Analysis of western lowland gorilla (Gorilla gorilla gorilla) specific Alu repeats

McLain et al.
Analysis of western lowland gorilla (Gorilla gorilla gorilla) specific Alu repeats

Adam T McLain, Glenn W Carman, Mitchell L Fullerton, Thomas O Beckstrom, William Gensler, Thomas J Meyer, Christopher Faulk and Mark A Batzer*

Abstract

Background: Research into great ape genomes has revealed widely divergent activity levels over time for Alu elements. However, the diversity of this mobile element family in the genome of the western lowland gorilla has previously been uncharacterized. Alu elements are primate-specific short interspersed elements that have been used as phylogenetic and population genetic markers for more than two decades. Alu elements are present at high copy number in the genomes of all primates surveyed thus far. The AluY subfamily and its derivatives have been recognized as the evolutionarily youngest Alu subfamily in the Old World primate lineage.

Results: Here we use a combination of computational and wet-bench laboratory methods to assess and catalog AluY subfamily activity level and composition in the western lowland gorilla genome (gorGor3.1). A total of 1,075 independent AluY insertions were identified and computationally divided into 10 subfamilies, with the largest number of gorilla-specific elements assigned to the canonical AluY subfamily.

Conclusions: The retrotransposition activity level appears to be significantly lower than that seen in the human and chimpanzee lineages, while higher than that seen in orangutan genomes, indicative of differential Alu amplification in the western lowland gorilla lineage as compared to other Homininae.

Keywords: Primate, Gorilla, SINE, Retrotransposon, Mobile elements

Background

Alu elements are a family of primate-specific SINEs (Short INterspersed Elements) of approximately 300 base pairs (bp) long and present in the genomes of all living primates [1-3]. Alu elements were derived from 7SL RNA, the RNA component of the signal recognition particle, in the common ancestor of all living primates [4]. In the past approximately 65 million years Alu elements have become widely distributed in primate genomes [1,5]. Alu elements are now present at copy numbers of >1,000,000 in all surveyed great ape genomes (Additional file 1) [1]. Despite their high copy number the majority of Alu elements are genomic fossils, non-propagating relics passed down over millions of years after earlier periods of replicative activity [1,6]. It is hypothesized that a relatively small number of ‘master’ elements are responsible for the continued spread of all active subfamilies [7,8].

As non-autonomous retrotransposons, Alu elements do not encode the enzymatic machinery necessary for self-propagation [1,2]. This is accomplished by appropriating the replication machinery [2,9] of a much larger, autonomous retrotransposon called LINE1 (L1) via a process termed target-primed reverse transcription (TPRT) [10-13].

The effective use of SINEs as phylogenetic markers was first demonstrated in 1993 in a study seeking to resolve relationships between Pacific salmonid species [14]. Subsequent to this study, SINE-based phylogenetic methods have been applied across a wide range of species to determine evolutionary relationships [15,16]. In particular, Alu elements have proven to be extremely useful tools for elucidating evolutionary relationships between primate species [1,17]. The essentially homoplasy free presence of an Alu element of the same subfamily at a given locus between two or more primate species is almost always definitive evidence of shared ancestry [18]. The possibility of confounding events is very small, and
easily resolved by the sequencing and examining of the element in question [1,18]. In the past 15 years Alu-based phylogenetic methods have been used with great success to resolve evolutionary relationships among the Tarsiers [19,20], New World [21] and Old World monkeys [22-24], gibbons [25], lemurs [26,27], and great apes [28].

In addition to phylogenetic applications Alu elements also function as effective markers for the study of population genetics via examination of polymorphic elements between members of the same species [2,29,30]. Alu elements are also linked to numerous genetic diseases, and the insertion of an element at an important genomic location can have grave consequences for the individual involved [3,31,32]. Additionally, Alu elements are thought to be a causal factor in genomic instability [33-36].

**Alu** elements are classified in multiple major subfamilies and numerous smaller, derivative subfamilies based on specific sequence mutations [37-40]. All extant primates share older elements, while all primate lineages examined also have younger, lineage-specific subfamilies [41]. Alu subfamily evolution is parallel, not linear, and various subfamilies have been found to be actively retrotransposing at the same time in all primate genomes surveyed; each primate lineage thus possesses its own Alu subfamilies [1,42,43].

The AluJ subfamily is the most ancient Alu lineage, and was largely active from approximately 65 million years ago to approximately 55 million years ago, at which point AluS evolved and supplanted AluJ as the predominant active subfamily [37,41]. Due to the antiquity of the lineage, AluJ subfamilies are present in all extant primates, including Strepsirrhines [27,44]. AluS, on the other hand, evolved from AluJ after the Strepsirrhine-Haplorrhine divergence, and so is only found in New World and Old World primates [2,37,45]. The AluX subfamily subsequently evolved from AluS in the Old World primate lineage, and remains the predominant active subfamily in catarrhines [1,41,45].

A number of AluY-derived subfamilies continue to be active in great apes [1], and polymorphic lineage-specific Alu elements have been well documented between existing human populations [2], indicating a continued activity level for these mobile elements. A rate of one new element in every approximately 20 live births has been proposed as the current rate of Alu element activity in the extant human population, but the large size of this population coupled with human generation time would make it very difficult for new elements to come to fixation outside of small population groups [46,47]. Research into Alu element activity in Sumatran and Bornean orangutans has indicated a comparatively low-level of continued retrotransposition activity in these apes [48], suggesting some alteration of the propagation of Alu within this lineage [49].

The western lowland gorilla (*Gorilla gorilla gorilla*), a subspecies of the western gorilla (*Gorilla gorilla*), is a critically endangered great ape endemic to the forests and lowland swamps of central Africa [50,51]. Western lowland gorillas are gregarious, living in family groups comprised of a dominant male, multiple females, subadult males, and juvenile offspring [52]. Western lowland gorillas are in danger of extinction due to human activity. Their wild population size is shrinking in the face of anthropogenic pressure and diseases such as Ebola [50]. Gorillas are a close evolutionary relative of humans and the *Pan* lineage of chimpanzees and bonobos, with the most widely accepted date for a common ancestor 6 to 9 million years ago [28,53-55], though a date as early as 10 million years ago has been recently proposed [56].

The genome of ‘Kamilah’, a female western lowland gorilla living at the San Diego Zoo, was initially assembled from 5.4 Gbp of capillary sequence and 166.8 Gbp of Illumina read pairs, and further refined using bacterial artificial chromosome (BAC) and fosmid end pair capillary technology [57]. This sequence is available from the Wellcome Trust-Sanger Institute.

Previous analyses of Alu elements in gorillas have been limited to analysis in the context of wider research projects [28,58-61] and have not focused specifically on subfamily analysis. Here we examine the western lowland gorilla genome (build gorGor3.1) [57] to identify gorilla-specific AluY subfamilies and assess the activity levels, copy number, and age of these subfamilies. Our final analysis resulted in the identification of 1,075 *Gorilla* specific Alu element insertions.

**Results and discussion**

**Computational examination of the western lowland gorilla genome**

A total of 1,085,174 Alu elements were identified in the genome of the western lowland gorilla (Additional file 1). Of these, 286,801 were identified as belonging to the ancient AluJ subfamily, and 599,237 were identified as members of the AluS subfamily. A total of 57,427 elements were too degraded or incompletely sequenced to be assigned a subfamily designation by RepeatMasker, and were simply identified as ‘Alu’. We identified 141,709 members of the AluX subfamily. This subfamily is of particular interest due to its relatively young age and known continued mobility in other great ape genomes [1,62]. Approximately one-third (57,458) of these putative AluX elements were >250 bp in length. Gorilla-specific elements were subsequently identified by comparison of orthologous loci in the genomes of human, common chimpanzee, and orangutan [63]. Putative unique, gorilla-specific AluX insertions were estimated at 4,127 copies. This number is similar (96.5%) to the 4,274 gorilla-specific Alu elements identified using other approaches [58]. Individual examination demonstrated that
the majority of our 4,127 loci were in fact shared insertions. These loci were manually examined for gorilla specificity using BLAT [64]. This manual examination excluded 2,858 loci from further analysis due to the presence of shared insertions missed by Lift Over (2,626 insertions) or the lack of orthologous flanking regions in the genomes of other species that preclude PCR verification (232 insertions). This resulted in a total of 1,269 likely gorilla-specific Alu insertion loci for inclusion in subfamily structure analysis.

These 1,269 loci were analyzed for subfamily structure using the COSEG program. COSEG removed 194 probable gorilla-specific Alu insertions from the dataset due to the presence of truncations or deletions in diagnostic regions of the element, leaving 1,075 probable gorilla-specific Alu insertion loci for further analysis Additional file 2. COSEG then divided the loci into 10 subfamilies based on diagnostic mutations in the sequence of the individual Alu elements and provided subfamily consensus sequences (Figure 1) [43]. The consensus sequences were then aligned with known human AluY subfamilies from the RepBase database of repetitive elements [65] (Figure 2). A gorilla-specific nomenclature system was created to designate subfamilies using the suffix 'Gorilla' preceded by the subfamily affiliation based on a comparison to identified human subfamilies (for example, ‘AluYc5a1_Gorilla’). Subfamilies were named in accordance with established practice for Alu subfamily nomenclature [41]. The first identified AluYc5-derived subfamily was, for example, designated AluYc5a3_Gorilla. The ‘a’ denotes the fact that this is the first Yc5-derived subfamily identified. The ‘3’ denotes the number of diagnostic mutations by which this gorilla-specific subfamily differs from the human AluYc5 consensus sequence [41]. Subfamily age estimates were calculated using the BEAST (Bayesian Evolutionary Analysis by Sampling Trees) program [66].

**AluY subfamily activity in the western lowland gorilla genome**

Computational and PCR analysis of the western lowland gorilla genome has identified 1,075 independent, gorilla-specific AluY insertion loci. Computational analysis of this dataset indicates the presence of 10 distinct subfamilies identifiable by the presence of diagnostic mutations specific to each lineage. The 1,075 elements identified in our study almost certainly do not represent the total number of AluY specific to western lowland gorilla genome. Any loci under our arbitrary length of >250 were excluded from our dataset. It is also likely that a number of AluY loci are located in portions of the genome where sequence data is incomplete; within repeat regions, for example. Additionally, some AluY loci were excluded when no orthologous genomic region was present in the species being used for comparison.

The largest newly identified gorilla-specific Alu subfamily was designated as AluY_Gorilla. This designation was established via computational evaluation and manual alignment of the 759 elements assigned to this subfamily. The consensus sequence for these elements was found to be 100% identical to the canonical AluY human consensus sequence (Figure 2). This subset of classic AluY elements continued to propagate in the Gorilla lineage after the divergence from the shared common ancestor with the Homo-Pan lineage. We assayed and verified a total of 135 loci from this subfamily via PCR (18%). The 43 elements belonging to the AluYa1_Gorilla subfamily differ from the AluY consensus sequence by one diagnostic mutation at nucleotide position 133. We assayed and verified via PCR 21 elements in this subfamily (49%). This sequence should not be confused with the Homo-Pan AluYa subfamily.

The AluYa1b4 subfamily is derived from AluYa1_Gorilla and is a small and very likely young subfamily of 13 elements that shared the diagnostic mutation at position 133 of Ya1 but has also accrued four additional diagnostic mutations. We assayed and verified via PCR seven elements in this subfamily (54%). A second identified AluY lineage in gorilla is the AluYc3_Gorilla subfamily. We assayed and verified via PCR 20 of the 69 elements in this subfamily (29%). The consensus sequence for the 69 members identified in this subfamily is a 100% match to the human AluYc3 subfamily consensus sequence (Figure 2).

Two additional gorilla-specific AluYc-derived subfamilies share the characteristic 12 bp deletion at positions 87–98 that is a hallmark of human AluYc5. These two subfamilies possess independent diagnostic mutations that make them distinct from the AluYc5 consensus sequence. These two subfamilies are designated as AluYc5a3_Gorilla (55 elements identified) and AluYc5b2_Gorilla (46 elements identified). AluYc5a3_Gorilla has three additional diagnostic mutations differentiating it from the AluYc5 consensus as a mark of identification. In keeping with Alu subfamily naming convention this subfamily has thus been deemed ‘Yc5a3’, ‘a’ as the first Yc5-like subfamily identified in the gorilla genome and ‘3’ for the three diagnostic mutations differentiating it from the canonical Yc5 consensus. We assayed and verified 27 members of this subfamily via PCR (49%). AluYc5b2 also shares the characteristic 12 bp deletion of the human AluYc5, but has two independent diagnostic mutations (Figure 2). We assayed and verified via PCR 19 members of this subfamily (41%). It is probable that AluYc5a3_Gorilla and AluYc5b2_Gorilla derived from AluYc5 around the time of the Homo_Pan-Gorilla speciation event.

A third lineage nearly identical to human AluYb3a2 was identified as AluYb3a2b2_Gorilla (25 elements identified). This Alu subfamily contains two additional diagnostic mutations. Termed AluYb3a2b2_Gorilla, this lineage is an independent evolution in the Gorilla gorilla gorilla genome and not a derivative of the human-specific AluYb3a2. The
Figure 1 (See legend on next page.)
AluYb lineage is human specific, meaning any identical or apparently derived Alu lineages in other primate genomes must be examples of independent evolution [67]. This is confirmed by the lack of orthologs at the same location in the human genome. We assayed and verified 14 members of this subfamily via PCR (56%). An additional subfamily present at only 17 copies and derived from AluYb3a2b2_Gorilla was identified and termed AluYb3a2b2a2_Gorilla, due to two diagnostic mutations separating these otherwise identical subfamilies. We assayed and verified via PCR nine elements in this subfamily (53%). The low copy number of these subfamilies coupled with their lack of impairing point mutations, even with the caveat that some subfamily members may have been overlooked, leads us to posit that they are among the youngest and potentially still active subfamilies in the western lowland gorilla genome.

Two additional subfamilies were identified that, while clearly AluY derived, do not follow the consensus sequences of established subfamilies available via RepBase. The first, termed AluY16_Gorilla is identified clearly by the presence of an A-rich insert at position 219 followed by a 16 bp deletion, and is present in 30 copies. We assayed and verified via PCR 10 members of this subfamily (33%). The second subfamily, apparently derived from the first and designated AluY16a4_Gorilla, is present in 18 copies and can be distinguished from AluY16_Gorilla by a 20 bp deletion occurring after the A-rich region at position 219. Seventeen elements from this subfamily were assayed via PCR (94%), with 100% of these 17 being verified as gorilla-specific. One locus (gorGor3.1 chrX:74544052–74544324) lacked sufficient orthologous 5’ sequence in non-gorilla outgroups to successfully design a working primer, but was included in the total count based on computational verification. The accumulation of non-diagnostic mutations in these two subfamilies may indicate that they are more ancient.

Approximately 25% of the 1,075 gorilla-specific AluY elements computationally identified in this study were verified by PCR, with the remaining approximately 75% verified by manual examination of computational data. It is important to note that we had no false positives in this study, and 100% of the elements computationally identified as gorilla-specific that were subsequently assayed via PCR were confirmed to be, in fact, gorilla-specific.

One means of identifying potential master elements [7] is to look for subfamily members with mutation-free polyA-tails [68]. A possible source element for the AluY_Gorilla subfamily, for instance, was identified on chrX:5135584–5135921, with a mutation-free 30 bp polyA-tail and intact promoter region. A posited source element for the AluYc5b2 subfamily was identified on chr9:17925753–17926051, also with a mutation-free 30 bp polyA-tail and intact promoter region.

AluY retrotransposition rates appear to be lower in the western lowland gorilla genome than in the human or chimpanzee genomes [69], while higher than that seen in the orangutan genome [48,49]. Factors influencing rates of retrotransposition are myriad [1,46]. Active retrotransposons are frequently polymorphic within a population, and are easily lost during events like speciation or population bottlenecks [70,71]. The number of active elements, and the amplification rate of elements surviving such an event, can vary greatly and impact overall retrotransposition activity in the host genome.

A possible explanation for this lower activity level include inhibition of retrotransposition in the Gorilla lineage by the interaction of host factors such as members of the APOBEC family of proteins with the enzymatic machinery of L1 [1,72]. Interference with L1 and Alu retrotransposition by APOBEC has been documented [72-74]. Analysis of the activity level of Gorilla-specific L1 elements could elucidate this, but has not yet been done. Additionally, environmental stress factors may impact retrotransposition rates [75]. It is possible that one or a combination of these retrotransposition-inhibiting factors could be responsible for the lower level of AluY activity in the western lowland gorilla genome.

A median joining tree of relationships between gorilla-specific AluY subfamilies was generated from a stepwise alignment [76] using the Network program (Figure 1) [42,77]. The tree generated supports the divergence of all gorilla-specific subfamilies from the AluY_Gorilla subfamily, and analysis of subfamily ages using BEAST places the date for this subfamily divergence at the stem of the Gorilla lineage. Initial divergence of gorilla-specific subfamilies occurred shortly after the speciation event separating the Gorilla lineage from the Homo-Pan lineage 6 to 9 million years ago [28,53-55], and master elements have continued to produce copies of each subfamily at varying rates since [7].

Divergence dates of gorilla-specific AluY subfamilies
BEAST analysis of individual subfamily ages suggests no delay or change in transposon activity in western
lowland gorilla following the divergence of the Gorilla and Homo-Pan lineages. The age of the gorilla-specific lineages ranges from 6.5-6.71 million years ago based on a baseline divergence of 7 million years ago for the most recent common ancestor of Gorilla and Homo-Pan. This indicates that all of the identified subfamilies originated around the time of the speciation event that separated these two lineages. This result is consistent with the ongoing propagation of these subfamilies before, during, and after the speciation event at a relatively constant rate. This indicates that the ‘master genes’ [7] from which these subfamilies are derived already existed and were retrotranspositionally active prior to the aforementioned speciation event, and have remained active subsequently. Examination of Alu elements indicates retrotranspositionally active elements are relatively rare, and that most Alu activity is the result of a small number of ‘master’ copies engaging in retrotranspositional activity over time [7]. Our results suggest that the 10 gorilla-specific AluY subfamilies identified in this study diverged and are still diverging from master elements already present in the genome of the common ancestor of the Gorilla and Homo-Pan lineages. A table listing each subfamily, the ‘master gene’ or ancestral Alu subfamily from which it was likely derived, the % divergence from the consensus sequence of the master element, copy number, and suggested age of the most recent common ancestral element are available in the Additional files section of this paper as Additional file 3.

Conclusions
AluY subfamily activity appears to be greatly reduced in the western lowland gorilla genome when compared to the human and chimpanzee genomes. The level of activity seen, while not as low as that observed in the genome of the orangutan, is consistent with a change in host surveillance or intrinsic retrotransposition capacity. Alu subfamily age estimates provide further support for the master gene model of Alu retrotransposition with a relatively low number of ancient lineages responsible for ongoing retrotranspositional activity. The 1,075 lineage specific AluY insertion loci and the 10 subfamilies identified should provide future researchers with a rich source of genetic systems for conservation biology and evolutionary genetics.

Methods
Computational methodology
The genome of the Western lowland gorilla (Gorilla gorilla gorilla) was downloaded and analyzed for the presence of Alu elements using an in-house installation of the RepeatMasker program [62]. The Gorilla gorilla gorilla genome is available for download and analysis via the website of the Wellcome Trust-Sanger Institute [78]. The resulting dataset was parsed into separate files based on the Alu subfamily designations assigned by RepeatMasker. The file containing elements designated as members of the AluY subfamily was then further parsed to remove 84,251 elements under the length of 250 bp using the estimation that shorter elements were likely to be older elements present in multiple species and therefore not useful for our analysis. The ‘Fetch Sequences’ function from the online version of the Galaxy suite of programs [63,79-81] was then used to retrieve the individual DNA sequence present at each of these loci using the gorilla
genome build gor Gor 3.1, and the Lift Over function was used to examine these loci for gorilla lineage specificity by comparison to the closely related genomes of human (Homo sapiens; hg19), chimpanzee (Pan troglodytes; panTro2), and Sumatran orangutan (Pongo pygmaeus abelii; ponAbe2). An additional 200 bp of flanking sequence on each side of the loci assayed was included in this analysis for validation of orthologous loci between the nine primate species considered in this study (Table 1).

Loci selected for verification were examined for further evidence of gorilla-specificity using the BLAST-Like Alignment Tool (BLAT) available at the UCSC Genome Browser website [82]. Putative gorilla-specific loci were compared to the available genomes of three other primate species, human (hg19), chimpanzee (panTro2), and orangutan (ponAbe2) [64,83]. Elements found to be absent in these species and with sufficient orthologous flanking across species were marked for PCR primer design and experimental validation. Loci determined to be shared insertions, as well as those lacking sufficient orthologous flanking for effective primer design, were removed from further consideration [64].

The COSEG program [84], designed to identify repeat subfamilies using significant co-segregating mutations, was then run on the remaining putative gorilla-specific insertions to identify and group specific subfamilies together. COSEG ignores non-diagnostic mutations during analysis, providing an accurate representation of relationships between subfamilies of elements by ignoring potentially misleading mutational events [43]. COSEG uses a minimum subfamily size of 50 elements as the default setting. We arbitrarily defined subfamilies as groups of >10 elements to increase the detail of subfamily structure resolved. A subset of a minimum of 10% of each identified subfamily was then chosen for verification using locus-specific PCR, with a total of 279 loci assayed and verified (Figure 1).

A multi-species alignment comprised of the species listed above was created for each locus using BioEdit [85]. Oligonucleotide primers for the PCR assays were designed in shared regions flanking each putative gorilla-specific element chosen for experimental verification using the Primer3Plus program [86]. These primers were then tested computationally against available primate genomes using the in-silico PCR tool on the UCSC Genome Bioinformatics website [83].

---

**Figure 3** Phylogenetic assay of a western lowland gorilla-specific Alu insertion (Primer Pair Gor112). An agarose gel chromatograph of the gorilla specific Alu insertion Gor112. The filled site is approximately 550 bp (lane 7) and the empty site is 250 bp (lanes 3 to 6 and 8 to 11). Lanes (1) 100 bp DNA ladder; (2) negative control; (3) human; (4) bonobo; (5) common chimpanzee; (6) northern white-cheeked gibbon; (7) western lowland gorilla; (8) Sumatran orangutan; (9) Bornean orangutan; (10) Rhesus macaque; (11) green monkey; (12) empty; (13) 100 bp DNA ladder.

---

**Table 1 DNA sample data of all species examined in this study**

| Taxonomic name               | Common name                  | Origin    | ID number   |
|------------------------------|------------------------------|-----------|-------------|
| Gorilla gorilla gorilla      | Western lowland gorilla      | Coriell   | AG05251     |
| Homo sapiens                 | Human, HeLa                  | ATCC      | HeLa CCL-2  |
| Pan troglodytes              | Common chimpanzee            | IPBIR     | NS06006     |
| Pan paniscus                 | Bonobo                       | IPBIR     | PR00661     |
| Pongo pygmaeus               | Bornean orangutan            | Coriell   | AG05252A    |
| Pongo abelii                 | Sumatran orangutan           | Coriell   | GM0621A     |
| Nomascus leucogenys          | Northern white-cheeked gibbon| Carbone Lab| NLL606      |
| Macaca mulatta               | Rhesus macaque               | Coriell   | NG07098     |
| Chlorocebus aethiops         | African green monkey         | ATCC      | CCL70       |

1Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103, USA.
2From cell lines provided by American Type Culture Collection (ATCC), P.O. Box 1549, Manassas, VA 20108, USA.
3Integrated Primate Biomaterials and Information Resource (IPBIR), http://ccr.coriell.org/Sections/Collections/.
4Laboratory of Dr. Lucia Carbone, Oregon Health & Science University, Beaverton, Oregon, http://carbone-lab.com/.

McLain et al. Mobile DNA 2013, 4:26 Page 7 of 10 http://www.mobilednajournal.com/content/4/1/26
Subfamily age estimates were calculated using the BEAST program [66,87]. BEAST has previously been used to estimate dates of divergence using transposon data [88]. For each subclade, the consensus sequence for each subfamily was determined from the COSEG output [43]. The progenitor element was determined by RepeatMasker analysis of each consensus sequence. Elements were aligned using the SeaView software program and MUSCLE algorithm [76,89]. The progenitor element was then used as an out-group to root the tree of each subclade. BEAST was calibrated with a baseline divergence date of 7 million years ago for the split between the Gorilla and Homo-Pan lineages. A divergence date of 7 million years ago is within the generally accepted 6 to 9 million years ago range for this divergence [28,53-55]. BEAST was run with the following parameters: Site Heterogeneity = ‘gamma’; Clock = ‘strict clock’; Species Tree Prior = ‘birth death process’; Prior for Time of Most Recent Common Ancestor (tmrca) = ‘Normal distribution’ with mean of 7.0 million years and 1.0 standard deviation; ucld.mean = ‘uniform model’ with initial rate set at 0.033; Length of Chain = ‘10,000,000’; all other parameters were left at default settings.

The Network program [90] was run on gorilla-specific AluY subfamily consensus sequences to generate a stepwise tree of relationships between identified subfamilies [42,77]. The consensus sequences for the gorilla-specific AluY subfamilies were aligned using the MUSCLE algorithm [76] and converted to the .rdf file format using the DNAsp program [91]. The .rdf file was then imported to Network, and a median-joining analysis was run. The resulting output file demonstrating evolutionary relationships between subfamilies is presented in Figure 1C.

PCR and DNA sequencing

To verify gorilla-specificity, locus specific PCR was performed with a nine-species primate panel comprised of DNA samples from the following species: Western lowland gorilla (Gorilla gorilla gorilla); Human HeLa (Homo sapiens); Common chimpanzee (Pan troglodytes); Bonobo (Pan paniscus); Bornean orangutan (Pongo pygmaeus); Sumatran orangutan (Pongo abelii); Northern white-cheeked gibbon (Nomascus leucogenys); Rhesus macaque (Macaca mulatta); African green monkey (Chlorocebus aethiops). Information on all DNA samples used in PCR analysis is listed in Table 1. PCR amplification of each locus was performed in 25 μL reactions using 15 ng of template DNA, 200 nM of each primer, 200 μM dNTPs in 50 mM KCl, 1.5 mM MgCl2, 10 mM Tris–HCl (pH 8.4), and 2 units of Taq DNA polymerase. PCR reaction conditions were as follows: an initial denaturation at 95°C for 1 min, followed by 32 cycles of denaturation at 95°C, annealing at the previously determined optimal annealing temperature (60°C with some exceptions), and extension at 72°C for 30 s each, followed by a final extension of 72°C for 1 min. PCR products were analyzed to confirm gorilla specificity of all loci on 2% agarose gels stained with 0.25 ug ethidium bromide and visualized with UV fluorescence (Figure 3). A list of all 279 assayed loci, corresponding primer pairs, and optimal annealing temperatures for each are available as Additional file 4 in the Additional files for this study. Additionally, all PCR tested loci containing unidentified bases in the original sequence data were subjected to chain-termination sequencing to verify bp composition [92]. Sequence data generated from this project for gorilla-specific AluY subfamilies has been deposited in GenBank under the accession numbers (KF668269-KF668278).

Additional files

Additional file 1: Enumeration of Alu elements in ape genomes. The RepeatMasker program was run on the ape genomes currently available for download via Genbank. An in-house Perl script was then used to tally Alu elements by total copy number, and total copy number per each of the three major subfamilies.

Additional file 2: A complete listing of all 1,075 verified gorilla-specific AluY insertions.

Additional file 3: Estimated age of gorilla-specific Alu subfamilies based on BEAST analysis. The BEAST program was run on each gorilla-specific subfamily with a baseline divergence age of 7 million years ago to determine the age of the subfamilies, the most likely progenitor or ancestral element, and the % divergence from the consensus sequence of the ancestral subfamily.

Additional file 4: All primer pairs used in this study listed with chromosomal location of the locus assayed and optimal annealing temperature.

Abbreviations

BAC: Bacterial artificial chromosome; BEAST: Bayesian evolutionary analysis sampling trees; BLAST: Basic local alignment search tool; BLAT: Blast-like alignment tool; bp: Base pair; Chr: Chromosome; Gpb: Gigabase-pair; LINE: Long interspersed element; PCR: Polymerase chain reaction; SINE: Short interspersed element; TPRT: Target-primed reverse transcription; TSD: Target-site duplication; UCSC: University of California Santa Cruz.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

ATM and MAB designed the research and wrote the paper. ATM, GWC, MLF, TOB, and WG performed the experiments. ATM, GWC, MLF, TOB, and WG designed the PCR primers. ATM, CF, and TJM performed the computational analyses. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank G. Cook, J.A. Walker, S. Herke, and M.K. Konkel for all of their helpful advice during the course of this project. Special thanks go to Sydney Szot (szot@tigers.lsu.edu) for the primate illustrations. We thank the American Type Culture Collection, The Coriell Institute for Medical Research, the Integrated Primate Biomaterials and Information Resource, and Dr. Lucia Carbone (http://carbonelab.com) for providing the DNA samples used in this study. This research was supported by National Institutes of Health Grant R01 GM59290 (MAB). ATM was supported in part by a Louisiana Board of Regents Graduate Fellowship and the Louisiana State University Graduate School Dissertation Fellowship. MLF was supported by the Louisiana Biomedical Research Network with funding from the National Center for Research Resources (Grant Number P20GM103424), and by the Louisiana Board of Regents Support Fund.
References

1. Konkel MK, Walker JA, Batzer MA. LINEs and SINES of primate evolution. Evol Anthropol 2010, 19:236–249.
2. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. Nat Rev Genet 2002, 3:370–379.
3. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009, 10:691–703.
4. Ullu E, Tischhui C. Alu sequences are processed 7SL RNA genes. Nature 1984, 312:171–172.
5. Roy-Engel AM, Batzer MA, Deininger PL. Evolution of human retrosequences: Alu, Encyclopedia of Life Sciences. Chichester: John Wiley & Sons, Ltd. 2008.
6. Deininger PL, Batzer MA, Hutchison CA 3rd, Edgell MH. Master genes in mammalian retroviral DNA amplification. Trends Genet 1992, 8:307–311.
7. 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 Res 2005, 15:653–664.
8. Dewarremoux M, Enault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 2003, 35:41–48.
9. Luan DO, Korman MH, Jakubczak JL, Eckbush 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.
10. Luan DO, Eckbush TH. RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element. Mol Cell Biol 1995, 15:3882–3891.
11. Cost GJ, Feng Q, Jacquier A, Boeke JD. Human L1 element target-primed reverse transcription in vitro. EMBO J 2002, 21:5899–5910.
12. Feng Q, Moran IV, Kazazian HH Jr, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 1996, 87:305–916.
13. Murata S, Takasaki N, Satoh M, Okada N. Determination of the phylogenetic relationships among pacific salmonids by using short interspersed elements (SINES) as temporal landmarks of evolution. Proc Natl Acad Sci U S A 1993, 90:6995–6999.
14. Shedlock AM, Okada N. Sine insertions: powerful tools for molecular systematics. Bioessays 2000, 22:148–160.
15. Shedlock AM, Takahashi K, Okada N. Sines of speciation: tracking lineages with retrotransposons. Trends Ecol Evol 2004, 19:545–553.
16. Minghetti PP, Dugaszyck A. The emergence of new DNA repeats and the divergence of primates. Proc Natl Acad Sci U S A 1993, 90:1872–1876.
17. Ray DA, Xing J, Salem AH, Batzer MA. SINES of a nearly perfect character. Syst Biol 2006, 55:928–935.
18. Zietkiewicz E, Richer C, Labuda D. Phylogenetic affinities of tarisier in the context of primate Alu repeats. Mol Phylogenet Evol 1999, 11:73–83.
19. Schmitz J, Ohme M, Zischler H. SINE insertions in cladistic analyses and the phylogenetic affiliations of tarsius bancanus to other primates. J Mol Evol 2005, 61:160–167.
20. Xing J, Wang H, Han K, Ray DA, Huang CH, Cennick LG, Ryder OA, Batzer MA. Alu insertion loci and platyhirine primate phylogeny. Mol Phylogenet Evol 2005, 35:175–186.
21. Xing J, Wang H, Zhang Y, Ray DA, Toai AJ, Distel TR, Batzer MA. A mobile element-based evolutionary history of guenons (tribe cercopithecini). BMC Biol 2007, 5:5.
22. Li J, Han K, Xing J, Kim HS, Rogers J, Ryder OA, Distel TR, Yue B, Batzer MA. Phylogeny of the macaques (cercopithecidae: macaca) based on Alu elements. Gene 2009, 448:242–249.
23. Meyer T, McInley AT, Oldenburg JI, Faulk C, Bourgeois MG, Conlin EM, Mootnick AR, De Jong PJ, Roos C, Carbone L, Batzer MA. An Alu-based phylogeny of gibbons (hylobatidae). Mol Biol Evol 2012, 29:9441–9450.
24. McInley AT, Meyer T, Faulk C, Herke SW, Oldenburg, JM, Bourgeois, MG, Abshire CF, Roos C, Batzer MA: An alu-based phylogeny of lemur (infraorder lemuriformes). PLoS One 2012, 7:e4035.
25. Roos C, Schmitz J, Zischler H. Primate jumping genes elucidate strepsirhine phylogeny. Proc Natl Acad Sci U S A 2004, 101:10650–10654.
26. Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Jorde LD, Garber RK, Walberspoon DJ, Jorde LB, Batzer MA. Alu elements and hominin phylogenetics. Proc Natl Acad Sci U S A 2003, 100:12787–12791.
27. Batzer MA, Stoneking M, Allegria-Hartman M, Bazan H, Kass DH, Shalki TH, Novick GE, Ioannou PA, Scheer WD, Herrera RE. African origin of human-specific polymorphic Alu elements. Proc Natl Acad Sci U S A 1994, 91:12288–12292.
28. Perna NT, Batzer MA, Deininger PL, Stoneking M. Alu insertion polymorphism: a new type of marker for human population studies. Hum Biol 1992, 64:641–648.
29. Deininger PL, Batzer MA. Alu repeats and human disease. Mol Genet Metab 1999, 67:183–193.
30. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev 2012, 22:191–203.
31. Cook GW, Konkel MK, Major JD 3rd, Walker JA, Han K, Batzer MA. Alu pair exclusions in the human genome. Mob DNA 2011, 2:10.
32. Hedges DJ, Deininger PL. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. Mutat Res 2007, 616:466–59.
33. Konkel MK, Batzer MA. A mobile threat to genome stability: the impact of non-LTR retrotransposons upon the human genome. Semin Cancer Biol 2010, 20:211–221.
34. Cook GW, Konkel MK, Walker JA, Bourgeois MG, Fullerton ML, Jessel TT, Herold HD, Batzer MA. A comparison of 100 human genes using an alu element-based instability model. PLoS One 2013, 8:e65188.
35. Jarre J, Smith T. A fundamental division in the Alu family of repeated sequences. Proc Natl Acad Sci U S A 1988, 85:4775–4778.
36. Stangel V, Flemington E, Traina-Dorge Y, Bradshaw H, Deininger P. Clustering and subfamily relationships of the Alu family in the human genome. Mol Biol Evol 1987, 4:19–29.
37. Willard C, Nguyen HT, Schmid CW. Existence of at least three distinct Alu subfamilies. J Mol Biol 1987, 261:80–186.
38. Britten RJ, Baron WP, Stout DB, Davidson EH. Sources and evolution of human Alu repeated sequences. Proc Natl Acad Sci U S A 1998, 95:4770–4774.
39. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jarra J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E. Standardized nomenclature for Alu repeats. J Mol Biol 1996, 263:3–6.
40. Cordaux R, Hedges DJ, Batzer MA. Retrotransposition of Alu elements: how many sources? Trends Genet 2004, 20:464–467.
41. Price AL, Eskin E, Pevzner PA. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res 2004, 14:2245–2252.
42. Liu GE, Alkan C, Jiang L, Zhao S, Eichert EE. Comparative analysis of Alu repeats in primate genomes. Genome Res 2009, 19:876–885.
43. Kapitonov V, Jurka J. The age of Alu subfamilies. J Mol Biol 1996, 42:59–65.
44. Cordaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human Alu elements. Gene 2006, 373:134–137.
45. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB. Mobile elements create structural variation: analysis of a complete human genome. Genome Res 2009, 19:1516–1526.
46. Locke DP, Hillier LW, Warren WC, Worley KC, Nazareth LV, Muzny DM, Yang SP, Wang Z, Chimwana AT, Winz P, Mitveva M, Cook L, Delahaunty KD, Frinick C, Schmidt H, Fulton LA, Fulton RS, Nelson JO, Magrini V, Pohl C, Graves TA, Markovic C, Coo A, Dinh HH, Hume J, Kayor CL, Fowler GR, Lunter G, Meader S, Heger A, et al. Comparative and demographic analysis of orangutan genomes. Nature 2011, 469:529–533.
47. Walker JA, Konkel MK, Ulmner B, Monceaux CP, Ryder OA, Hubley R, Smir AF, Batzer MA. Orangutan Alu quiescence reveals possible source support: evidence for ancient backseat drivers. Mob DNA 2012, 3:0.
50. Strier KB: Primate behavioral ecology. 3rd edition. Boston, MA: Allyn and Bacon, 2007.
51. Fleagle JG: Primate adaptation and evolution. 2nd edition. San Diego, CA: Academic, 1999.
52. Fleagle JG, Janson CH, Reed KE: Primate communities. Cambridge: Cambridge University Press, 1990.
53. Snipen ME, Young NH: Primate molecular divergence dates. Mol Phylogenet Evol 2006, 41:384–394.
54. Giajio GV, Nein M: Estimation of divergence times for major lineages of primate species. Mol Biol Evol 2003, 20:424–434.
55. Chen FC, Li WH: Mammalian non-LTR retrotransposons: major players in intracellular defence. Apobec3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans 2007, 35:637–642.
56. Fleagle JG: Primate molecular divergence dates. Mol Phylogenet Evol 2006, 41:384–394.
57. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, McLain et al. Mobile DNA.
58. Schumann GG: From the margins of the genome: mobile elements shape primate evolution. Bioessays 2005, 27:785–794.
59. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
60. Sen SK, Han K, Wang J, Li WH: Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. PLoS One 2008, 3:e4047.
61. Hormozdiari F, Konkel MK, Prado-Martinez J, Chiatante G, Herraez IH, Walker JA, Bansu S, Vigilant L: Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci U S A 2012, 109:15716–15721.
62. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, Herrero J, Hobolth A, McLain et al. Mobile DNA.
63. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Kent WJ: Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol 2010, Chapter 19:1–21.
64. Hellen EH, Brookfield JF: APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans 2007, 35:637–642.
65. Hulme AE, Bogerd HP, Cullen BR, Moran JV: Selective inhibition of Alu retrotransposition by APOBEC3G. Gene 2007, 390:199–205.
66. Drummond AJ, Rambaut A: BEAST: bayesian evolutionary analysis by sampling trees. BMC Evol Biol 2007, 7:214.
67. 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, 1:167–178.
68. Roy-Engel AM, Salem AH, Dyeniran DO, Deininger L, Hedges DJ, Kilroy GE, Batzer MA, Deininger PL: Active Alu element "A-tails" size does matter. Genome Res 2002, 12:1333–1344.
69. Hedges DJ, Callinan PA, Cordaux R, King J, Barnes E, Batzer MA: Differential alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 2004, 14:1068–1075.
70. Hedges DJ, Batzer MA: From the margins of the genome: mobile elements shape primate evolution. Bioessays 2005, 27:785–794.
71. Hulme AE, Bogerd HP, Cullen BR, Moran JV: Selective inhibition of Alu retrotransposition by APOBEC3G. Gene 2007, 390:199–205.
72. Bogerd HP, Weigand HL, Hulme AE, Garcia-Perez JL, O'Shea KS, Moran JV, Cullen BR: Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proc Natl Acad Sci U S A 2006, 103:8780–8785.
73. Farkash EA, Kao GD, Hornan SR, Prak ET: Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. Nucleic Acids Res 2006, 34:1196–1204.
74. Edgar RC: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004, 32:1792–1797.
75. Li WH: Mammalian non-LTR retrotransposons: major players in intracellular defence. Apobec3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. Biochem Soc Trans 2007, 35:637–642.
76. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
77. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
78. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
79. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
80. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
81. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
82. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
83. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
84. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
85. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
86. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
87. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.
88. Schumann GG: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 2008, 18:343–358.