1-1-2010

Laboratory methods for the analysis of primate mobile elements

David A. Ray

Kyudong Han

Jerilyn A. Walker

Mark A. Batzer

Recommended Citation
Ray, D., Han, K., Walker, J., & Batzer, M. (2010). Laboratory methods for the analysis of primate mobile elements. *Methods in Molecular Biology, 628*, 153-179. https://doi.org/10.1007/978-1-60327-367-1_9

Published in final edited form as:
Methods Mol Biol. 2010 ; 628: 153–179. doi:10.1007/978-1-60327-367-1_9.

Laboratory methods for the analysis of primate mobile elements

David A. Ray[a], **Kyudong Han**[b], **Jerilyn A. Walker**[b], and **Mark A. Batzer**[b]

[a]Department of Biology, West Virginia University, Morgantown, WV 26505

[b]Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA.

Abstract

Mobile elements represent a unique and powerful set of tools for understanding the variation in a genome. Methods exist not only to utilize the polymorphisms among and within taxa to various ends but also to investigate the mechanism through which mobilization occurs. The number of methods to accomplish these ends is ever growing. Here we present several protocols designed to assay mobile element-based variation within and among individual genomes.

1. Introduction

Mobile elements are interspersed repetitive DNA sequences with the unique ability to spread copies of themselves throughout the genome they occupy. As a result, these sequences can comprise a large proportion of the genomes in which they are found (1,2). Mobile elements may be divided into two classes depending on how they mobilize and the type of intermediate they use. Class I elements include the retrotransposons, which utilize an RNA intermediate during retrotransposition, while DNA transposons, Class II, utilize a DNA intermediate during mobilization (3). While DNA transposons have had periods of activity early in primate evolution, all major recent activity in the human lineage has been retrotransposon-based (1, 4). Thus, we will focus on these for this chapter.

Retrotransposons from the human lineage include L1 (a Long INterspersed Element), *Alu* (a primate-specific Short INterspersed Element) and SVA (a composite retrotransposon). Together these elements have had significant impacts on the architecture of primate genomes (5). They comprise over ~40% of the human genome by mass and are the most abundant interspersed elements therein (6). Because of their high copy number, these interspersed repeats have been a significant source of variation as a result of insertion, transduction, and post-integration recombination among elements (6,7). During retrotransposition the RNA copy is reverse transcribed by target primed reverse transcription (TPRT) and subsequently integrated into the host genome (8,9,10). Unable to retrotranspose autonomously, *Alu* and SVA elements are thought to borrow the enzymatic factors required for their mobilization from L1 elements (8,11,12,13,14,15), which encode a protein complex with endonuclease and reverse transcriptase activity (16,17).

Over the millions of years of primate evolution, retrotransposons have tended to accumulate in a hierarchical manner. This pattern is a direct result of the mechanisms through which they mobilize and insert, a modified version of the master gene model (18,19,20,21). Evidence suggests that a subfamily will accumulate copies for a certain time period and then become quiescent. Other newer subfamilies subsequently become active and the pattern repeats itself. This pattern is well illustrated by the *Alu* family of SINEs. Over time, the *Alu* element has diversified into a variety of subfamilies, each with its own set of diagnostic sequence characteristics and period of activity. For example, during the early stages of primate evolution, *AluJ* subfamilies were active. The activity of these subfamilies was later reduced (if not

extinguished) and the *Alu*S subfamilies, derivatives of *Alu*J, became active. Thus, while *Alu*J elements are found in all primates, *Alu*S elements are found only in anthropoid primates (tarsiers, Platyrrhini, Catarrhini). The *Alu*Y subfamilies (22) are even more taxonomically specific in that they began their expansion in catarrhine primates after the platyrrhine-catarrhine split (23). Thus, each taxon has a unique pattern of insertions some of which are shared with other closely related taxa and others that are unique to that lineage. For example, the most recent *Alu* elements to mobilize in our own genome belong to a series of *Alu*Y subfamilies (*Alu*Yb8, *Alu*Ya5a2, etc.) that are exclusively or primarily specific to the human branch of the primate tree (24,25,26,27).

As genetic markers, retrotransposons of all sorts offer certain advantages over more commonly used genetic characters such as microsatellites and sequence data. First and foremost is the observation that these markers are an essentially homoplasy-free set of characters (28,29,30, 31,32). Unlike many other genetic markers, they tend to exist as character states for no other reason than inheritance from a common ancestor. Thus, they are almost invariably identical by descent, not just identical by state. As a result, they can be used to provide an extremely accurate picture of evolutionary and population relationships (33,34,35,36,37,38,39). We also know that the ancestral state at any locus is the absence of the element and once the element is inserted it typically remains there indefinitely. These characteristics result in a relatively simple evolutionary model to be applied when interpreting the data.

SINEs and other retrotransposons share other desirable characteristics as well. The vast majority of mobile element insertions in the host genome are neutral residents(40). The process of genotyping individuals to determine insertion presence or absence at any given set of loci is a relatively simple task involving easily distinguishable fragments on a simple agarose gel stained with ethidium bromide. Multiplexing of loci is possible (41,42) and fluorescently labeled primers may also be used if one is interested in automated analysis (43). These features make the analysis of *Alu* elements a robust tool for tracking human geographic origins.

In the following pages we will describe several techniques that have been used to investigate aspects of primate biology and mobile element biology in the primate lineage. We will not only focus on the human lineage but will also mention some techniques that are widely applicable to other taxa, especially non-human primates. We will describe the laboratory techniques required to investigate questions from the fields of forensics, population biology, phylogenetics, genome evolution, and the biology of the elements themselves. One advantage primate researchers have over many other taxa is the availability of a variety of primate genome sequences to serve as a resource and reference in their work. Many of the laboratory techniques to be described benefit from the availability of these sequences and we suggest the reader reference the companion chapter in this volume dedicated to computational analysis of primate/human mobile elements.

1.1. Forensic Applications

Many of the unique properties of mobile elements make them ideally suited for a variety of forensic applications. This section will focus on the *Alu* element, the most abundant class of SINE in the human genome. Most *Alu* elements have become permanent residents of the human genome and are “fixed present”, meaning that all individuals are homozygous for the insertion at a particular locus. The continued expansion of *Alu* elements throughout primate evolution has created several recently integrated “young” subfamilies that are present in the human genome, but largely absent from nonhuman primates (24,25,27,44,45). Some members of these young *Alu* subfamilies have inserted in the human genome recently enough that individuals remain polymorphic for the insertion presence/absence. Both fixed and polymorphic *Alu* elements have been utilized successfully as robust forensic tools.

Forensic DNA analysis typically begins with the quantitation of human-specific DNA obtained from the biological sample. This is essential to determine the most appropriate autosomal and Y chromosome analysis strategies to perform (46). Highly sensitive methods for quantitation of human DNA based on *Alu* elements have been reported (46,47,48,49,50,51,52,53). These methods take advantage of the high copy number of fixed *Alu* elements in the human genome to maximize sensitivity. Human DNA quantitation based on *Alu* elements is evolving as the preferred method in the forensic community (50). The method described in this chapter utilizes a subfamily of *Alu* elements, enriched in the human genome as compared to other primate species, to maximize human specificity (52,53).

Another important forensic use for *Alu* elements is human gender identification (54). Fixed *Alu* insertions on either the X or the Y chromosome provide a simple and reliable system to identify them. *AluSTXa* and *AluSTYa* loci demonstrate 100% accuracy in X and Y chromosome identification. The combination of these two markers provides added assurance that gender identification results are accurate since two completely independent mutations would have to occur to affect the outcome.

When one thinks about forensic DNA analysis what typically comes to mind is obtaining a "match" between a crime scene DNA sample and an alleged criminal suspect, thus "solving the case." Frequently however, tools that narrow the potential pool of suspects are essential precursors to a positive identification. The inferred ancestral origin of a DNA specimen is one type of predictor evidence which can advance a criminal investigation (55). Polymorphic *Alu* insertions have been widely used to study human genetic variation in the world populations (6,56,57,58,59,60).

1.2. Taxonomic Applications

One of the most productive areas of mobile element application has been in the arena of phylogenetic inference. Numerous difficult questions regarding the evolutionary history of the primate lineage have been successfully addressed using *Alu* elements as tools. For example, Salem et al. (37) confidently resolved the human-chimpanzee-gorilla trichotomy and Ray et al. (38) successfully determined the controversial branching order of three families of platyrrhine (New World) primates. Utilizing retrotransposons as phylogenetic markers has been described a number of times. However, phylogenetic analysis of the primate lineage is unique due to the existence of several 'reference' genomes. The human (1), chimpanzee (61), and macaque (62) genomes have been released and the marmoset and orang-utan genomes will likely be released in the near future. These genomes provide a valuable resource in determining potentially informative insertions and primer design.

One important consequence of the hierarchical accumulation of retrotransposons in the genome is the ability to target subfamilies of the retrotransposon family that were active during the evolutionary period of interest. For example, if a researcher's interest is in the recent evolutionary history of tamarins, he or she would want to focus on elements belonging to the *Alu*Ta subfamilies instead of *Alu*Y, *Alu*S or *AluJ*: the reason being that all of the latter families were either inactive during that period or never proliferated in that lineage. *Alu*Ta, on the other hand has been active in the tamarin lineage over the last 15-20 million years and many informative insertions will likely be present. Methods described in the companion chapter on computational analysis can aid researchers in determining the sequences that should be targeted for any particular question.

In laboratories dealing with primate genetics, it is critical that researchers be sure that they are handling DNA from the appropriate taxon. For instance, very often researchers collect or receive DNA that was collected in a 'non-invasive' manner (i.e. 'divorced' tissues such as hair or feces) (63,64,65). This is especially true during investigations of the illegal wildlife trade

and identification of seized products (64,66,67). Even when laboratories produce their own 'in-house' genomic 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 result in the loss of their labels. Future analyses based on these mistaken identities can be compromised. We will review an *Alu*-based dichotomous key for the resolution of primate sample identity for researchers in this area.

1.3. Structural Impact of Retrotransposons

Among mobile elements, retrotransposons (e.g.,L1, *Alu*, and SVA elements) are major endogenous contributors to the creation of structural variation in primate genomes. The tempo and mode of their amplification during the primate radiation have been shown to be lineage-specific events and thus, retrotransposons have had an extensive impact on the evolutionary history of different primate lineages through shaping of their genomic landscape (1,61,68,69, 70). Computational analyses of genomic sequence, along with the use of newly developed cell culture assays, suggest that the overall contribution of retrotransposon-mediated genomic variation involves not only the initial integration event but also a variety of recombination events occurring after that integration (e.g., *Alu* retrotransposition-mediated deletions, L1 insertion-mediated deletion, and *Alu* recombination-mediated deletions) (68,71,72,73).

Completion of the human and chimpanzee reference genomes allowed whole-genome comparison studies of L1 and *Alu* insertion-mediated variation in these primate lineages. The results showed that 24 (~1.3%) of the total ~1800 human-specific L1 insertions are involved in genomic deletions and are directly responsible for the loss of ~18 kb from the human genome (72), whereas, only ~0.2% of human-specific *Alu* insertions are involved in genomic deletions and are responsible for the loss of ~9 kb from the human genome (71). Post-insertion recombination events, however, were shown to have greater genomic impact. Sen et al. (73) identified 492 *Alu* recombination-mediated genomic deletions which resulted in the loss of ~400 kb of human genomic sequence, and ~60% of these deletions are involved in known or predicted genes. Three events actually deleted functional exons from human genes as compared to orthologous chimpanzee genes (73).

Genome alignment studies such as these have helped us to understand the distribution of retrotransposons and provide insight into their impact on host genomes, but tell us little about their mobilization. It has been the development of *in vitro* cell culture based assays which have allowed us to study the mobilization dynamics of retrotransposons. A companion chapter in this volume is dedicated to computational methods for the analysis of primate/human mobile elements. Therefore, in this section we will focus on methods which utilize recently developed cell culture assays to study retrotransposition events and consider their genomic impact in cultured human cells.

The transient cultured cell retrotransposition assay was developed by Moran and his colleagues (74,75). L1.2A was isolated as a potential progenitor of disease-producing L1 insertions into the factor VIII from patient JH-27 (hemophilia A) (76). To investigate whether the L1.2A has the capacity of an autonomous retrotransposon, the sequence was cloned and subcloned into a pCEP4 expression vector including a *mneol* reporter cassette which is comprised of an antisense copy of a *neo* selectable marker, the heterologous SV40 promoter, and a polyadenylation sequence. The *neo* gene is disrupted by an intron in the opposite transcriptional orientation (74). This genetic system could display L1 retrotransposition in cultured cell lines and help to estimate the frequency of L1 autonomous retrotransposition. On the basis of these achievements, 82 out of 89 L1s with intact ORFs that exist in the human genome were cloned, and the retrotranspositional capability of each was predicted in cultured human 143B TK- osteosarcoma cells (77). Moreover, the characterization of new daughter L1 inserts generated by synthetic retrotransposition-competent L1s in cultured human cells demonstrated that L1

retrotransposition events cause genomic instability such as deletions, duplications, translocations, and intra-L1 rearrangements (78,79,80) and have the potential to provide the host genome with new gene families through L1-mediated transduction (74,81).

Through the L1-mediated *Alu* retrotransposition assay, the retrotransposed *Alu* elements and their flanking sequences were investigated to confirm the fact that *Alu* elements are indeed mobilized *in trans* by using the L1 enzymatic machinery. As a result, the new daughter *Alu* inserts derived from a *neo*^{Tet}-marked *Alu* construct were intact without deletion. Their pre-insertion sites were predominantly close to an L1 endonuclease cleavage site consensus (TT^AAAA) and on each side of the *Alu* inserts were the presence of target site duplications (TSDs), one hallmark of authentic *Alu* retrotransposition, generated by the target-site primed reverse transcription process (10,17,82). Moreover, it was noteworthy that only ORF2p products (endonuclease and reverse transcriptase domains) of L1-encoded proteins are essential for the *Alu* retrotransposition (14).

The fact that L1 retrotransposition can create genomic deletions in the human genome was revealed by the systems of L1 retrotransposition in cultured human cells and the plasmid-based rescue technique (*see* **Subheading 3.3.3.**). It revealed that ~ 20% of *de novo* L1 insertions recognized through cultured cell retrotransposition assays caused genomic deletions at the integration site and the size of DNA sequences deleted through these events ranged up to 71 kb (78,79,80).

The enormous difference in genomic variation observed between *in vitro* and *in vivo* forms of investigation could be caused by evolutionary forces (e.g., selection pressure, the number of retrotransposition-competent L1s, and effective population size) and host defense mechanisms (e.g., RNAi, APOBEC, and methylation).

2. Materials

2.1 Forensic Applications

2.1.1. *AluSTYa* and *AluSTXa* loci for human gender identification—

1. Oligonucleotide PCR primers for each locus: *AluSTYa*, Forward 5′-CATGTATTTGATGGGGATAGAGG-3′and Reverse 5′-CCTTTTCATCCAACTACCACTGA-3′, Primers for the *Alu* insertion on X chromosomes, *AluSTXa*, Forward 5′-TGAAGAAATTCAGTTCATAGCTTGT-3′ and Reverse 5′-CAGGAGATCCTGAGATTATGTGG-3. For both loci, males are distinguished as having two DNA amplicons (X and Y chromosomes), while females (two X chromosomes) have only a single amplicon (Fig. 1).
2. Standard PCR reagents, a thermal-cycler PCR machine, single channel pipettes
3. Horizontal gel electrophoresis unit and power supply; agarose, TBE buffer
4. Ethidium bromide and UV light source to record gel image

2.1.2. Intra-*Alu*Yb*8* PCR Assay for Human DNA detection and quantitation—

1. PCR primers: Forward 5′-CTTGCAGTGAGCCGAGATT-3′ and Reverse 5′-GAGACGGAGTCTCGCTCTGTC-3′.
2. TaqMan-MGB probe: 5′FAM or VIC-ACTGCAGTCCGCAGTCCGGCCT-3′-MGBNFQ (Applied Biosystems, Inc.).
3. ABI 7000 Sequence detection system or equivalent and TaqMan PCR core reagents (Kit No. 4304439 or N8080228; Applied Biosystems, Inc).

4. Human Genomic DNA standard (examples: Promega G3041; Novagen #69237)
5. Optical PCR plates and lids (Cat. No. N8010560 and 4360954, respectively; ABI)

2.1.3. Inference of human geographic origins—

1. PCR primers for a set of 100 *Alu* insertion polymorphisms and the database of genotypes for 715 individuals of known geographic ancestry from sub-Saharan Africa, East Asia, Europe and India (83). These files are available for free download at: (http://www.batzerlab.lsu.edu; publication #158, Supplementary Data) (55).
2. The program *Structure2.2*: a free software package for using multi-locus genotype data to investigate population structure (84,85). It is available for free download at http://pritch.bsd.uchicago.edu/software.html
3. Standard PCR reagents, a thermal cycler PCR machine, single channel pipettes
4. Horizontal gel electrophoresis unit and power supply; agarose, TBE buffer
5. Ethidium bromide and UV light source to record gel image

2.2. Taxonomic Applications

2.2.1. Locus Identification—

1. Linker oligonucleotides.
2. *Alu* subfamily-specific oligonucleotides primers.
3. Standard PCR reagents, thermal cycler, single channel pipettes.
4. Restriction enzyme compatible to linker oligonucleotides.
5. Genomic DNA from taxa of interest.
6. Horizontal gel electrophoresis unit and power supply; agarose, TBE buffer.
7. Ethidium bromide and UV light source to analyze PCR products.

2.2.2. Phylogeny inference—

1. Oligonucleotide primers for specific *Alu* insertion loci.
2. Standard PCR reagents, thermal cycler, single channel pipettes.
3. Genomic DNA from taxa of interest.
4. Horizontal gel electrophoresis unit and power supply; agarose, TBE buffer.
5. Ethidium bromide and UV light source to analyze PCR products.

2.2.3. Dichotomous key identification—

1. Standard set of oligonucleotides primers from Herke et al. (86). Individual researchers must determine if the entire set of oligonucleotides or some subset is required for their particular research.
2. Standard PCR reagents, thermal cycler, single channel pipettes.
3. Genomic DNA from taxa of interest.
4. Horizontal gel electrophoresis unit and power supply; agarose, TBE buffer.
5. Ethidium bromide and UV light source to analyze PCR products.

2.3. Patterns and Processes of Transpositions in Cultured Cells and within a Genome

2.3.1 Transient cultured cell retrotransposition assay (70,71)—

1. Obtain *L1.2mneol* expression vector (74): L1.2A, isolated by Dombroski et al., (76), was engineered to create *L1.2mneol* expression vector. To leave a unique *Bam*HI site flanking its 3′ end, the *Bam*HI restriction site at position 4836 of L1.2A was disrupted by site-directed mutagenesis. A *Not*I and a *Sma*I restriction sites were introduced upstream of its 5′ UTR and into 3′ UTR at position 5980, respectively. A blunt-ended 2.1 kb *Eco*RI-*Bam*HI fragment bearing the *neo* indicator cassette (87) was ligated to the *Sma*I site resulting in pJCC9 or pJCC8. Thus, the two plasmids contained a tagged L1.2A element, but pJCC8 lacked the L1.2A 5′ UTR. The pJCC9 was restricted with two restriction enzymes of *Not*I and *Bam*HI, generating the 8.1 kb *Not*I-*Bam*HI fragment which was subcloned into pCEP4 expression vector (Invitrogen) to create pJM101.
2. *Neo*^{Tet}-marked *Alu* element (14): The *Alu* sequence, integrated into intron 5 of neurofibromatosis type 1 (88), was inserted between 7SL RNA gene enhancer and termination signal using the pDL41-48 plasmid (89). Next, the *neo*^{Tet} reporter gene (controlled by the SV40 promoter) was inserted upstream of the right monomer poly (A) tail by cleaving the *Alu* sequence-containing plasmid with *Tth*111I (5′-GACNNNGTC-3′) restriction enzyme and ligating with *neo*^{Tet} PCR product.
3. CMV L1-RP expression vector (14): The cloned L1.2A (76) was inserted, as a blunt-ended *Not*I–*Nsi*I fragment, between the CMV promoter and the SV40 polyadenylation site of pCMV$_{\beta}$ (Clontech) to create CMV L1.2 expression vector (90). Next, the L1.2 sequence of the CMV L1.2 expression vector was replaced with L1$_{RP}$ sequence (91) resulting in the CMV L1-RP expression vector.

2.3.2. L1-Mediated *Alu* Retrotransposition in Cultured Human Cells—

1. Liquid N_2 is used to preserve cell lines, either in the vapor phase (−156°C) or in the liquid phase (−196°C). Hela cells (ATCC CCL2) are grown at 37 °C an atmosphere containing 5-7% CO_2 and 100% humidity in high glucose Dulbecco's modified Eagle's medium (DMEM) lacking pyruvate (GIBCO®). DMEM was supplemented with 10% fetal bovine calf serum, 0.4 m*M* glutamine, and 20U/mL penicillin-streptomycin (DMEM-complete).
2. Phosphate-buffered saline (PBS) solution: PBS contains 136.8 m*M* NaCl, 2.5 m*M* KCl, 0.8 m*M* Na_2HPO_4, 1.47 m*M* KH_2PO_4, 0.9 m*M* $CaCl_2$, and 0.5 m*M* $MgCl_2$ ($6H_2O$) in distilled water. The solution is sterilized by using 0.22-μm filter (Millipore) and is stored at room temperature.
3. Geneticin (GIBCO®): Geneticin powder is dissolved in PBS to make a 125 mg/mL stock solution, which is sterilized by using 0.22-μm filter (Millipore) and stored at −20°C.
4. FIX solution: FIX solution contains 2% formaldehyde (of a 37% stock solution in ddH_2O) and 0.2% glutaraldehyde (of a 50% stock solution in ddH_2O) in 1 × PBS and is stored at 4°C.

2.3.3. Rescue of L1 integrants from G418^R foci (Fig. 5) (74,92)—

1. HeLa genomic DNAs are isolated using the blood and/or cell Midi Prep kit (Qiagen) or the cell and tissue DNA isolation kit (Puregen; Gentra).
2. Plasmid DNAs are purified on Qiagen midi prep columns.

3. For transfection experiments, DNA superhelicity is tested by electrophoresis on 0.6–0.7% agarose-ethidium bromide gels.

3. Methods

3.1. Forensic Applications

3.1.1. Human Gender Identification—

1. Dilute *AluSTYa* and *AluSTXa* stock primers to 2u*M* in 10mM Tris / 0.1m*M* EDTA (TLE) to make a 10X working solution.
2. Obtain DNA from a human male and a human female control (if possible) and dilute all DNA samples to 5ng/uL for PCR.
3. Set-up PCR reactions with 5ul (25ng) of DNA template per 25uL reaction volume. Prepare a master mix containing PCR reagents per reaction: 1X PCR buffer, 0.2u*M* each oligonucleotide primer, 200u*M* dNTP mix, 1.5m*M* $MgCl_2$ and 1U *Taq* DNA polymerase. Add sterile water for a final volume of 20uL of mix per reaction. Prepare 20% excess master mix (if you have 10 PCR samples, make enough mix for 12 to insure accurate transfer of 20uL of mix per well).
4. Perform PCR reactions using the following conditions: Initial denaturation for 90 seconds at 94°C followed by 30-32 cycles of 95°C for 30 seconds, anneal for 30 seconds at 58°C (*AluSTYa*) or 60°C (*AluSTXa*), extension at 72°C for 30 seconds followed by a final extension at 72°C for 2 minutes.
5. Prepare a 2% agarose gel in 1X TBE buffer containing 0.2ug/mL ethidium bromide. Load 20ul of PCR product on gel, size separate PCR amplicons by electrophoresis at 150 V for 1 hour and visualize the genotypes using UV illumination (Fig. 1).

3.1.2. Human DNA Detection and Quantitation—

1. Prepare a 10-fold serial dilution of human DNA standard from 100ng to 0.01pg by first making an aliquot of 20ng/uL, then diluting a portion of that 1:10 in TLE (10m*M* Tris/ 0.1m*M* EDTA) to make 2ng/uL, diluting a portion of that 1:10, and so on serially.
2. Use 5uL of each of the above in duplicate (*see* **Note 4.1**) to prepare a standard curve from 100ng to 0.01pg.
3. Use 5uL of each DNA sample being tested (the unknowns), also in duplicate, in a 50uL PCR reaction volume.
4. Prepare a master mix using TaqMan PCR core reagents (ie. Kit No. 4304439) from Applied Biosystems, Inc. (or equivalent) according to the manufacturer's instructions. Each quantitative PCR reaction includes 1X TaqMan PCR buffer, 0.5U AmpErase UNG, 1u*M* Intra-*Alu*Yb8 primers from **subheading 2.1.2**, 100n*M* TaqMan probe from **subheading 2.1.3**, 0.5m*M* dNTPs, 5.0m*M* $MgCl_2$, and 2.5U AmpliTaq Gold DNA polymerase.
5. Add 45 μL master mix to each well containing 5 μL DNA template and carefully seal the optical plate using optical adhesive film (Cat. No. 4360954). Use a plastic sealing spatula or equivalent to avoid touching the optical film with hands.
6. If using an ABI 7000 Prism Sequence Detection System, open a new absolute quantitation file and select the "Setup" icon. Designate each sample well according to standard, unknown, FAM, VIC, etc. Next, select the "Instrument" icon and confirm

the PCR cycling conditions listed in the next step, then save the file before starting the run.

7. Perform quantitative PCR using universal PCR cycling conditions as described: 2 minutes at 50°C for activation of the AmpErase UNG, followed by a denaturation step of 10 minutes at 95°C to activate the AmpliTaq Gold DNA polymerase, then 40 amplification cycles of denaturation at 95°C for 15 seconds and 1 minute of anneal/ extension at 60°C.

8. Following amplification, select the "Results" icon and "amplification plot." Select the wells containing the standards from **Step 1** and drag the green threshold bar until it crosses the amplification signal of the standards in the linear phase of amplification (Fig. 2a). Select "Analyze"

9. The ABI Prism 7000 SDS software will calculate the value of each unknown based on the standard curve DNA concentrations.

10. Export the data from the ABI Prism 7000 SDS software into a Microsoft Excel spreadsheet. Calculate the mean and standard deviation for each point on the standard curve and use the Excel "trendline" option to construct the standard curve. Plot each unknown (mean ± SD) along the standard curve to calculate the DNA quantity (Fig. 2b).

3.1.3. Inference of human geographic origins—

1. Dilute 100 *Alu* stock primers to 2 u*M* in 10m*M* Tris/ 0.1m*M* EDTA (TLE) to make a 10X working solution for each.

2. Perform PCR reactions with at least 10ng of DNA template per 25ul reaction volume for 30-32 cycles using the conditions downloaded in **subheading 2.1.3**.

3. Prepare a 2% agarose gel in 1X TBE buffer containing 0.2ug/mL ethidium bromide. Load 20uL of PCR product on gel, size separate PCR amplicons by electrophoresis at 170 V for 1 hour and visualize the genotypes using UV illumination (Fig. 3).

4. Record genotype data in an Excel spreadsheet as: 1, 1 (homozygous present); 1, 0 (heterozygous); 0, 0 (homozygous absent) in rows for each sample as shown in the reference database downloaded in **subheading 2.1.3**.

5. Prepare the data input file for *Structure* analysis by pasting the genotypes collected for the unknown subjects into a copy of the reference database. Use population code "0" for each unknown.

6. Open *Structure2.2* software by double click on the "gear icon" then select File; New Project (*see* **Note 4.2**).

7. Follow the New Project Wizard steps 1-4: Step 1: Name file and location; Step 2: Number of individuals equals 715 plus the number of your unknowns; Ploidy of data = 2; Number of loci = 100 and Missing data value = −9; Step 3: select box "data file stores data for individuals in a single line"; Step 4: select the following boxes: "individual ID for each individual"; "putative population origin for each individual"; "other extra columns" = 2 for population ID column and continental origin column. Click "Finish" to complete the data input process.

8. Select "Parameter Set"; "New" from the toolbar. Length of Burnin period = 10,000; number of MCMC reps after Burnin = 10,000. On the "Ancestry Model" page select box "Use Population Information." Use the default settings for the remaining tabs and select "OK".

9. Select “Parameter Set”; “Run”; enter number of K populations = 4 at the prompt and hit “OK”

10. The time required to complete the *Structure* analysis run varies depending on your personal computer, the number of individuals being analyzed, and particular parameter settings. Once complete, assess the population assignment of your unknown individuals. Probability of assignment to one of the 4 pre-defined clusters of at least 80% is a strong indicator of individual ancestry.

11. If the probability of assignment to one of the 4 pre-defined clusters is less than 80% with significant admixture from one of more of the other clusters, re-run the *Structure* analysis, first assigning the individual to one cluster, and then to each of the other admixed clusters.

3.2. Taxonomic Applications

3.2.1. Locus identification—Using knowledge of subfamily diagnostic sites, it is a relatively simple task to design primers to experimentally mine a genome with reference to a full genome sequence. The key to this process is to ensure that the diagnostic sites that define the subfamily of interest are well-represented in the primer to be used. Furthermore, it would be advantageous to have the most 3′ base be unique to the subfamily targeted. Several software packages exist to aid in this process. Identification of potentially informative loci involves the generation of ‘half-sites’ from the genomes of interest. Specifically, a linker ligation protocol first suggested by Munroe et al. (93) and refined by Roy et al. (94) and Ray et al. (38) (Fig. 4) is used to clone the sequences neighboring one side of an insertion. This process involves the digestion of genomic DNA in such a way that an overhang is produced. That overhang is matched to a set of annealed linkers, which are ligated to the digested genome fragments.

If you have some information on the consensus sequence of the *Alu* subfamily you are targeting, use an alignment of that subfamily consensus and other *Alu* subfamilies to design primers to be used. Primers should be as specific as possible to the subfamily of interest and preferably end with a subfamily specific base. Standard primer design criteria regarding length, GC content and annealing temperature should be considered (*see* **Notes 4.3** and **4.4**).

1. Perform a restriction digest of the genome of interest. Five hundred nanograms of genomic DNA from each taxon should be digested using a restriction enzyme leaving an appropriate overhand for the linkers to be ligated to the resulting fragments. For example, we often use the enzyme *Nde*I (CA^TATG), which leaves a 5′ TA overhang. *Nde*I is also a good choice because it does not cut within any known *Alu* subfamily and has a six-base restriction site. This provides an advantage over four-base cutters by producing longer fragments and thus, longer flanking sequences for later computational searches of the reference genome(s). Reactions should be conducted in 120 μl volumes and be followed by heat inactivation of the enzyme at 65 °C for 20 minutes.

2. Produce double stranded linkers by incubating 1 nmole of each linker oligonucleotide (top and bottom) at 94°C for 10 minutes in a solution of 2X SSC, 10 m*M* Tris (pH 8). Allow the solution to cool slowly to room temperature. It is important at this point to ensure that your top linker sequences are complementary to the restriction sites that will be produced. For example, when using *Nde*I, we will utilize linkers with the following sequences: *Nde*I_top – TAGAAGGAGAGGACGCTGTCTGTCGAAGG, Universal_bottom – GAGCGAATTCGTCAACATAGCATTTCTGTCCTCTCCTTC. Note the underlined bases in the top linker that will complement the overhang created by the *Nde*I digest.

3. Ligate 12 pmoles of the double stranded linkers 0.25 μg of the digested genomic DNA using the ligase manufacturers protocol.

4. Amplify half-sites in 20 μL reactions consisting of the appropriate 1X buffer, 1.5 m*M* $MgCl_2$, 200 m*M* dNTPs, 0.25 m*M* primers (the *Alu*-specific primer and the linker primer, LNP (5′-GAATTCGTCAACATAGCATTTCT-3′)) and 1.5 U Taq polymerase. Amplification conditions that typically work for us follow this temperature regime: 94°C – 2min, 94°C– 20s, 62°C – 20 s, 72°C – 1min, 10s, for 5 cycles; 94°C – 20 s, 55°C – 20 s, 72°C – 1min, 10 s, for 25 cycles; 72°C – 3min.

5. The PCR products will span a range of sizes. Because smaller products (i.e. products with shorter flanking sequences) will be cloned preferentially, we have found it useful to use gel purification to select for fragments of 500-1000 bp. This ensures that we will obtain enough flanking sequence to increase the probability of finding a single orthologous sequence in the reference genome. We separate the products on a 2% agarose gel and excise the appropriate range. The fragments are then purified using a standard kit such as the Wizard gel purification kit from Promega.

6. Clone the purified PCR products using the TOPO-TA cloning kit for sequencing (Invitrogen) and raise the colonies overnight at 37°C.

7. Select colonies for sequencing by using a sterile toothpick to pick the colonies and incubate in 2-3 ml of Luria Broth (LB) broth overnight with shaking (200rpm) at 37° C.

8. Purify the cultures using any of several standard kits. We typically use the Wizard Plus SV Miniprep kit from Promega.

9. Sequencing is performed using any standard method. Our laboratory utilizes the BigDye sequencing reagents from Applied Biosystems and an ABI 3130xl, also from Applied Biosystems. The objective of this step is to obtain enough sequence to verify the presence of the *Alu* insertion and identify the orthologous location in the reference genome.

10. Once the sequence for any given clone has been obtained, the next task is to identify the orthologous sequence in the reference genome. This is typically possible using the web-based Blast-like Alignment Tool (BLAT) hosted at http://genome.ucsc.edu. The search itself is trivial. However, with the expanding number of primate reference genomes available, the choice of genome is important. As of this writing, reference genomes for human, chimpanzee, and macaque might be used. Simply select the reference genome of most closely related to the taxa of interest and input the sequence from the cloned fragment. One of several possible results will be obtained. It is possible that no orthologous sequence will be retrieved. This is not unexpected as there will have been some evolutionary change since the divergence of the two genomes. Often a query will yield a multiplicity of hits. This is often due to the flanking sequence itself being a repetitive region of the genome. For example, a small percentage of the cloned products will likely contain flanking sequences that consist solely of L1 sequence. Unfortunately, these sequences are unlikely to be of much value when designing primers as the resulting primers will have multiple annealing sites in the genome.

The most productive hits are single, high-scoring hits from the genome of interest in which the flanking sequence is unique. When these are encountered, BLAT can be used to expand the coverage of the genome region to determine two important pieces of information. First, you can immediately discover whether the insertion you recovered from the genome of interest is present in the reference genome. This in itself may be useful information in resolving your phylogeny. Second, you can identify the opposing flanking sequence of the SINE insertion in

the reference genome. Using the opposing flank and the flanking sequence from the genome of interest, oligonucleotides primers can be designed using standard methodologies. Primer design should take into account the potential presence of other mobile elements in the flanks. These should be avoided as priming sites for reasons stated above.

3.2.2. Phylogeny inference—

1. Prepare a panel of template DNAs to perform your phylogenetic analysis. The panel should include all taxa of interest as well as an appropriate outgroup and negative control (water). Template DNA concentration will be variable depending on the standards of individual laboratories but should be consistent among taxa being examined.
2. Perform amplifications on the panel using appropriate conditions for each primer pair designed using the locus identification protocol above (*see* **Subheading 2.2.1.**). Annealing temperatures, $MgCl_2$ concentration, and other factors may differ among primer sets.
3. Use agarose gel electrophoresis to determine insertion patterns for the insertions at each locus. Fig. 4 illustrates one pattern obtained from an analysis of New World primate taxa.
4. Each band should be scored as 1 (insertion present) or 0 (insertion absent) for all taxa for which amplification was obtained.
5. While small, there is the possibility that size variation among taxa can be due to some other event that mimics the pattern expected by presence or absence of the *Alu* being assayed. Thus, some method should be used to verify that the source of any size variation is indeed due to the presence or absence of the *Alu* element. DNA sequence analysis provides the most information on each locus but can be cost-prohibitive. One alternative may be to perform hybridization analysis using probes designed from the *Alu* sequence and from the flanking sequences.
6. A matrix of presence/absence data can be analyzed using any of several available phylogenetic analysis packages including PAUP* (95) and PHYLIP (96). Specific considerations for phylogeny analysis using SINE insertion data has been previously discussed by Okada and colleagues (28).

3.2.3. An *Alu*-based Dichotomous Key—

1. Dilute *Alu* stock primers from Herke et al. (86) to 2 μ*M* in 10m*M* Tris/ 0.1m*M* EDTA (TLE) to make a 10X working solution for each. The sequences are available for free download at: (http://www.batzerlab.lsu.edu ; publication #190, Supplementary Data).
2. Perform PCR reactions with at least 10ng of DNA template per 25uL reaction volume for 30-32 cycles using the conditions downloaded in **subheading 2.2.3**.
3. Prepare a 2% agarose gel in 1X TBE buffer containing 0.2ug/mL ethidium bromide. Load 20ul of PCR product on gel, size separate PCR amplicons by electrophoresis and visualize the genotypes using UV illumination.

3.3. Patterns and Processes of Transpositions in Cultured Cells and within a Genome

3.3.1. The transient cultured cell retrotransposition assay (75)—

1. Plate HeLa cells at 2×10^5 HeLa cells/well in 6-well plates and culture for ~8-14 hours at 37 °C an atmosphere containing 5-7% CO_2 and 100% humidity in high glucose DMEM lacking pyruvate (GIBCO®).

2. Transfect cells with 3μL of FuGene 6 nonliposomal transfection reagents (Roche) and 1 μg of DNA per transfection of HeLa cell in 6-well plates.
3. Co-transfect one set of 6-well plates with equal amount of a reporter plasmid (pGreen Lantern) and an L1 allele tagged with the *mneol* indicator cassette while the others are transfected with only the L1 construct.
4. At three days post-transfection, trypsinize the HeLa cells in the first set of plates and analyze them by using flow cytometry.
    a. Remove the spent media with a sterile Pasteur pipette.
    b. Wash the cells with 2 mL PBS (one or two times).
    c. Remove the PBS and add 0.2 – 0.3 mL of a modified Versense solution (5m*M* EDTA in PBS) which has been pre-warmed to 37°C, and then incubate the plates for 10 minutes.
    d. Gently remove the adherent cells from the 6-well plates.
    e. Transfer the cell suspension to polystyrene tubes by passage through cells snap caps (Falcon) and keep them on ice until flow cytometry analysis.
    f. Quantify the cells with a Becton Dickinson flow cytometer using a 15 mWatt argon ion laser (488 nm) and fluorescein filter sets (530/30 bandpass).
    g. Perform data analysis using the CellQuest software.
    h. The percentage of GFP cells is used to determine the transfection efficiency of each sample.
5. Seed the remaining set at 2×10^5 cells/well in 6-well plates and add 400 μg/mL of G418 to the cells for the selection of L1 retrotransposition.
6. Aspirate the selection media after 12 days (daily re-feeding) and wash the cells in 1 × PBS.
7. Fix the $G418^R$ foci by incubation in FIX solution for 30 minutes at 4 °C.
8. Stain the fixed cells for 30 minutes with crystal violet (0.2% crystal violet in 5% acetic acid, 2.5% isopropanol) at room temperature and washed with PBS.
9. Determine the retrotransposition efficiency (the number of $G418^R$ foci/the number of transfected cells) using an Oxford Optronics ColCount colony counter.

3.3.2. L1-Mediated *Alu* Retrotransposition in Cultured Human Cells—

1. Plate HeLa cells at 5×10^5 HeLa cell/60-mm dish and grow ~8-14 hours at 37 °C an atmosphere containing 5-7% CO_2 and 100% humidity in high glucose DMEM lacking pyruvate (GIBCO®).
2. Co-transfect the dish with 12 μL Lipofectamine and 8 μL Reagent (GIBCO®), 2 μg neo^{Tet}-marked *Alu* and 2 μg CMV L1-RP expression vector.
3. For seven days, culture and seed the transfected cells at 5×10^5 cells/100-mm dish.
4. Add 560 μg/mL of Geneticin (GIBCO®) to the cells for G418 selection.
5. After 14 days (daily re-feeding), aspirate the selection media and wash the cells in 1X PBS.
6. Fix the $G418^R$ foci by incubation in FIX solution for 30 minutes at 4 °C.

7. Stain the fixed cells for 30 minutes with crystal violet (0.2% crystal violet in 5% acetic acid, 2.5% isopropanol) at room temperature and washed with PBS.
8. Determine the retrotransposition efficiency (the number of $G418^R$ foci/the number of transfected cells) using an Oxford Optronics ColCount colony counter.

3.3.3. Rescue of L1 integrants from $G418^R$ foci (Fig. 5) (74,92)—

1. Extract HeLa genomic DNA from either a single $G418^R$ focus or small pool (10 to 250) of $G418^R$ foci by using the Puregene cell and tissue DNA isolation kit (Gentra).
2. Perform a restriction digest of the genomic DNA as follow:
    a. 10 μg of genomic DNA
    b. 5 μL of 10 × buffer (specific to restriction enzyme)
    c. 20U of restriction enzyme (HindIII, BglII, BclI or BamHI (New England Biolabs))
    d. Add distilled water for a final volume of 50 μL
3. Put the tube in a thermomixer or water bath at 37°C for 2 hours (or overnight).
4. Inactivate the restriction enzyme by heating or the Wizard DNA cleanup kit (Promega).
5. Dilute the digested genomic DNA pieces.
6. Prepare the self ligation reaction as follow: (intra-molecule, hopefully)
    a. 5 μL of the digested genomic DNA (from step 5)
    b. 50 μL of 10X ligation buffer
    c. 1 μL of T4 DNA ligase (New England Biolabs)
    d. Add distilled water for a final volume of 500 μL
7. Put the tube overnight at 14 °C.
8. Centrifuge the ligation mixtures through a Microcon-100 at 500 × g for 14 minutes.
9. Transform 1mL of XL1-Blue MRF' $CaCl_2$ competent cells (efficiencies of >1 × 10^8 cfu/μg; Stratagene) with the total concentrated ligation (~1 μg) or perform the inverse PCR using the ligation as a template.
10. Several transformants are visible after overnight growth at 37 °C on LB agar plates with 50 μg/mL kanamycin.
11. Extract the plasmid DNA from the resistant clones and then perform restriction mapping, PCR or DNA sequencing analyses.

The pre-integration site of a *de novo* L1 insert would be identified by searching BLAT (e.g., hg18; Mar. 2006 freeze) with its each upstream and downstream flanking sequence obtained from above rescue procedure. The acquisition of pre-integration sequences confers the opportunity of additional analyses such as endonuclease cleavage sites, TSD structures, and target sequence alterations derived from the L1 retrotransposition.

4. Notes

1. Quantitative PCR is best prepared using a single channel pipette, electronic repeater style is ideal for duplicates. A typical multi-channel pipette is typically not consistent enough between channels for accurate duplicates.
2. Once *Structure* is open, do not minimize the window. If you leave *Structure* to use another application, minimize that application window when finished and you will return to the *Structure* sub-window to continue the set up. DO NOT click on the *Structure* window – it won't let you continue because the initial sub-window is still open.
3. Primer3: The Primer3 software (http://frodo.wi.mit.edu/) (97) is a useful web interface for designing oligonucleotides primers. The software offers users a variety of options such as the size range of PCR product, primer size, GC content of the primer, and oligonucleotide-melting temperature, which allows users to design optimal primers. In addition, it is linked with BLAST-Like Alignment Tool (BLAT) web browser (http://genome.ucsc.edu/cgi-bin/hgBlat) showing the genomic positions and sequences of PCR products which could be amplified by the primers and thus users are readily able to figure out whether the primers are accurate or specific to their genomic target region.
4. OligoCalc (Oligonucleotide Properties Calculator): The OligoCalc (98) is a web-accessible (http://www.basic.northwestern.edu/biotools/oligocalc.html) and estimates properties for single-stranded DNA and RNA. Features important to consider include self-complementarity, potential hairpin loop formation, and oligonucleotide-melting temperature with and without salt conditions.

Acknowledgments

Our research is supported by National Science Foundation BCS-0218338 (MAB) and EPS-0346411 (MAB), National Institutes of Health RO1 GM59290 (MAB) and PO1 AG022064 (MAB), and the State of Louisiana Board of Regents Support Fund (MAB). DAR is supported by the Eberly College of Arts and Sciences at West Virginia State University.

References

1. 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, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921. [PubMed: 11237011]
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, Antonarakis SE, Attwood J, Baertsch R, Bailey J, Barlow K, Beck S, Berry E, Birren B, Bloom T, Bork P, Botcherby M, Bray N, Brent MR, Brown DG, Brown SD, Bult C, Burton J, Butler J, Campbell RD, Carninci P, Cawley S, Chiaromonte F, Chinwalla AT, Church DM, Clamp M, Clee C, Collins FS, Cook LL, Copley RR, Coulson A, Couronne O, Cuff J, Curwen V, Cutts T, Daly M, David R, Davies J, Delehaunty KD, Deri J, Dermitzakis ET, Dewey C, Dickens NJ, Diekhans M, Dodge S, Dubchak I, Dunn DM, Eddy SR, Elnitski L, Emes RD, Eswara P, Eyras E, Felsenfeld A, Fewell GA, Flicek P, Foley K, Frankel WN, Fulton LA, Fulton RS, Furey TS, Gage D,

Gibbs RA, Glusman G, Gnerre S, Goldman N, Goodstadt L, Grafham D, Graves TA, Green ED, Gregory S, Guigo R, Guyer M, Hardison RC, Haussler D, Hayashizaki Y, Hillier LW, Hinrichs A, Hlavina W, Holzer T, Hsu F, Hua A, Hubbard T, Hunt A, Jackson I, Jaffe DB, Johnson LS, Jones M, Jones TA, Joy A, Kamal M, Karlsson EK, et al. Initial sequencing and comparative analysis of the mouse genome. Nature 2002;420:520–62. [PubMed: 12466850]

3. Berg, DE.; Howe, MM., editors. Mobile DNA. American Society for Microbiology; Washington, DC: 1989.
4. Pace JK 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Research 2007;17:422–32. [PubMed: 17339369]
5. Deininger PL, Batzer MA. Evolution of retroposons. Evolutionary Biology 1993;27:157–196.
6. Batzer MA, Deininger PL. *Alu* repeats and human genomic diversity. Nature Reviews Genetics 2002;3:370–9.
7. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab 1999;67:183–93. [PubMed: 10381326]
8. Kajikawa M, Okada N. LINEs mobilize SINEs in the eel through a shared 3′ sequence. Cell 2002;111:433–44. [PubMed: 12419252]
9. Kazazian HH Jr. Moran JV. The impact of L1 retrotransposons on the human genome. Nature Genetics 1998;19:19–24. [PubMed: 9590283]
10. Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 1993;72:595–605. [PubMed: 7679954]
11. Wang W, Kirkness EF. Short interspersed elements (SINEs) are a major source of canine genomic diversity. Genome Res 2005;15:1798–808. [PubMed: 16339378]
12. Sinnett D, Richer C, Deragon JM, Labuda D. *Alu* RNA transcripts in human embryonal carcinoma cells. Model of post-transcriptional selection of master sequences. Journal of Molecular Biology 1992;226:689–706. [PubMed: 1507221]
13. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. American Journal of Human Genetics 2003;73:1444–51. [PubMed: 14628287]
14. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nature Genetics 2003;35:41–8. [PubMed: 12897783]
15. Boeke JD. LINEs and Alus--the polyA connection. Nature Genetics 1997;16:6–7. [PubMed: 9140383]
16. Feng Q, Moran JV, Kazazian HH Jr. Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 1996;87:905–16. [PubMed: 8945517]
17. Jurka J. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proceedings of the National Academy of Sciences of the United States of America 1997;94:1872–7. [PubMed: 9050872]
18. Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH. Master genes in mammalian repetitive DNA amplification. Trends in Genetics 1992;8:307–11. [PubMed: 1365396]
19. Cordaux R, Hedges DJ, Batzer MA. Retrotransposition of Alu elements: how many sources? Trends in Genetics 2004;20:464–7. [PubMed: 15363897]
20. Matera AG, Hellmann U, Hintz MF, Schmid CW. Recently transposed Alu repeats result from multiple source genes. Nucleic Acids Research 1990;18:6019–23. [PubMed: 2172925]
21. Shen MR, Batzer MA, Deininger PL. Evolution of the master *Alu* gene(s). Journal of Molecular Evolution 1991;33:311–20. [PubMed: 1774786]
22. 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. Genetic variation of recent *Alu* insertions in human populations. Journal of Molecular Evolution 1996;42:22–9. [PubMed: 8576959]
23. Kapitonov V, Jurka J. The age of *Alu* subfamilies. Journal of Molecular Evolution 1996;42:59–65. [PubMed: 8576965]

24. 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. Large-scale analysis of the *Alu* Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. Journal of Molecular Biology 2001;311:17–40. [PubMed: 11469855]
25. Carter AB, Salem AH, Hedges DJ, Keegan CN, Kimball B, Walker JA, Watkins WS, Jorde LB, Batzer MA. Genome-wide analysis of the human Alu Yb-lineage. Human Genomics 2004;1:167–78. [PubMed: 15588477]
26. Han K, Xing J, Wang H, Hedges DJ, Garber RK, Cordaux R, Batzer MA. Under the genomic radar: the stealth model of Alu amplification. Genome Research 2005;15:655–64. [PubMed: 15867427]
27. 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. Analysis of the human Alu Ya-lineage. Journal of Molecular Biology 2004;342:109–18. [PubMed: 15313610]
28. Okada, N.; Shedlock, AM.; Nikaido, M. Mobile Genetic Elements: Protocols and Genomic Applications. Vol. Vol. 260. Humana Press; Totowa, NJ: 2004. Retroposon mapping in molecular systematics; p. 189-226.
29. Ray DA. SINEs of progress: Mobile element applications to molecular ecology. Molecular Ecology 2007;16:19–33. [PubMed: 17181718]
30. Ray DA, Xing J, Salem A-H, Batzer MA. SINEs of a *nearly* perfect character. Systematic Biology 2006;55:928–935. [PubMed: 17345674]
31. Shedlock AM, Okada N. SINE insertions: powerful tools for molecular systematics. Bioessays 2000;22:148–60. [PubMed: 10655034]
32. Shedlock AM, Takahashi K, Okada N. SINEs of speciation: tracking lineages with retroposons. Trends in Ecology and Evolution 2004;19:545–553. [PubMed: 16701320]
33. Zietkiewicz E, Richer C, Labuda D. Phylogenetic affinities of tarsier in the context of primate Alu repeats. Molecular Phylogenetics and Evolution 1999;11:77–83. [PubMed: 10082612]
34. Watanabe M, Nikaido M, Tsuda TT, Inoko H, Mindell DP, Murata K, Okada N. The rise and fall of the CR1 subfamily in the lineage leading to penguins. Gene 2006;365:57–66. [PubMed: 16368202]
35. Takahashi K, Terai Y, Nishida M, Okada N. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Molecular Biology and Evolution 2001;18:2057–66. [PubMed: 11606702]
36. Sasaki T, Takahashi K, Nikaido M, Miura S, Yasukawa Y, Okada N. First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. Molecular Biology and Evolution 2004;21:705–15. [PubMed: 15014157]
37. Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA. *Alu* elements and hominid phylogenetics. Proceedings of the National Academy of Sciences of the United States of America 2003;100:12787–91. [PubMed: 14561894]
38. 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. Alu insertion loci and platyrrhine primate phylogeny. Molecular Phylogenetics and Evolution 2005;35:117–26. [PubMed: 15737586]
39. Murata S, Takasaki N, Saitoh M, Okada N. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proceedings of the National Academy of Sciences of the United States of America 1993;90:6995–9. [PubMed: 8346208]
40. Cordaux R, Lee J, Dinoso L, Batzer MA. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. Gene 2006;373:138–44. [PubMed: 16527433]
41. Kass DH. Generation of human DNA profiles by *Alu*-based multiplex polymerase chain reaction. Analytical Biochemistry 2003;321:146–9. [PubMed: 12963068]
42. Thomas E, Herrera RJ. Multiplex polymerase chain reaction of *Alu* polymorphic insertions. Electrophoresis 1998;19:2373–9. [PubMed: 9820952]
43. Flavell AJ, Knox MR, Pearce SR, Ellis TH. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 1998;16:643–50. [PubMed: 10036780]
44. Salem AH, Ray DA, Hedges DJ, Jurka J, Batzer MA. Analysis of the human Alu Ye lineage. BMC Evolutionary Biology 2005;5:18. [PubMed: 15725352]

45. Xing J, Salem AH, Hedges DJ, Kilroy GE, Watkins WS, Schienman JE, Stewart CB, Jurka J, Jorde LB, Batzer MA. Comprehensive analysis of two *Alu* Yd subfamilies. Journal of Molecular Evolution 2003;57:S76–S89. [PubMed: 15008405]
46. Shewale JG, Schneida E, Wilson J, Walker JA, Batzer MA, Sinha SK. Human genomic DNA quantitation system, H-Quant: development and validation for use in forensic casework. J Forensic Sci 2007;52:364–70. [PubMed: 17316233]
47. Nicklas JA, Buel E. Development of an Alu-based, real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 2003;48:936–44. [PubMed: 14535658]
48. Nicklas JA, Buel E. Quantification of DNA in forensic samples. Anal Bioanal Chem 2003;376:1160–7. [PubMed: 12739098]
49. Nicklas JA, Buel E. An Alu-based, MGB Eclipse real-time PCR method for quantitation of human DNA in forensic samples. J Forensic Sci 2005;50:1081–90. [PubMed: 16225212]
50. Nicklas JA, Buel E. Simultaneous determination of total human and male DNA using a duplex real-time PCR assay. J Forensic Sci 2006;51:1005–15. [PubMed: 17018077]
51. Sifis ME, Both K, Burgoyne LA. A more sensitive method for the quantitation of genomic DNA by *Alu* amplification. Journal of Forensic Sciences 2002;47:589–92. [PubMed: 12051342]
52. Walker JA, Kilroy GE, Xing J, Shewale J, Sinha SK, Batzer MA. Human DNA quantitation using *Alu* element-based polymerase chain reaction. Analytical Biochemistry 2003;315:122–8. [PubMed: 12672420]
53. Walker JA, Hedges DJ, Perodeau BP, Landry KE, Stoilova N, Laborde ME, Shewale J, Sinha SK, Batzer MA. Multiplex polymerase chain reaction for simultaneous quantitation of human nuclear, mitochondrial, and male Y-chromosome DNA: application in human identification. Anal Biochem 2005;337:89–97. [PubMed: 15649380]
54. Hedges DJ, Walker JA, Callinan PA, Shewale JG, Sinha SK, Batzer MA. Mobile element-based assay for human gender determination. Analytical Biochemistry 2003;312:77–9. [PubMed: 12479838]
55. Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, Christian AT, Turteltaub K, Batzer MA. Inference of human geographic origins using Alu insertion polymorphisms. Forensic Science International 2005;153:117–24. [PubMed: 16139099]
56. Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB. Human population genetic structure and inference of group membership. Am J Hum Genet 2003;72:578–89. [PubMed: 12557124]
57. Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. Am J Hum Genet 2000;66:979–88. [PubMed: 10712212]
58. Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL. Alu insertion polymorphisms for the study of human genomic diversity. Genetics 2001;159:279–90. [PubMed: 11560904]
59. Salem AH, Kilroy GE, Watkins WS, Jorde LB, Batzer MA. Recently integrated Alu elements and human genomic diversity. Mol Biol Evol 2003;20:1349–61. [PubMed: 12777511]
60. 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. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. Genome Res 2003;13:1607–18. [PubMed: 12805277]
61. CSAC. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 2005;437:69–87. [PubMed: 16136131]
62. 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, et al. Evolutionary and biomedical insights from the rhesus macaque genome. Science 2007;316:222–34. [PubMed: 17431167]

63. Kohn M, Knauer F, Stoffella A, Schroder W, Paabo S. Conservation genetics of the European brown bear--a study using excremental PCR of nuclear and mitochondrial sequences. Mol Ecol 1995;4:95–103. [PubMed: 7711958]
64. Matsubara M, Basabose AK, Omari I, Kaleme K, Kizungu B, Sikubwabo K, Kahindo M, Yamagiwa J, Takenaka O. Species and sex identification of western lowland gorillas (Gorilla gorilla gorilla), eastern lowland gorillas (Gorilla beringei graueri) and humans. Primates 2005;46:199–202. [PubMed: 15635453]
65. Taberlet P, Camarra JJ, Griffin S, Uhres E, Hanotte O, Waits LP, Dubois-Paganon C, Burke T, Bouvet J. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. Mol Ecol 1997;6:869–76. [PubMed: 9301075]
66. Yan P, Wu XB, Shi Y, Gu CM, Wang RP, Wang CL. Identification of Chinese alligators (Alligator sinensis) meat by diagnostic PCR of the mitochondrial cytochrome b gene. Biological Conservation 2005;121:45–51.
67. Domingo-Roura X, Marmi J, Ferrando A, Lopez-Giraldez F, Macdonald DW, Jansman HAH. Badger hair in shaving brushes comes from protected Eurasian badgers. Biological Conservation 2006;128:425–430.
68. Han K, Lee J, Meyer TJ, Wang J, Sen SK, Srikanta D, Liang P, Batzer MA. Alu recombination-mediated structural deletions in the chimpanzee genome. PLoS Genet 2007;3:1939–49. [PubMed: 17953488]
69. Khan H, Smit A, Boissinot S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. Genome Res 2006;16:78–87. [PubMed: 16344559]
70. Lee J, Cordaux R, Han K, Wang J, Hedges DJ, Liang P, Batzer MA. Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. Gene 2007;390:18–27. [PubMed: 17055192]
71. Callinan PA, Wang J, Herke SW, Garber RK, Liang P, Batzer MA. Alu retrotransposition-mediated deletion. Journal of Molecular Biology 2005;348:791–800. [PubMed: 15843013]
72. Han K, Sen SK, Wang J, Callinan PA, Lee J, Cordaux R, Liang P, Batzer MA. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. Nucleic Acids Res 2005;33:4040–52. [PubMed: 16034026]
73. Sen SK, Han K, Wang J, Lee J, Wang H, Callinan PA, Dyer M, Cordaux R, Liang P, Batzer MA. Human Genomic Deletions Mediated by Recombination between Alu Elements. Am J Hum Genet 2006;79:41–53. [PubMed: 16773564]
74. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. Cell 1996;87:917–27. [PubMed: 8945518]
75. Wei W, Morrish TA, Alisch RS, Moran JV. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem 2000;284:435–8. [PubMed: 10964437]
76. Dombroski BA, Mathias SL, Nanthakumar E, Scott AF, Kazazian HH Jr. Isolation of an active human transposable element. Science 1991;254:1805–8. [PubMed: 1662412]
77. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. Proceedings of the National Academy of Sciences of the United States of America 2003;100:5280–5. [PubMed: 12682288]
78. Gilbert N, Lutz S, Morrish TA, Moran JV. Multiple fates of L1 retrotransposition intermediates in cultured human cells. Mol Cell Biol 2005;25:7780–95. [PubMed: 16107723]
79. Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. Cell 2002;110:315–25. [PubMed: 12176319]
80. Symer DE, Connelly C, Szak ST, Caputo EM, Cost GJ, Parmigiani G, Boeke JD. Human l1 retrotransposition is associated with genetic instability in vivo. Cell 2002;110:327–38. [PubMed: 12176320]

81. Moran JV, DeBerardinis RJ, Kazazian HH Jr. Exon shuffling by L1 retrotransposition. Science 1999;283:1530–4. [PubMed: 10066175]
82. Cost GJ, Boeke JD. Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. Biochemistry 1998;37:18081–93. [PubMed: 9922177]
83. 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. Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. Genome Research 2003;13:1607–18. [PubMed: 12805277]
84. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155:945–59. [PubMed: 10835412]
85. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164:1567–87. [PubMed: 12930761]
86. Herke SW, Xing J, Ray DA, Zimmerman JW, Cordaux R, Batzer MA. A SINE-based dichotomous key for primate identification. GENE 2007;390
87. Holmes SE, Dombroski BA, Krebs CM, Boehm CD, Kazazian HH Jr. A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. Nat Genet 1994;7:143–8. [PubMed: 7920631]
88. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS. A de novo Alu insertion results in neurofibromatosis type 1. Nature 1991;353:864–6. [PubMed: 1719426]
89. Ullu E, Tschudi C. *Alu* sequences are processed 7SL RNA genes. Nature 1984;312:171–2. [PubMed: 6209580]
90. Dhellin O, Maestre J, Heidmann T. Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. Embo J 1997;16:6590–602. [PubMed: 9351839]
91. Kimberland ML, Divoky V, Prchal J, Schwahn U, Berger W, Kazazian HH Jr. Full-length human L1 insertions retain the capacity for high frequency retrotransposition in cultured cells. Hum Mol Genet 1999;8:1557–60. [PubMed: 10401005]
92. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nature Genetics 2002;31:159–65. [PubMed: 12006980]
93. Munroe DJ, Haas M, Bric E, Whitton T, Aburatani H, Hunter K, Ward D, Housman DE. IRE-bubble PCR: a rapid method for efficient and representative amplification of human genomic DNA sequences from complex sources. Genomics 1994;19:506–14. [PubMed: 8188293]
94. Roy AM, Carroll ML, Kass DH, Nguyen SV, Salem AH, Batzer MA, Deininger PL. Recently integrated human *Alu* repeats: finding needles in the haystack. Genetica 1999;107:149–61. [PubMed: 10952208]
95. Swofford, DL. PAUP: Phylogenetic Analysis Using Parsimony (*and Other Methods). 4.0b10 edit. Sinauer Associates; Sunderland, Massachusetts: 2002.
96. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 1989;5:164–166.
97. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 2000;132:365–86. [PubMed: 10547847]
98. Kibbe WA. OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 2007;35:W43–6. [PubMed: 17452344]

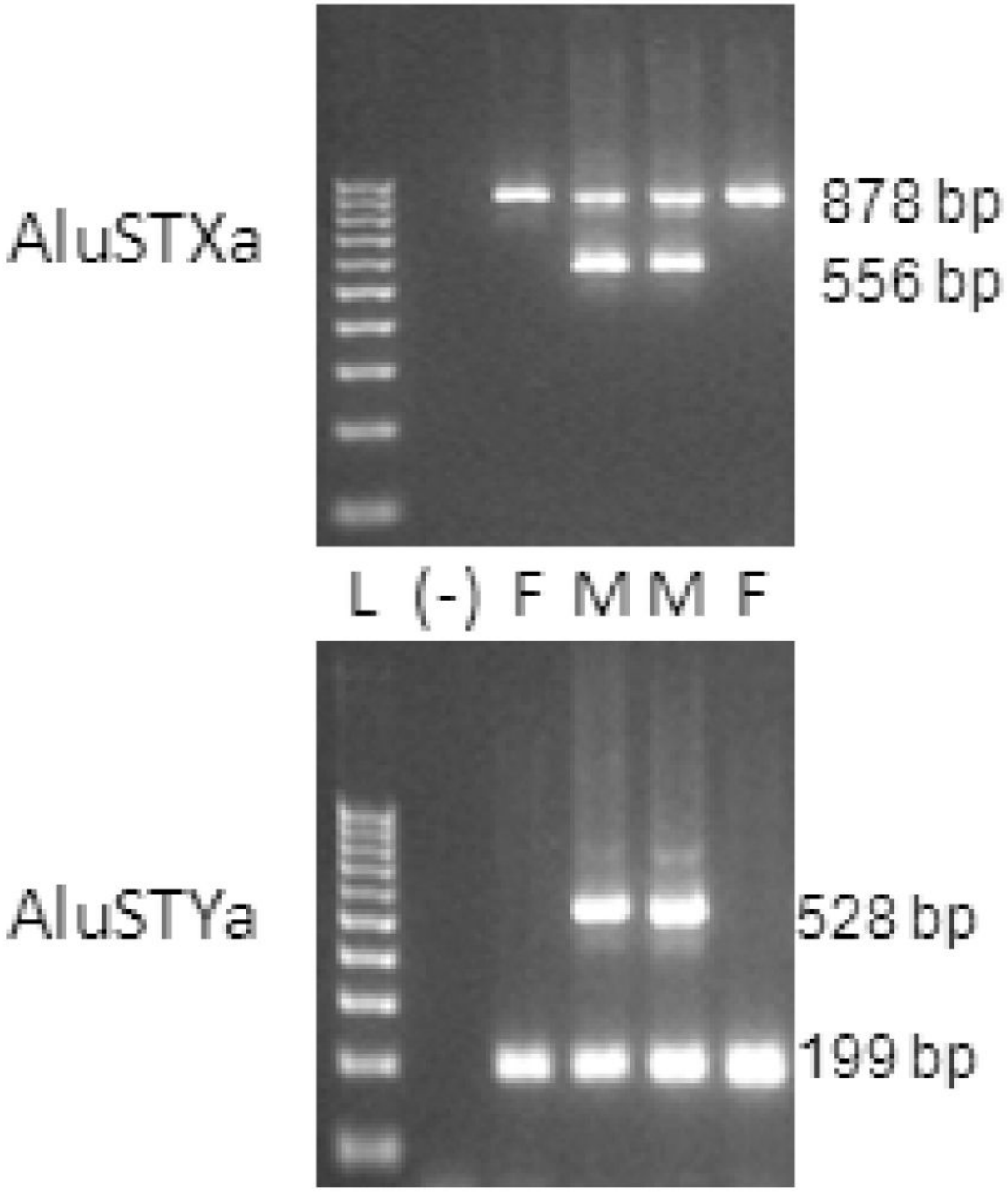

Figure 1.
Mobile element-based human gender determination. An agarose gel chromatograph from the analysis of four individuals using the genetic systems. *AluSTXa* and *AluSTYa* loci are shown. Males are distinguished by the presence of two DNA fragments, while females have a single amplicon. F (female) and M (male) on each lane indicate the gender. L – 100 bp DNA ladder, (−) – negative control consisting of a water template.

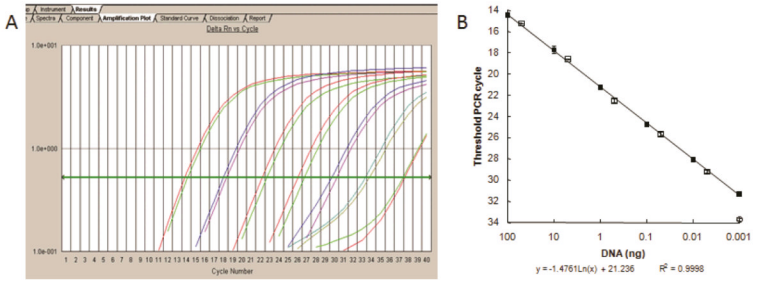

Figure 2.

Quantitative PCR using *Alu* subfamily-specific amplification. Example of a 10-fold serial dilution of DNA duplicates using the ABI Prism 7000 Sequence Detection System. (A) Fluorescent signal is plotted against PCR cycle number. The threshold cycle (Ct) is defined as the cycle at which the signal crosses the threshold (represented by the horizontal line) during the linear phase of amplification. (B) Ct values are exported into a Microsoft Excel spreadsheet where the mean and standard deviation are calculated for each point on the standard curve. Unknown DNA samples are quantified by comparison to the standard curve.

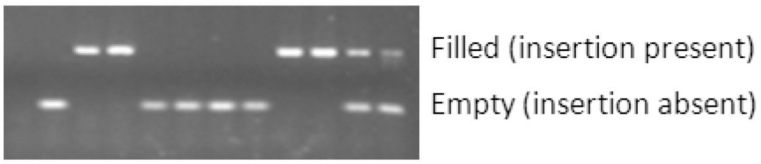

Figure 3.
Gel electrophoresis results using *Alu* insertion loci for human geographic affiliation analysis. The upper band is seen for filled sites and the lower band for empty sites. Individuals exhibiting two bands are presumed to be heterozygous. Individuals for whom only one band is visible indicate a homozygous genotype for either of the alternative states for the locus.

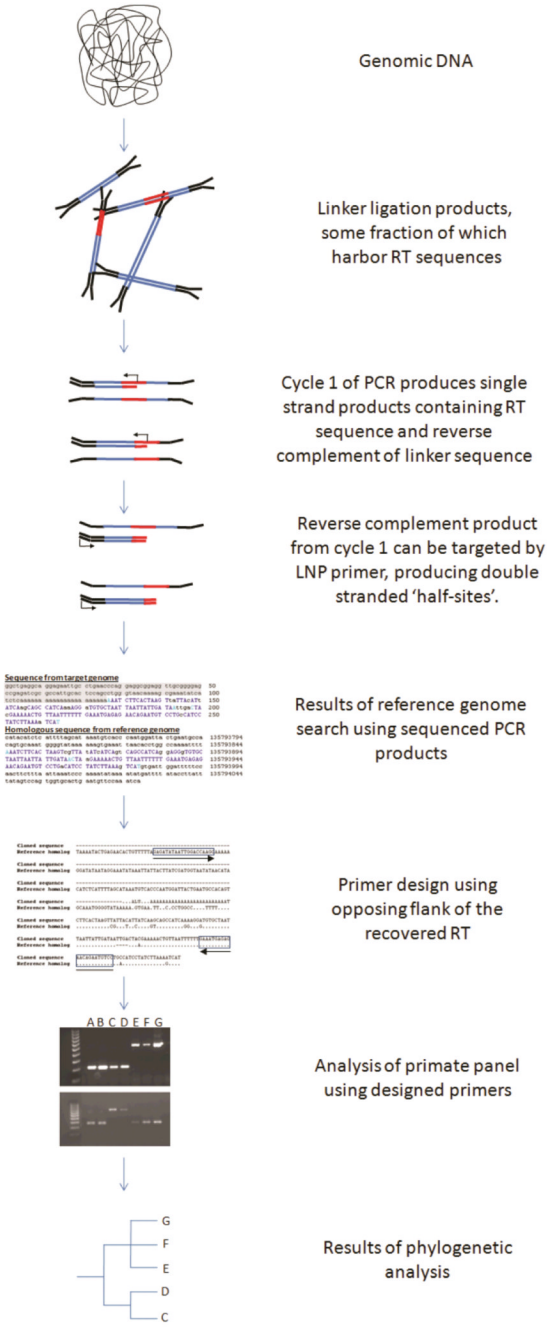

Figure 4.
Schematic representing the steps required to perform a *de novo* phylogenetic analysis using *Alu* insertion loci.

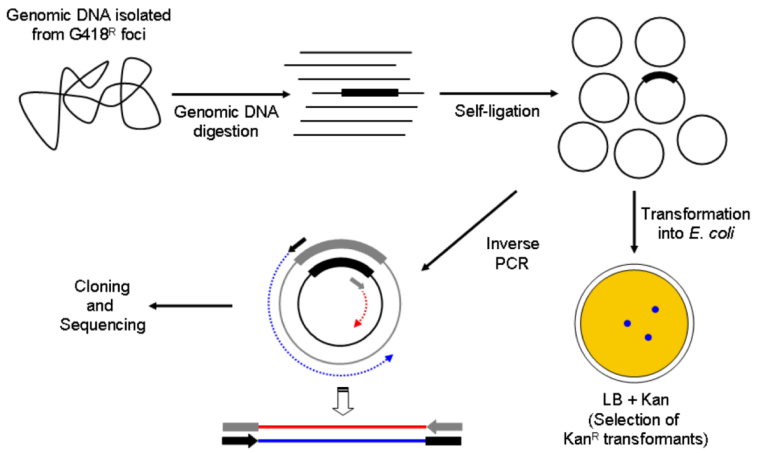

Figure 5.
Rescue of L1 integrants from G418^R foci (74,92).
Genomic DNA is isolated from HeLa cells that have G418-resistance (G418^R) derived from *de novo* L1 inserts. The new L1 elements (thick black boxes) are recovered by either transforming into *E.coli* or performing an inverse PCR.