank the editor and the three anonymous reviewers for their constructive and valuable comments. In addition, valuable comments and assistance were provided by Liz Marchani, Alan Rogers, Kazima Bulayeva, and W. Scott Watkins.

LITERATURE CITED

Andrews TD, Jermiin LS, Easteal S. 1998. Accelerated evolution of cytochrome *b* in simian primates: adaptive evolution in concert with other mitochondrial proteins? J Mol Evol 47:249–257.

Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, Yunis JJ, Yunis EJ, Damodaran C, Martinez de Pancorbo M, Caeiro JL, Puzyrev VP, Herrera RJ. 2002. Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23:3346–3356.

Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc Natl Acad Sci USA 99:8151–8156.

Bailey JA, Carrel L, Chakravarti A, Eichler EE. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc Natl Acad Sci USA 97:6634–6639.

Bamshad M, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BV, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, Jorde LB. 2001. Genetic evidence on the origins of Indian caste populations. Genome Res 11:994–1004.

Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. 2003. Human population genetic structure and inference of group membership. Am J Hum Genet 72:578–589.

Barbalic M, Pericic M, Skaric-Juric T, Narancic NS. 2004. Ace Alu insertion polymorphism in Croatia and its isolates. Coll Antropol 28:603–610.

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ,

Deininger PL, Stoneking M. 1996. Genetic variation of recent Alu insertions in human populations. J Mol Evol 42:22–29.
Batzer MA, Deininger PL. 1991. A human-specific subfamily of Alu sequences. Genomics 9:481–487.
Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379.
Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL. 1991. Amplification dynamics of human-specific (HS) Alu family members. Nucleic Acids Res 19:3619–3623.
Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL. 1994. African origin of human-specific polymorphic Alu insertions. Proc Natl Acad Sci USA 91:12288–12292.
Boissinot S, Entezam A, Young L, Munson PJ, Furano AV. 2004. The insertional history of an active family of L1 retrotransposons in humans. Genome Res 14:1221–1231.
Brinkmann B. 2002. Is the amelogenin sex test valid? Int J Legal Med 116:63.
Brosius J, Gould SJ. 1992. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc Natl Acad Sci USA 89:10706–10710.
Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci USA 100:5280–5285.
Budowle B, Shea B, Niezgoda S, Chakraborty R. 2001. CODIS STR loci data from 41 sample populations. J Forensic Sci 46:453–489.
Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL. 1999. Mitochondrial DNA regions HVI and HVII population data. Forensic Sci Int 103:23–35.
Bulayeva K, Jorde LB, Ostler C, Watkins S, Bulayev O, Harpending H. 2003. Genetics and population history of Caucasus populations. Hum Biol 75:837–853.
Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. 2003. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 48:908–911.
Callinan PA, Batzer MA. 2006. Transposable elements and human disease In: Volff JN, editor. Genome dynamics. Basel: Karger. p 104–115.
Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. 2005. Alu retrotransposition-mediated deletion. J Mol Biol 348:791–800.
Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde LB, Deininger PL, Batzer MA. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J Mol Biol 311:17–40.
Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, Walker JA, Watkins WS, Jorde LB, Batzer MA. 2004. Genome-wide analysis of the human Alu Yb-lineage. Hum Genomics 1:167–178.
Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. 1999. The utility of short tandem repeat loci beyond human identification: implications for development of new DNA typing systems. Electrophoresis 20:1682–1696.
Chen FC, Li WH. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. Am J Hum Genet 68:444–456.
Chen JM, Ferec C, Cooper DN. 2006. LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease: mutation detection bias and multiple mechanisms of target gene disruption. J Biomed Biotechnol 2006:56182.
Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW. 1998. Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol Cell Biol 18:58–68.
Churakov G, Smit AF, Brosius J, Schmitz J. 2005. A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). Mol Biol Evol 22:886–893.
Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, Sajantila A. 2000. Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. Hum Genet 107:312–319.
Comas D, Schmid H, Braeuer S, Flaiz C, Busquets A, Calafell F, Bertranpetit J, Scheil HG, Huckenbeck W, Efremovska L, Schmidt H. 2004. Alu insertion polymorphisms in the Balkans and the origins of the Aromuns. Ann Hum Genet 68:120–127.
Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH Jr. 2005. Two independent retrotransposon insertions at the same site within the coding region of BTK. Hum Mutat 25:324–325.
Cordaux R, Lee J, Dinoso L, Batzer MA. 2006a. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene 373:138–144.
Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA. 2007. In search of polymorphic Alu insertions with restricted geographic distributions. Genomics 90:154–158.
Cordaux R, Udit S, Batzer MA, Feschotte C. 2006b. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc Natl Acad Sci USA 103:8101–8106.
Cotrim NH, Auricchio MT, Vicente JP, Otto PA, Mingroni-Netto RC. 2004. Polymorphic Alu insertions in six Brazilian African-derived populations. Am J Hum Biol 16:264–277.
CSAC. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.
Davenport TR, Stanley WT, Sargis EJ, De Luca DW, Mpunga NE, Machaga SJ, Olson LE. 2006. A new genus of African monkey, Rungwecebus: morphology, ecology, and molecular phylogenetics. Science 312:1378–1381.
de Pancorbo MM, Lopez-Martinez M, Martinez-Bouzas C, Castro A, Fernandez-Fernandez I, de Mayolo GA, de Mayolo AA, de Mayolo PA, Rowold DJ, Herrera RJ. 2001. The Basques according to polymorphic Alu insertions. Hum Genet 109:224–233.
Deininger PL, Batzer MA. 1993. Evolution of retroposons. In: Hect M, MacIntyre RJ, Clegg M, editors. Evolutionary biology. New York: Plenum. p 157–196.
Deininger PL, Batzer MA. 1999. Alu repeats and human disease. Mol Genet Metab 67:183–193.
Deininger PL, Batzer MA. 2002. Mammalian retroelements. Genome Res 12:1455–1465.
Deininger PL, Roy-Engel AM. 2002. Mobile elements in animal and plant genomes. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, DC: ASM Press.
Delson E. 1992. Evolution of Old Word monkeys. In: Johns JS, Martin RD, Pilbeam D, Bunney S, editors. The Cambridge encyclopedia of human evolution. Cambridge: Cambridge University Press. p 217–222.
Dewannieux M, Esnault C, Heidmann T. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35: 41–48.
Disotell TR. 2000. Molecular systematics of the Cercopithecidae. In: Whitehead PF, Jolly CJ, editors. Old World monkeys. Cambridge: Cambridge University Press. p 29–56.
Disotell TR. 2003. Primates: phylogenetics. Encyclopedia of the human genome. London: Nature.
Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. 1991. Isolation of an active human transposable element. Science 254:1805–1808.
Domingo-Roura X, Marmi J, Ferrando A, Lopez-Giraldez F, Macdonald DW, Jansman HAH. 2006. Badger hair in shaving brushes comes from protected Eurasian badgers. Biol Conserv 128:425–430.
Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM. 2004. Mitochondrial DNA and Alu insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. Am J Hum Biol 16:479–488.
Dunn DS, Romphruk AV, Leelayuwat C, Bellgard M, Kulski JK. 2005. Polymorphic Alu insertions and their associations with MHC class I alleles and haplotypes in the northeastern Thais. Ann Hum Genet 69:364–372.

Economou-Pachnis A, Tsichlis PN. 1985. Insertion of an Alu SINE in the human homologue of the Mlvi-2 locus. Nucleic Acids Res 13:8379–8387.

Ejima Y, Yang L. 2003. Tran mobilization of genomic DNA as a mechanism for retrotransposon-mediated exon shuffling. Hum Mol Genet 12:1321–1328.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, Remington KA, Strausberg RL, Venter JC, Wilson RK, Batzer MA, Bustamante CD, Eichler EE, Hahn MW, Hardison RC, Makova KD, Miller W, Milosavljevic A, Palermo RE, Siepel A, Sikela JM, Attaway T, Bell S, Bernard KE, Buhay CJ, Chandrabose MN, Dao M, Davis C, Delehaunty KD, Ding Y, Dinh HH, Dugan-Rocha S, Fulton LA, Gabisi RA, Garner TT, Godfrey J, Hawes AC, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Kirkness EF, Cree A, Fowler RG, Lee S, Lewis LR, Li Z, Liu YS, Moore SM, Muzny D, Nazareth LV, Ngo DN, Okwuonu GO, Pai G, Parker D, Paul HA, Pfannkoch C, Pohl CS, Rogers YH, Ruiz SJ, Sabo A, Santibanez J, Schneider BW, Smith SM, Sodergren E, Svatek AF, Utterback TR, Vattathil S, Warren W, White CS, Chinwalla AT, Feng Y, Halpern AL, Hillier LW, Huang X, Minx P, Nelson JO, Pepin KH, Qin X, Sutton GG, Venter E, Walenz BP, Wallis JW, Worley KC, Yang SP, Jones SM, Marra MA, Rocchi M, Schein JE, Baertsch R, Clarke L, Csuros M, Glasscock J, Harris RA, Havlak P, Jackson AR, Jiang H, Liu Y, Messina DN, Shen Y, Song HX, Wylie T, Zhang L, Birney E, Han K, Konkel MK, Lee J, Smit AF, Ullmer B, Wang H, Xing J, Burhans R, Cheng Z, Karro JE, Ma J, Raney B, She X, Cox MJ, Demuth JP, Dumas LJ, Han SG, Hopkins J, Karimpour-Fard A, Kim YH, Pollack JR, Vinar T, Addo-Quaye C, Degenhardt J, Denby A, Hubisz MJ, Indap A, Kosiol C, Lahn BT, Lawson HA, Marklein A, Nielsen R, Vallender EJ, Clark AG, Ferguson B, Hernandez RD, Hirani K, Kehrer-Sawatzki H, Kolb J, Patil S, Pu LL, Ren Y, Smith DG, Wheeler DA, Schenck I, Ball EV, Chen R, Cooper DN, Giardine B, Hsu F, Kent WJ, Lesk A, Nelson DL, O'Brien WE, Prufer K, Stenson PD, Wallace JC, Ke H, Liu XM, Wang P, Xiang AP, Yang F, Barber GP, Haussler D, Karolchik D, Kern AD, Kuhn RM, Smith KE, Zwieg AS. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–234.

Gonzalez-Perez E, Via M, Esteban E, Lopez-Alomar A, Mazieres S, Harich N, Kandil M, Dugoujon JM, Moral P. 2003. Alu insertions in the Iberian Peninsula and north west Africa—genetic boundaries or melting pot? Coll Antropol 27:491–500.

Goodman M. 1999. The genomic record of Humankind's evolutionary roots. Am J Hum Genet 64:31–39.

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598.

Groves C. 2001. Primate taxonomy. Washington, DC: Smithsonian Press.

Hamdi HK, Nishio H, Tavis J, Zielinski R, Dugaiczyk A. 2000. Alu-mediated phylogenetic novelties in gene regulation and development. J Mol Biol 299:931–939.

Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, Huang CT, Sandifer E, Hebert K, Barnes EW, Hubley R, Miller W, Smit AF, Ullmer B, Batzer MA. 2007. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. Science 316:238–240.

Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA. 2005. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res 33:4040–4052.

Hansen RS. 2003. X inactivation-specific methylation of LINE-1 elements by DNMT3B: implications for the Lyon repeat hypothesis. Hum Mol Genet 12:2559–2567.

Hasagawa M, and Kishino H. 1991. DNA sequence analysis and evolution of Hominoidea. In: Kimura M, Takahata N, editors. New aspects of the genetics of molecular evolution. Tokyo: Japan Scientific Societies Press. p 303–317.

Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. PLoS Biol 2:e312.

Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA. 2004. Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 14:1068–1075.

Hedges DJ, Cordaux R, Xing J, Witherspoon DJ, Rogers AR, Jorde LB, Batzer MA. 2005. Modeling the amplification dynamics of human Alu retrotransposons. PLoS Comput Biol 1:e44.

Hedges DJ, Walker JA, Callinan PA, Shewale JG, Sinha SK, Batzer MA. 2003. Mobile element-based assay for human gender determination. Anal Biochem 312:77–79.

Herke SW, Xing J, Ray DA, Zimmerman JW, Cordaux R, Batzer MA. 2007. A SINE-based dichotomous key for primate identification. Gene 390:39–51.

Herrera RJ, Rojas DP, Terreros MC. 2007. Polymorphic Alu insertions among Mayan populations. J Hum Genet 52:129–142.

Hillis DM. 1999. SINEs of the perfect character. Proc Natl Acad Sci USA 96:9979–9981.

Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr. 1994. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimeric insertion. Nat Genet 7:143–148.

Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc Natl Acad Sci USA 92:532–536.

IHMC. 2005. A haplotype map of the human genome. Nature 437:1299–1320.

Jones T, Ehardt CL, Butynski TM, Davenport TR, Mpunga NE, Machaga SJ, De Luca DW. 2005. The highland mangabey Lophocebus kipunji: a new species of African monkey. Science 308:1161–1164.

Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC. 1997. Microsatellite diversity and the demographic history of modern humans. Proc Natl Acad Sci USA 94:3100–3103.

Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 66:979–988.

Kajikawa M, Okada N. 2002. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 111:433–444.

Kass DH, Batzer MA, Deininger PL. 1995. Gene conversion as a secondary mechanism of short interspersed element (SINE) evolution. Mol Cell Biol 15:19–25.

Kay RF, Williams BA, Ross CF, Takai M, and Shigehara N. 2004. Anthropoid origins: a phylogenetic analysis. In: Ross CF, Kay RF, editors. Anthropoid origins: new visions. New York: Kluwer. p 91–135.

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.

Krull M, Brosius J, Schmitz J. 2005. Alu-SINE exonization: en route to protein-coding function. Mol Biol Evol 22:1702–1711.

Kuryshev VY, Skryabin BV, Kremerskothen J, Jurka J, Brosius J. 2001. Birth of a gene: locus of neuronal BC200 snmRNA in three prosimians and human BC200 pseudogenes as archives of change in the Anthropoidea lineage. J Mol Biol 309:1049–1066.

Labrador M, Corces VG. 2002. Interactions between transposable elements and the host genome. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, DC: ASM Press.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough

R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J. 2001. Initial sequencing and analysis of the human genome. Nature 409:860–921.

Lev-Maor G, Sorek R, Shomron N, Ast G. 2003. The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. Science 300:1288–1291.

Liu G, Zhao S, Bailey JA, Sahinalp SC, Alkan C, Tuzun E, Green ED, Eichler EE. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. Genome Res 13:358–368.

Lorenc A, Makalowski W. 2003. Transposable elements and vertebrate protein diversity. Genetica 118:183–191.

Lorenz JG, Jackson WE, Beck JC, 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.

Lyon MF. 1998. X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet 80:133–137.

Maca-Meyer N, Villar J, Perez-Mendez L, Cabrera de Leon A, Flores C. 2004. A tale of aborigines, conquerors and slaves: Alu insertion polymorphisms and the peopling of Canary Islands. Ann Hum Genet 68:600–605.

Majumder PP, Roy B, Banerjee S, Chakraborty M, Dey B, Mukherjee N, Roy M, Thakurta PG, Sil SK. 1999. Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet 7:435–446.

Mansoor A, Mazhar K, Khaliq S, Hameed A, Rehman S, Siddiqi S, Papaioannou M, Cavalli-Sforza LL, Mehdi SQ, Ayub Q. 2004. Investigation of the Greek ancestry of populations from northern Pakistan. Hum Genet 114:484–490.

Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004. The effects of human population structure on large genetic association studies. Nat Genet 36:512–517.

Mastana S, Lee D, Singh PP, Singh M. 2003. Molecular genetic variation in the East Midlands, England: analysis of VNTR, STR and Alu insertion/deletion polymorphisms. Ann Hum Biol 30:538–550.

Mateus Pereira LH, Socorro A, Fernandez I, Masleh M, Vidal D, Bianchi NO, Bonatto SL, Salzano FM, Herrera RJ. 2005. Phylogenetic information in polymorphic L1 and Alu insertions from East Asians and Native American populations. Am J Phys Anthropol 128:171–184.

Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, Kakol J, Buyske S, Chui B, Cohen P, de Toma C, Ehm M, Glanowski S, He C, Heil J, Markianos K, McMullen I, Pericak-Vance MA, Silbergleit A, Stein L, Wagner M, Wilson AF, Winick JD, Winn-Deen ES, Yamashiro CT, Cann HM, Lai E, Holden AL. 2003. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. Am J Hum Genet 73:271–284.

Matsubara M, Basabose AK, Omari I, Kaleme K, Kizungu B, Sikubwabo K, Kahindo M, Yamagiwa J, Takenaka O. 2005. Species and sex identification of western lowland gorillas (*Gorilla gorilla gorilla*), eastern lowland gorillas (*Gorilla beringei graueri*) and humans. Primates 46:199–202.

Minghetti PP, Dugaiczyk A. 1993. The emergence of new DNA repeats and the divergence of primates. Proc Natl Acad Sci USA 90:1872–1876.

Moran JV, DeBerardinis RJ, Kazazian HH Jr. 1999. Exon shuffling by L1 retrotransposition. Science 283:1530–1534.

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

Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet 31:159–165.

Murata S, Takasaki N, Saitoh M, Okada N. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci USA 90:6995–6999.

Murata S, Takasaki N, Saitoh M, Tachida H, Okada N. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics 142:915–926.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409:614–618.

Myers JS, Vincent BJ, Udall H, Watkins WS, Morrish TA, Kilroy GE, Swergold GD, Henke J, Henke L, Moran JV, Jorde LB, Batzer MA. 2002. A comprehensive analysis of recently integrated human Ta L1 elements. Am J Hum Genet 71:312–326.

Myers S, Spencer CC, Auton A, Bottolo L, Freeman C, Donnelly P, McVean G. 2006. The distribution and causes of meiotic recombination in the human genome. Biochem Soc Trans 34:526–530.

Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, Stoneking M. 2001. Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. Eur J Hum Genet 9:267–272.

Nicklas JA, Buel E. 2003a. Development of an Alu-based. QSY 7-labeled primer PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 48:282–291.

Nicklas JA, Buel E. 2003b. Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 48:936–944.

Nicklas JA, Buel E. 2003c. Quantification of DNA in forensic samples. Anal Bioanal Chem 376:1160–1167.

Nicklas JA, Buel E. 2005. An Alu-based, MGB Eclipse real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 50:1081–1090.

Nikaido M, Rooney AP, 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 USA 96:10261–10266.

Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. Proc Natl Acad Sci USA 103:9929–9934.

Nouaud D, Quesneville H, Anxolabehere D. 2003. Recurrent exon shuffling between distant P-element families. Mol Biol Evol 20:190–199.

Novick GE, Novick CC, Yunis J, Yunis E, Antunez de Mayolo P, Scheer WD, Deininger PL, Stoneking M, York DS, Batzer MA, Herrera RJ. 1998. Polymorphic Alu insertions and the Asian origin of Native American populations. Hum Biol 70:23–39.

O'HUigin C, Satta Y, Takahata N, Klein J. 2002. Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. Mol Biol Evol 19:1501–1513.

Okada N. 1991. SINEs. Curr Opin Genet Dev 1:498–504.

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

Ostertag EM, Kazazian HH Jr. 2001. Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538.

Otieno AC, Carter AB, Hedges DJ, Walker JA, Ray DA, Garber RK, Anders BA, Stoilova N, Laborde ME, Fowlkes JD, Huang CH, Perodeau B, Batzer MA. 2004. Analysis of the human Alu Ya-lineage. J Mol Biol 342:109–118.

Pace JK 2nd, Feschotte C. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17:422–432.

Page SL, Chiu C, Goodman M. 1999. Molecular phylogeny of Old World monkeys (Cercopithecidae) as inferred from γ-globin DNA sequences. Mol Phylogenet Evol 13:348–359.

Page SL, Goodman M. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol Phylogenet Evol 18:14–25.

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

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103–1108.

Perna NT, Batzer MA, Deininger PL, Stoneking M. 1992. Alu insertion polymorphism: a new type of marker for human population studies. Hum Biol 64:641–648.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rasmussen DT, Nekaris KA. 1998. Evolutionary history of lorisiform primates. Folia Primatol 69:250–285.

Ray DA, Batzer MA. 2005. Tracking Alu evolution in New World primates. BMC Evol Biol 5:51.

Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, Christian AT, Turteltaub K, Batzer MA. 2005a. Inference of human geographic origins using Alu insertion polymorphisms. Forensic Sci Int 153:117–124.

Ray DA, Xing J, Hedges DJ, Hall MA, Laborde ME, Anders BA, White BR, Stoilova N, Fowlkes JD, Landry KE, Chemnick LG, Ryder OA, Batzer MA. 2005b. Alu insertion loci and platyrrhine primate phylogeny. Mol Phylogenet Evol 35:117–126.

Ray DA, Xing J, Salem AH, Batzer MA. 2006. SINEs of a nearly perfect character. Syst Biol 55:928–935.

Reich DE, Gabriel SB, Altshuler D. 2003. Quality and completeness of SNP databases. Nat Genet 33:457–458.

Rogers AR, Wooding S, Huff CD, Batzer MA, Jorde LB. 2007. Ancestral alleles and population origins: inferences depend on mutation rate. Mol Biol Evol 24:990–997.

Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, Sherry ST, Stoneking M, Batzer MA, Barbujani G. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12:602–612.

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

Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, Feldman MW. 2005. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet 1:e70.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. Science 298:2381–2385.

Ross C, Williams B, Kay RF. 1998. Phylogenetic analysis of anthropoid relationships. J Hum Evol 35:221–306.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, Havlak P, Hodgson A, Metzker ML, Richards S, Scott G, Steffen D, Sodergren E, Wheeler DA, Worley KC, Ainscough R, Ambrose KD, Ansari-Lari MA, Aradhya S, Ashwell RI, Babbage AK, Bagguley CL, Ballabio A, Banerjee R, Barker GE, Barlow KF, Barrett IP, Bates KN, Beare DM, Beasley H, Beasley O, Beck A, Bethel G, Blechschmidt K, Brady N, Bray-Allen S, Bridgeman AM, Brown AJ, Brown MJ, Bonnin D, Bruford EA, Buhay C, Burch P, Burford D, Burgess J, Burrill W, Burton J, Bye JM, Carder C, Carrel L, Chako J, Chapman JC, Chavez D, Chen E, Chen G, Chen Y, Chen Z, Chinault C, Ciccodicola A, Clark SY, Clarke G, Clee CM, Clegg S, Clerc-Blankenburg K, Clifford K, Cobley V, Cole CG, Conquer JS, Corby N, Connor RE, David R, Davies J, Davis C, Davis J, Delgado O, Deshazo D, Dhami P, Ding Y, Dinh H, Dodsworth S, Draper H, Dugan-Rocha S, Dunham A, Dunn M, Durbin KJ, Dutta I, Eades T, Ellwood M, Emery-Cohen A, Errington H, Evans KL, Faulkner L, Francis F, Frankland J, Fraser AE, Galgoczy P, Gilbert J, Gill R, Glockner G, Gregory SG, Gribble S, Griffiths C, Grocock R, Gu Y, Gwilliam R, Hamilton C, Hart EA, Hawes A, Heath PD, Heitmann K, Hennig S, Hernandez J, Hinzmann B, Ho S, Hoffs M, Howden PJ, Huckle EJ, Hume J, Hunt PJ, Hunt AR, Isherwood J, Jacob L, Johnson D, Jones S, de Jong PJ, Joseph SS, Keenan S, Kelly S, Kershaw JK, Khan Z, Kioschis P, Klages S, Knights AJ, Kosiura A, Kovar-Smith C, Laird GK, Langford C, Lawlor S, Leversha M, Lewis L, Liu W, Lloyd C, Lloyd DM, Loulseged H, Loveland JE, Lovell JD, Lozado R, Lu J, Lyne R, Ma J, Maheshwari M, Matthews LH, McDowall J, McLaren S, McMurray A, Meidl P, Meitinger T, Milne S, Miner G, Mistry SL, Morgan M, Morris S, Muller I, Mullikin JC, Nguyen N, Nordsiek G, Nyakatura G, O'Dell CN, Okwuonu G, Palmer S, Pandian R, Parker D, Parrish J, Pasternak S, Patel D, Pearce AV, Pearson DM, Pelan SE, Perez L, Porter KM, Ramsey Y, Reichwald K, Rhodes S, Ridler KA, Schlessinger D, Schueler MG, Sehra HK, Shaw-Smith C, Shen H, Sheridan EM, Shownkeen R, Skuce CD, Smith ML, Sotheran EC, Steingruber HE, Steward CA, Storey R, Swann RM, Swarbreck D, Tabor PE, Taudien S, Taylor T, Teague B, Thomas K, Thorpe A, Timms K, Tracey A, Trevanion S, Tromans AC, d'Urso M, Verduzco D, Villasana D, Waldron L, Wall M, Wang Q, Warren J, Warry GL, Wei X, West A, Whitehead SL, Whiteley MN, Wilkinson JE, Willey DL, Williams G, Williams L, Williamson A, Williamson H, Wilming L, Woodmansey RL, Wray PW, Yen J, Zhang J, Zhou J, Zoghbi H, Zorilla S, Buck D, Reinhardt R, Poustka A, Rosenthal A, Lehrach H, Meindl A, Minx PJ, Hillier LW, Willard HF, Wilson RK, Waterston RH, Rice CM, Vaudin M, Coulson A, Nelson DL, Weinstock G, Sulston JE, Durbin R, Hubbard T, Gibbs RA, Beck S, Rogers J, Bentley DR. 2005. The DNA sequence of the human X chromosome. Nature 434:325–337.

Roy AM, Carroll ML, Nguyen SV, Salem AH, Oldridge M, Wilkie AO, Batzer MA, Deininger PL. 2000. Potential gene conversion and source genes for recently integrated Alu elements. Genome Res 10:1485–1495.

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL. 2001. Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159:279–290.

Rozmahel R, Heng HH, Duncan AM, Shi XM, Rommens JM, Tsui LC. 1997. Amplification of CFTR exon 9 sequences to multiple locations in the human genome. Genomics 45:554–561.

Ruvolo M, Disotell TR, Allard MW, Brown WM, Honeycutt RL. 1991. Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proc Natl Acad Sci USA 88:1570–1574.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928–933.

Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA. 2003a. Recently integrated Alu elements and human genomic diversity. Mol Biol Evol 20:1349–1361.

Salem AH, Myers JS, Otieno AC, Watkins WS, Jorde LB, Batzer MA. 2003b. LINE-1 preTa elements in the human genome. J Mol Biol 326:1127–1146.

Salem AH, Ray DA, Batzer MA. 2005. Identity by descent and DNA sequence variation of human SINE and LINE elements. Cytogenet Genome Res 108:63–72.

Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA. 2003c. Alu elements and hominid phylogenetics. Proc Natl Acad Sci USA 100:12787–12791.

Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC. 2003. SNP and haplotype variation in the human genome. Mutat Res 526:53–61.

Santos FR, Pandya A, Tyler-Smith C. 1998. Reliability of DNA-based sex tests. Nat Genet 18:103.

Satta Y, Klein J, Takahata N. 2000. DNA archives and our nearest relative: the trichotomy problem revisited. Mol Phylogenet Evol 14:259–275.

Schmid CW. 1998. Does SINE evolution preclude Alu function? Nucleic Acids Res 26:4541–4550.

Schmid CW, Jelinek WR. 1982. The Alu family of dispersed repetitive sequences. Science 216:1065–1070.

Schmitz J, Ohme M, Suryobroto B, Zischler H. 2002a. The colugo (*Cynocephalus variegatus*, Dermoptera): the primates' gliding sister? Mol Biol Evol 19:2308–2312.

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, Ohme M, Zischler H. 2002b. The complete mitochondrial sequence of *Tarsius bancanus*: evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol Biol Evol 19:544–553.

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

Schmitz J, Zischler H. 2003. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating dermopterans from primates. Mol Phylogenet Evol 28:341–349.

Schneider H. 2000. The current status of the New World monkey phylogeny. An Acad Bras Cienc 72:165–172.

Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. 2006. Human genomic deletions mediated by recombination between Alu elements. Am J Hum Genet 79:41–53.

Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays 22:148–160.

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

Sheen FM, Sherry ST, Risch GM, Robichaux M, Nasidze I, Stoneking M, Batzer MA, Swergold GD. 2000. Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. Genome Res 10:1496–1508.

Shen L, Wu LC, Sanlioglu S, Chen R, Mendoza AR, Dangel AW, Carroll MC, Zipf WB, Yu CY. 1994. Structure and genetics of the partially duplicated gene RP located immediately upstream of the complement C4A and the C4B genes in the HLA class III region. Molecular cloning, exon–intron structure, composite retroposon, and breakpoint of gene duplication. J Biol Chem 269:8466–8476.

Sherry ST, Harpending HC, Batzer MA, Stoneking M. 1997. Alu evolution in human populations: using the coalescent to estimate effective population size. Genetics 147:1977–1982.

Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388:666–670.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, Ferrell RE. 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. Am J Hum Genet 60:957–964.

Sifis ME, Both K, Burgoyne LA. 2002. A more sensitive method for the quantitation of genomic DNA by Alu amplification. J Forensic Sci 47:589–592.

Singer MF. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28:433–434.

Singer SS, Schmitz J, Schwiegk C, Zischler H. 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol Phylogenet Evol 26:490–501.

Sinha A, Datta A, Madhusudan MD, Mishra C. 2005. Macaca munzala: a new species from western Arunachal Pradesh, northeastern India. Int J Primatol 26:977–989.

Smit AF. 1996. The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748.

Smit AF, Riggs AD. 1996. Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci USA 93:1443–1448.

Smit AF, Toth G, Riggs AD, Jurka J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246:401–417.

Song M, Boissinot S. 2007. Selection against LINE-1 retrotransposons results principally from their ability to mediate ectopic recombination. Gene 390:206–213.

Sorek R, Ast G, Graur D. 2002. Alu-containing exons are alternatively spliced. Genome Res 12:1060–1067.

Steinlechner M, Berger B, Niederstatter H, Parson W. 2002. Rare failures in the amelogenin sex test. Int J Legal Med 116:117–120.

Steiper ME, Ruvolo M. 2003. New World monkey phylogeny based on X-linked G6PD DNA sequences. Mol Phylogenet Evol 27:121–130.

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA. 1997. Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071.

Stringer CB, Andrews P. 1988. Genetic and fossil evidence for the origin of modern humans. Science 239:1263–1268.

Sullivan KM, Mannucci A, Kimpton CP, Gill P. 1993. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X–Y homologous gene amelogenin. Biotechniques 15:636–638,640–641.

Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J. 1997. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Mol Ecol 6:869–876.

Takahashi K, Terai Y, Nishida M, Okada N. 1998. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. Mol Biol Evol 15:391–407.

Takahashi K, Terai Y, Nishida M, Okada N. 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol Biol Evol 18:2057–2066.

Terreros MC, Martinez L, Herrera RJ. 2005. Polymorphic Alu insertions and genetic diversity among African populations. Hum Biol 77:675–704.

Thangaraj K, Reddy AG, Singh L. 2002. Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? Int J Legal Med 116:121–123.

Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340.

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

Vidal F, Mougneau E, Glaichenhaus N, Vaigot P, Darmon M, Cuzin F. 1993. Coordinated posttranscriptional control of gene expression by modular elements including Alu-like repetitive sequences. Proc Natl Acad Sci USA 90:208–212.

Vincent BJ, Myers JS, Ho HJ, Kilroy GE, Walker JA, Watkins WS, Jorde LB, Batzer MA. 2003. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol Biol Evol 20:1338–1348.

Vishwanathan H, Edwin D, Usharani MV, Majumder PP. 2003. Insertion/deletion polymorphisms in tribal populations of southern India and their possible evolutionary implications. Hum Biol 75:873–887.

Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform Ser Workshop Genome Inform 12:141–154.

Walker JA, Hughes DA, Anders BA, Shewale J, Sinha SK, Batzer MA. 2003a. Quantitative intra-short interspersed element PCR for species-specific DNA identification. Anal Biochem 316:259–269.

Walker JA, Kilroy GE, Xing J, Shewale J, Sinha SK, Batzer MA. 2003b. Human DNA quantitation using Alu element-based polymerase chain reaction. Anal Biochem 315:122–128.

Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. 1991. A de novo Alu insertion results in neurofibromatosis type 1. Nature 353:864–866.

Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. 2005. SVA elements: a hominid-specific retroposon family. J Mol Biol 354:994–1007.

Watkins WS, Ricker CE, Bamshad MJ, Carroll ML, Nguyen SV, Batzer MA, Harpending HC, Rogers AR, Jorde LB. 2001. Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. Am J Hum Genet 68:738–752.

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BV, Reddy PG, Das PK, Batzer MA, Jorde LB. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 13:1607–1618.

Weiner AM, Deininger PL, Efstratiadis A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661.

Witherspoon DJ, Marchani EE, Watkins WS, Ostler CT, Wooding SP, Anders BA, Fowlkes JD, Boissinot S, Furano AV, Ray DA, Rogers AR, Batzer MA, Jorde LB. 2006. Human population genetic structure and diversity inferred from polymorphic L1(LINE-1) and Alu insertions. Hum Hered 62:30–46.

Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, and Jorde LB. 2007. Genetic similarities within and between human populations. Genetics 176:351–359.

Xiao FX, Yang JF, Cassiman JJ, Decorte R. 2002. Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China. Hum Biol 74:555–568.

Xing J, Salem AH, Hedges DJ, Kilroy GE, Watkins WS, Schienman JE, Stewart CB, Jurka J, Jorde LB, Batzer MA. 2003. Comprehensive analysis of two Alu Yd subfamilies. J Mol Evol 57 (Suppl 1):S76–S89.

Xing J, Wang H, Belancio VP, Cordaux R, Deininger PL, Batzer MA. 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. Proc Natl Acad Sci USA 103:17608–17613.

Xing J, Wang H, Han K, Ray DA, Huang CH, Chemnick LG, Stewart CB, Disotell TR, Ryder OA, Batzer MA. 2005. A mobile element based phylogeny of Old World monkeys. Mol Phylogenet Evol 37:872–880.

Xing J, Wang H, Zhang Y, Ray DA, Tosi AJ, Disotell TR, Batzer MA. 2007. A mobile element-based evolutionary history of guenons (tribe Cercopithecini). BMC Biol 5:5.

Yan P, Wu XB, Shi Y, Gu CM, Wang RP, Wang CL. 2005. Identification of Chinese alligators (Alligator sinensis) meat by diagnostic PCR of the mitochondrial cytochrome b gene. Biol Conserv 121:45–51.

Yoder AD. 1997. Back to the future: a synthesis of strepsirrhine systematics. Evol Anthropol 6:11–22.

Yoder AD. 2003. The phylogenetic position of genus Tarsius: whose side are you on? In: Wright PC, Simons EL, and Gursky S, editors. Tarsiers: past, present, and future. New Brunswick, NJ: Rutgers University Press. p 161–175.

Zampicinini G, Blinov A, Cervella P, Guryev V, Sella G. 2004. Insertional polymorphism of a non-LTR mobile element (NLRCth1) in European populations of *Chironomus riparius* (Diptera, Chironomidae) as detected by transposon insertion display. Genome 47:1154–1163.

Zhi D. 2007. Sequence correlation between neighboring Alu instances suggests post-retrotransposition sequence exchange due to Alu gene conversion. Gene 390:117–121.

Zietkiewicz E, Richer C, Labuda D. 1999. Phylogenetic affinities of tarsier in the context of primate Alu repeats. Mol Phylogenet Evol 11:77–83.