4-2014

Examining Alu Elements in the Squirrel Monkey Genome

Michelle Limbaugh

Examining *Alu* Elements in the Squirrel Monkey Genome

by

Michelle Limbaugh

Undergraduate honors thesis under the direction of

Dr. Mark A. Batzer

Department of Biological Sciences

Submitted to the LSU Honors College in partial fulfillment of

the Upper Division Honors Program

April 2014

Louisiana State University

& Agricultural and Mechanical College

Baton Rouge, Louisiana

ABSTRACT

The squirrel monkey genome (*Saimiri boliviensis boliviensis*) is the second New World monkey genome assembled and made available online. Recent studies have attempted to resolve the phylogeny between the callitrichines, the cebids (capuchins and squirrel monkeys), and the owl monkeys. Here, we have started a phylogenetic study in New World monkeys using squirrel monkey *Alu*S elements.

Alu elements are ideal markers in phylogenetic studies because their presence or absence is easy to identify, and their insertion is unidirectional. We tested squirrel monkey *Alu*S elements that were absent in common marmoset on a New World monkey panel with three outgroup species. PCR products were visualized using gel electrophoresis to determine the presence or absence of the element at each locus on the panel. Confounding events were further investigated with chain termination DNA sequencing.

The majority of the 140 *Alu* elements tested on the New World monkey panel were found to be squirrel monkey specific. We sequenced 12 confounding events and confirmed a shared insertion in capuchin monkey and 8 near parallel insertions. Further research into the squirrel monkey genome could shed more light onto the branching of the callitrichines, cebids, and owl monkey, as well as contribute to a subfamily analysis of squirrel monkey-specific *Alu* elements.

INTRODUCTION

I. Mobile Elements

The initial discovery of genomic mobile elements is accredited to Barbara McClintock. While studying the property of mosaicism in maize kernels, she determined that heritability of this trait was unstable and did not follow any known patterns of inheritance (McClintock 1950). She asserted that mosaicism was controlled by the interaction of two loci, *Dissociator* (*Ds*) and *Activator* (*Ac*). McClintock hypothesized that these "controlling elements" could transpose in the genome through an unknown mechanism (McClintock 1956). Her contemporaries rejected her postulation of these transposons, but she was later proven correct.

The field of mobile elements is rapidly expanding. Since coding DNA only accounts for 1.5% of the human genome, there is a new focus on the genome content that constitutes the remaining 98.5% (Cordaux and Batzer 2009, Lander *et al*. 2001). A portion of this remainder is composed of a subset of highly repetitive endogenous elements, such as microsatellites. Mobile elements, or transposons, account for between half to two-thirds of the human genome (Lander *et al*. 2001, de Koning *et al*. 2011). These conserved DNA sequences are capable of moving throughout the genome by different mechanisms, and mobile elements are found in all eukaryotic genomes. The remaining non-coding DNA consists of unclassified nucleotides lacking any type of pattern.

Transposable elements are separated into two classes based on their mode of mobilization and insertion: DNA transposons and retrotransposons. We will briefly discuss DNA transposons here and then focus on retrotransposons, the class to which *Alu* elements belong and that constitutes the basis of this research. DNA transposons move by a "cut-

and-paste" mechanism in the genome: endonuclease proteins excise the element, and transcriptase proteins reinsert it into a new locus in the genome (Craig *et al.* 2002). This mode of mobilization does not generate new elements; rather, the mechanism moves only existing DNA transposon copies, thus accounting for their small contribution to the percentage of all mobile elements (Cordaux and Batzer 2009). DNA transposons are not essentially active in primate genomes today, but they mobilized throughout early primate evolution until about 37 million years ago (Smit 1999, Pace and Feschotte 2007).

II. Retrotransposons

Retrotransposons are the second class of genomic mobile elements. DNA is transcribed into an RNA intermediate that is then reverse transcribed as DNA in another part of the genome (Konkel *et al.* 2010, Smit *et al*. 1995). This mechanism is referred to as a "copy-and-paste" mechanism where each event creates a duplicate copy of the retrotransposon in the genome. This type of mobilization increases retrotransposon copy numbers in their host genomes. Thus, retrotransposons make up at least 45% of the human genome (Lander *et al*. 2001). Retrotransposons can be further divided into two groups based on the presence or absence of long terminal repeats (LTRs): LTR retrotransposons and non-LTR retrotransposons. LTR retrotransposons are longer than non-LTRs, they closely resemble a simple retrovirus, and they are less numerous in the genome than non-LTR retrotransposons (Deininger *et al.* 2003). However, we will focus on non-LTR retrotransposons since they include *Alu* elements.

Non-LTR retrotransposons are currently active in the human genome (Cordaux and Batzer 2009). They include LINEs (long-interspersed elements) and SINEs (short-interspersed elements). Non-LTR retrotransposons can also be subdivided into

autonomous and nonautonomous elements. Autonomous elements contain regions that code for their own enzymatic machinery that allow them to independently retrotranspose throughout the genome. Nonautonomous retrotransposons lack this machinery and must instead hijack that of the autonomous elements to move throughout the genome (Dewannieux *et al.* 2003, Mathias *et al*. 1991).

III. Alu *Elements*

Alu elements are a type of retrotransposon that arose in primates around 65 million years ago (Deininger and Daniels 1986, Ulla and Tschudi 1984). Due to their continued mobilization, there are over 1 million *Alu* elements in the human genome; they are the most successful mobile element by copy number (Lander *et al.* 2001). Full-length *Alu* elements are typically ~300 base pairs long, but there is a modest range of length due to the variable-length adenosine-rich tail, or oligo(dA)-rich tail, found at the end of an *Alu* (Kazazian 2004, Kriegs *et al*. 2007). *Alu* elements are dimeric, consisting of two fused monomers. The left monomer contains the A and B boxes which have a promoter for RNA polymerase III, and an A-rich region of varying length separates the left monomer from the right monomer (Konkel *et al.* 2010). The right monomer ends with the variable-length oligo(dA)-rich tail at the 3' end of the element, a sequence of repeating adenine nucleotides occasionally interspersed with other bases (Batzer and Deininger 2002) (Figure 1).

Figure 1. Structure of an *Alu* Element. The element begins with the GGCCGGG sequence at the 5' end and ends with an adenosine-rich tail of variable length at the 3' end. *Alu* elements are dimeric, with their left and right monomers connected by the A-rich region shown. The A and B boxes in the left monomer contain the promoters recognized by RNA polymerase III that allow initiation of transcription of the *Alu*. The *Alu* then moves through target-primed reverse transcription, resulting in the target site duplications (TSDs) that flank the element. A full length *Alu* is approximately 300 base pairs long.

Alu elements are a type of SINE (Singer 1982). They are nonautonomous elements and must use LINE-1 (or L1) endonuclease and reverse transcriptase machinery to move throughout the genome (Dewannieux *et al*. 2003, Mathias *et al*. 1991). *Alu* elements move via target-primed reverse transcription (TPRT) using an RNA intermediate (Esnault *et al*. 2000, Luan *et al*. 1993).

Prior to TPRT, the RNA polymerase III promoter in the A and B boxes of the *Alu* element's left monomer attracts this polymerase to initiate transcription of the RNA intermediate (Batzer and Deininger 2002). This RNA intermediate then hijacks the endonuclease and reverse transcriptase enzymes produced by L1. In TPRT, the L1 endonuclease nicks the 5' TTTTAA 3' sequence at the target integration site (Luan *et al*. 1993, Feng *et al*. 1996, Jurka 1997). The oligo(dA)-rich tail of the RNA intermediate anneals to the TTTT sequence, and the L1 reverse transcriptase initiates the reverse transcription of the RNA intermediate into DNA (Cost *et al*. 2002, Konkel *et al*. 2010). The complementary strand of target DNA is then nicked, and the DNA sequence of the *Alu* element is incorporated into this site (Esnault *et al*. 2000). Target site duplications (TSDs) are evidence of TPRT as the mechanism of *Alu* element mobilization. TSDs result from the staggered cut with "sticky ends" made in the target DNA prior to insertion; the polymerase fills the gaps in sequence caused by these sticky ends, and two short, identical sequences (6 to about 20 base pairs long) are thus generated at each end of the element (Cost *et al.* 2002) (Figure 2).

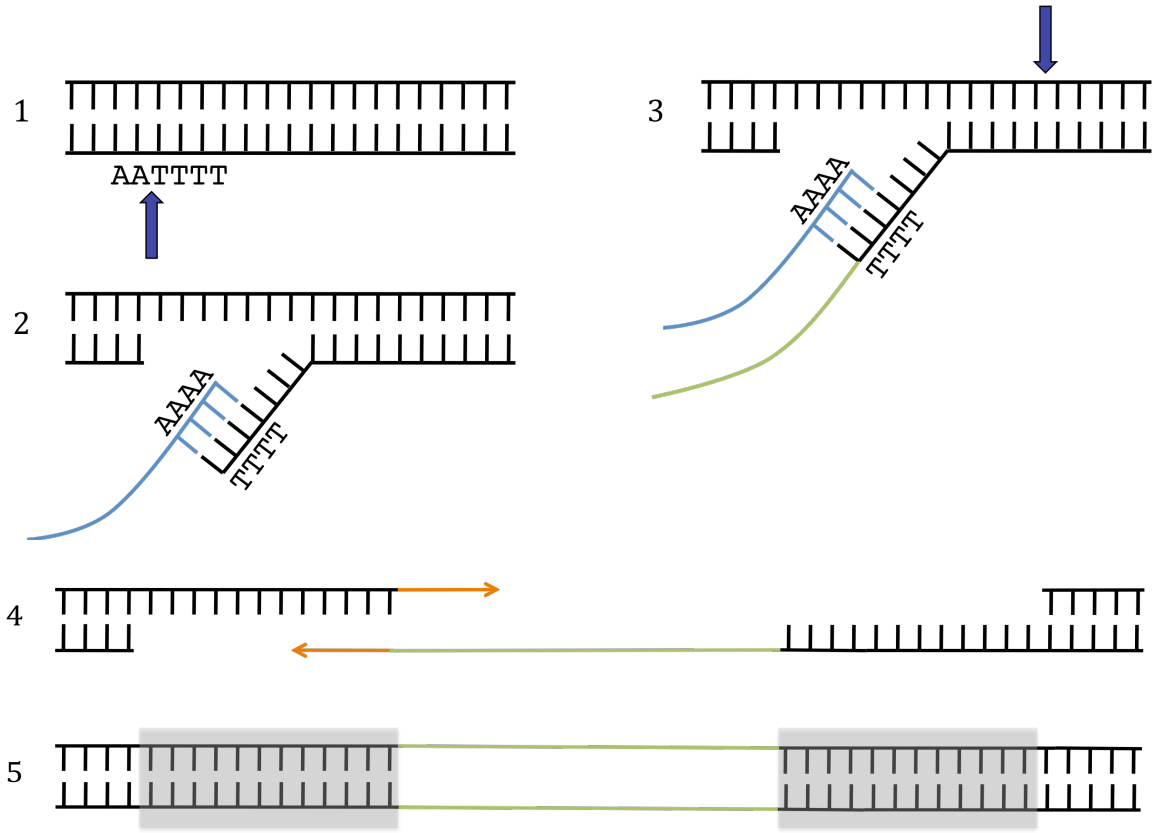

Figure 2. Target-Primed Reverse Transcription. (1) The target DNA is nicked by L1 endonuclease at the target integration site. (2) The oligo(dA)-rich tail of the RNA intermediate (shown in blue) complementarily base pairs with the TTTT sequence on the target DNA. (3) The L1 reverse transcriptase begins transcription of the RNA intermediate into DNA (shown in green). Then a second nick is made on the complementary strand of DNA. (4) The DNA sequence of the *Alu* is incorporated into the target site, and its complementary base pair sequence is transcribed via an unknown mechanism. The orange arrows indicate extension of the sequence on both strands with DNA polymerase. (5) The gaps caused by the staggered cuts are filled in by polymerase, creating the target site duplications (gray boxes) surrounding the *Alu* element (image based off of figures from Konkel *et al.* 2010, Cordaux and Batzer 2009).

IV. Alu *Elements as Phylogenetic Markers*

Alu elements are often used as genetic markers to elucidate phylogenetic relationships among primate species. Since the ancestral state of any given locus is the absence of an insertion, the interrelatedness of different species can be examined by determining the insertion patterns at a particular locus in their respective genomes (Xing *et al.* 2005, Li *et al*. 2009). In other words, if three different species are examined at a specific locus, and if species A and species B share an insertion that is absent in species C,

then this provides evidence for the hypothesis that A and B are more closely related to each other than either is to C.

Furthermore, *Alu* elements have little homoplasy, meaning that a shared insertion between species implies a common ancestor. There are four confounding events associated with *Alu* element-based phylogeny: parallel insertions, precise deletions, incomplete lineage sorting, and paralogous insertions (Ray *et al.* 2006). Both precise deletions and parallel independent insertions of *Alu* elements are rare. Incomplete lineage sorting occurs when a polymorphic insertion in a common ancestor is variably and randomly fixed or absent among the daughter species (Nei 1987, Ray *et al*. 2006). A paralogous insertion may occur when a region of the genome is duplicated but an insertion occurred at only one of the duplicates (Ray *et al.* 2006). However, paralogous insertions are usually resolvable through sequence analysis.

V. New World Monkeys

Platyrrhini is a parvorder of neotropical, anthropoid primates also known as the New World monkeys. They represent a conclusive monophyletic clade, as evidenced by three independent transpositional events unique to New World monkeys (Singer *et al.* 2003). Researchers agree on the three families Cebidae (small, clawed monkeys), Pitheciidae (seed predator monkeys), and Atelidae (large monkeys with prehensile tails) (Schneider *et al.* 1996, Ray *et al.* 2005). The Cebidae and Atelidae families are grouped to the exclusion of the Pitheciidae family (Ray *et al*. 2005). These taxonomic families also represent the three monophyletic clades of platyrrhines (Figure 3). The New World monkeys are found in Central and South America as well as parts of southern Mexico (Steiper and Young 2006). It is believed that a rafting event occurred 26 to 40 million years

ago involving the Simiiformes infraorder, causing the split between platyrrhines and catarrhines (catarrhines include apes and Old World monkeys) (Perez *et al.* 2012). This split and the subsequent isolated evolution of New World monkeys are easily visualized by morphological differences between platyrrhines and catarrhines: platyrrhines have broad, flat noses (a characteristic so prominent that it is used to name them), four extra premolars, no opposable thumbs, and all but howler monkeys (genus *Alouatta*) lack trichromatic vision (Groves 2005).

Following their migration event, the platyrrhines underwent extensive and rapid radiation into the fifteen genera that comprise the clade, resulting in immense phenotypic diversity among the New World monkeys (Aristide *et al*. 2013). Today, for example, New World monkeys range in height from the pygmy marmoset (15 cm) to the southern muriqui (55-70cm) (Smith and Jungers 1997). All New World monkeys are diurnally active except for owl monkey (genus *Aotus*); interestingly, owl monkeys lack a tapetum lucidum (the layer of the eye seen in most nocturnal mammals), suggesting that this nocturnal activity represents an independent evolutionary event (Wildman *et al.* 2009). Capuchin monkeys (genus *Cebus*) have the highest primate encephalization quotient behind humans (Harvey *et al*. 1987) and have very long life spans (up to 50 years) (Judge and Carey 2000). These traits may be useful in studying evolution of higher order thinking in humans (Wildman *et al*. 2009). Thus, the diversity seen in the New World monkeys can lead to many new areas of research, placing high importance on correctly interpreting their phylogenetic tree.

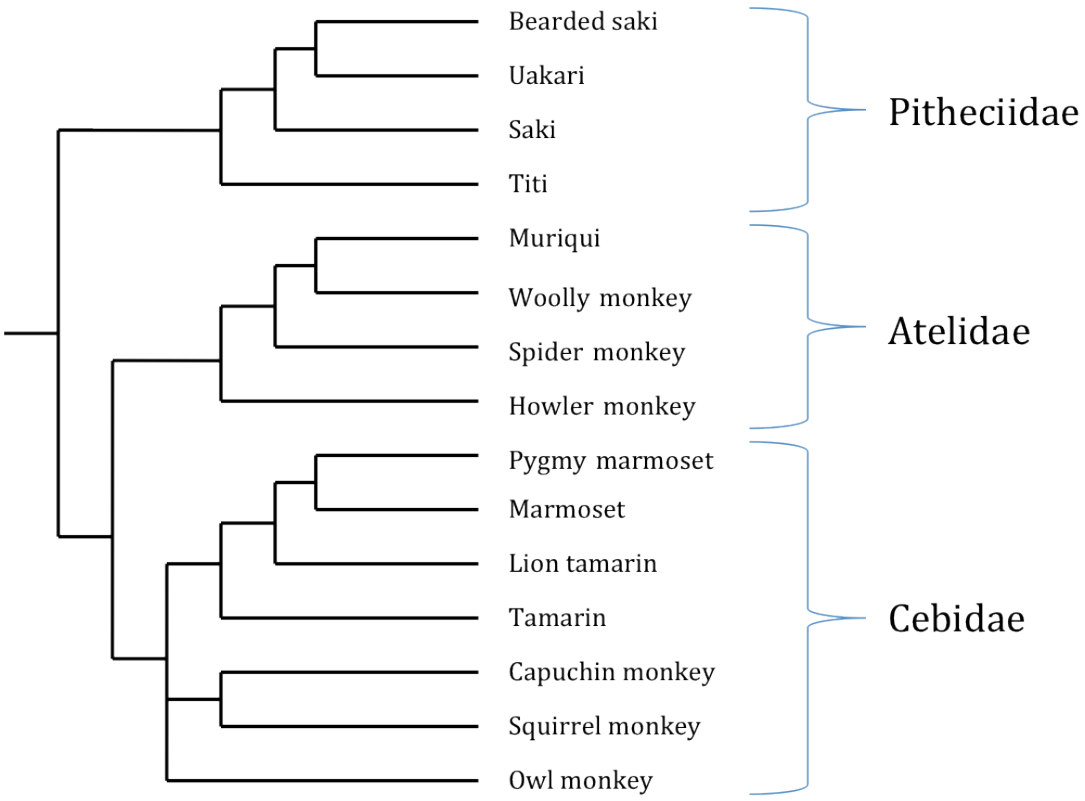

Figure 3: The New World Monkey Phylogenetic Tree. The New World monkeys are a monophyletic clade because they arose from the same common ancestor. Each node on the tree designates a common ancestor shared by all of the taxa that branch off of the node. Taxa that share a node are called sister taxa. For example, bearded saki, uakari, and saki are sister taxa; however, bearded saki and uakari are also considered sister taxa to the exclusion of saki because they share a more recent common ancestor which each other than either of them share with saki. There are 15 genera in the New World monkey clade. They are separated into three families: Pitheciidae, Atelidae, and Cebidae (tree based off of Wildman *et al*. 2009 comprehensive data set).

VI. Phylogenetic Study of the New World Monkeys

The rapid radiation within the New World monkey families has made it difficult to elucidate platyrrhine evolutionary history and phylogeny. Within this monophyletic clade, the high rate of speciation events confounds efforts to organize its members. Furthermore, recent molecular clock data show that early divergence of the three families Pitheciidae, Atelidae, and Cebidae occurred over just a few million years (Opazo *et al.* 2006). Rapid radiation can cause homologous phenotypes, genotypes, or DNA sequences to be shared between non-sister taxa, (i.e. paraphyletically as opposed to monophyletically) (Barbulescu

et al. 2001). Hybridization between diverging populations in relatively close proximity muddles results of phenotypic and genotypic comparisons (Osterholz *et al.* 2008). Additionally, incomplete lineage sorting has most likely occurred during the speciation events of the New World monkeys, possibly distorting the calculated coalescence times (Hobolth *et al*. 2011). However, homoplasious events only occur at a rate of ~1.4% in New World monkeys (Xing *et al*. 2005, Ray *et al*. 2005, Salem *et al*. 2003a, Salem *et al*. 2003b). Thus, the largely infrequent nature of incomplete lineage sorting in relation to non-homoplasious events should not prevent the eventual resolution of platyrrhine phylogeny. To this day, studies based on morphology, behavioral patterns, evolutionary past, and mitochondrial DNA have fallen short of fully resolving the branching of the platyrrhines on a genus and species level.

In general, these studies have encountered the following question: what are the phylogenetic relationships between the callitrichines (common marmoset and pygmy marmoset), the cebids (capuchin monkey and squirrel monkey), and owl monkey? (Singer *et al.* 2003). The two most highly supported hypotheses for this branching are represented in Figure 4. Two notable studies highlighting the difficulty of answering this question follow.

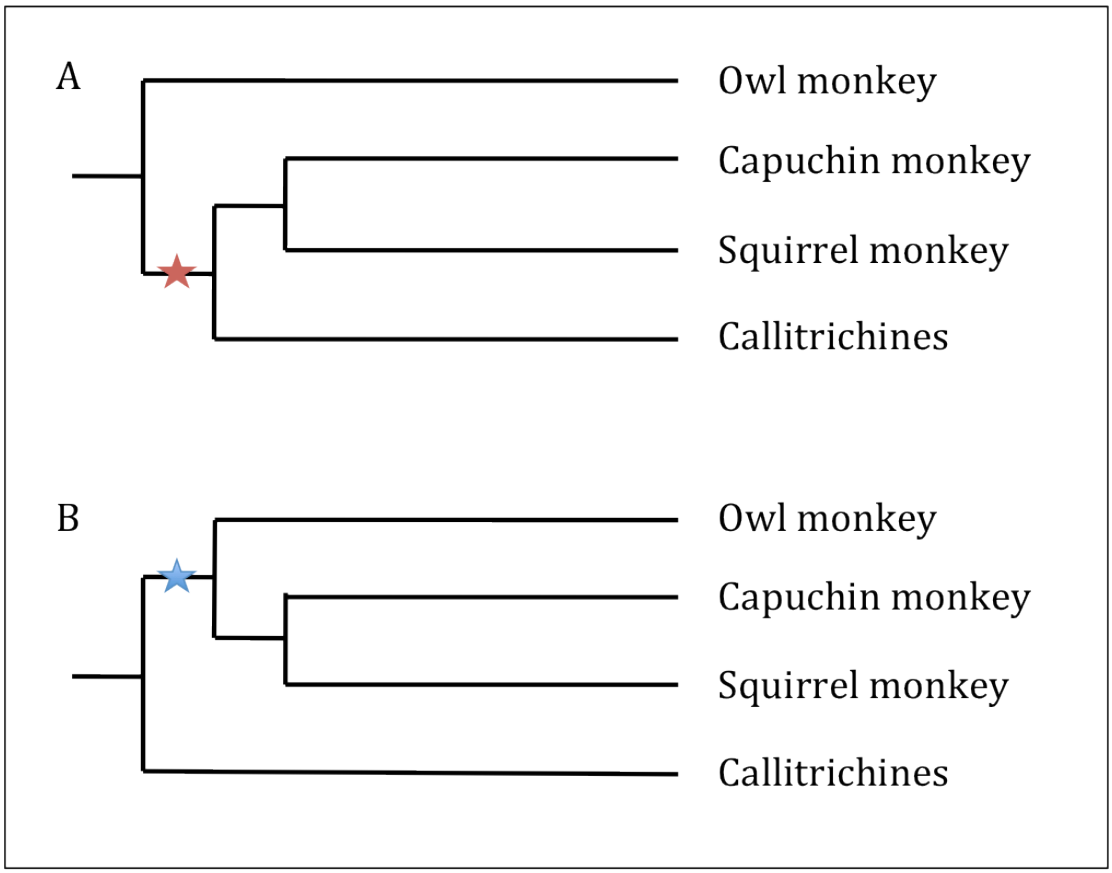

Figure 4: New World Monkey Phylogeny Hypotheses. These two hypotheses show the unresolved branching order. The hypothetical tree A could begin to be supported by finding a shared *Alu* element between capuchin monkey, squirrel monkey, and the callitrichines that was also absent in owl monkey. Such an element would have arisen in a hypothetical most recent common ancestor population at the hypothetical node between the callitrichines and the cebids (red star). By contrast, the hypothetical tree B could begin to be supported by a shared *Alu* element between capuchin monkey, squirrel monkey, and owl monkey that is absent in the callitrichines. This element would have arisen in the hypothetical most recent common ancestor of both owl monkey and the cebids (blue star). Since we selected for squirrel monkey *Alu* elements not present in common marmoset, the direct support of the first hypothetical tree is impossible.

Singer *et al.* (2003) addressed the question of branching order between the callitrichines, the cebids, and owl monkey using *Alu* elements. A significant result was the identification of an *Alu* element that arose in the most recent common ancestor of the cebids, callitrichines, and owl monkey, supporting this clade as monophyletic. However, the study was not able to fully resolve the polytomy between the cebids, owl monkey, and the callitrichines. Wildman *et al*. (2009) addressed this question using novel markers (non-

genic, non-coding, non-repetitive) and traditional markers (*Alu* elements) in New World monkeys. Wildman *et al*. constructed a consensus tree grouping owl monkey with the cebids to the exclusion of the callitrichines. However, bootstrap support for the node between owl monkey and the cebids was weak (55% for maximum parsimony and 53% for maximum likelihood); the other possibility (that is, the callitrichines grouped with the cebids to the exclusion of owl monkey) is almost equally likely.

VII. An Analysis of Squirrel Monkey-Specific Alu *Elements*

The squirrel monkey (genus *Saimiri*) inhabits a range spreading over the tropical lowland forest of Central and South America (Steiper and Young 2006). They often associate peaceably with capuchin monkeys in mixed-species groups where they benefit from the detailed alarm calling system of the capuchins (Sussmann 2000). The Bolivian squirrel monkey (*Saimiri boliviensis*) and the common marmoset (*Callithrix jacchus*) genomes are the only New World monkey genome draft assemblies currently made available. In this study, research was based off of one of two possible solutions for the aforementioned polytomy. Here, we identified *Alu* elements derived from *Alu*S elements present in the squirrel monkey genome and absent in common marmoset. We sought to support a putative phylogenetic tree placing owl monkey as a sister taxon to the cebids to the exclusion of the callitrichines (see Figure 4).

MATERIALS & METHODS

I. Collection of Alu *Elements and BLAT Querying*

To locate potential squirrel monkey-specific loci, we used the draft genome assembly of the Bolivian squirrel monkey (*Saimiri boliviensis boliviensis*—saiBol1) made

available by the Broad Institute (Cambridge, MA) on the UCSC Genome Browser. The genome was assembled from contigs of DNA from a female Bolivian squirrel monkey (Broad sample ID #3227) and is not yet assembled on a chromosomal level. Using the UCSC Genome Browser's Table Browser function, we selected *Alu*S elements from the squirrel monkey genome along with 500 base pairs of flanking on either side of each element. Using the online program BLAT (UCSC Genome Bioinformatics, Kent 2002), we compared squirrel monkey *Alu*S elements against the common marmoset genome (calJac3). Squirrel monkey elements that were absent in common marmoset were then queried against three outgroup genomes: human (hg19), chimpanzee (panTro2), and rhesus macaque (rheMac2). In order for a locus to be used for further analysis, it had to meet the following conditions:

- The *Alu* element present at that locus must not be present in the orthologous locus of any of the outgroup genomes (common marmoset, human, common chimpanzee, and rhesus macaque).
- The upstream and downstream flanking sequence of the element must have matching orthologous sequences in at least common marmoset and one of the Great Apes (human or chimpanzee).
- Upon alignment with the outgroup genomes at that locus, there must be enough uninterrupted, non-repetitive upstream and downstream flanking sequence to design appropriate forward and reverse primers around the *Alu* element.
- Due to time constraints and desire to preserve the quality of our study, any elements that did not fit the above criteria were not used for further analysis.

Using these methods, we computationally analyzed 1, 548 squirrel monkey *Alu*S elements. Of these, 290 were computationally determined to be squirrel monkey-specific elements (i.e. they were initially determined to be absent in the common marmoset genome, then determined to be absent in the human, common chimpanzee, and rhesus macaque genomes), and 140 *Alu* elements fit the above criteria and were thus eligible for primer design.

II. Primer Design

We aligned the flanking of the squirrel monkey-specific *Alu* elements with the four outgroup sequences using the ClustalW feature on the BioEdit Sequence Alignment Editor (Hall 1999). We used the most conserved regions of the flanking sequence—at least 50 base pairs away from either side of the element—to design forward and reverse oligonucleotide primers for the locus. We used RepeatMasker on the squirrel monkey element and flanking to avoid designing primer pairs within repetitive elements in the flanking (Smit *et al*. 1996).

We designed orthologous primers using Primer3 (Rozen and Skaletsky 2000). Suitable primers were between 20-27 base pairs in length, had an annealing temperature between 57^0C and 61^0C, and contained four or fewer consecutive identical nucleotides (max poly-X of 4). Using the *in silico* PCR function in BLAT, we obtained and recorded the expected sizes of the PCR products in all five species, and we confirmed that the primer pair yielded only one product in each species. We also used BLAT to confirm the uniqueness of the primer pair within the squirrel monkey, common marmoset, human, and common chimpanzee genomes to avoid using primers that could anneal with a similar sequence elsewhere in the genome (primer promiscuity). PCR primers were obtained from

Sigma Aldrich (Woodlands, TX, USA). The stock primer was re-suspended and diluted using TLE buffer (10 mM Tris-HCl/0.1 mM EDTA), and primers were then tested using HeLa DNA (from human cell line HeLa CCL-2 from ATCC) in a PCR reaction.

III. Polymerase Chain Reaction

Once tested for amplification using HeLa DNA in a PCR reaction, the squirrel monkey *Alu* elements were then tested against a New World monkey panel containing 14 species and a negative control of TLE buffer (Table 1). The New World monkey panel consisted of three Old World outgroup species and 11 New World monkey species. PCR amplification was performed under the following conditions: 25 µl total volume using 15 ng of template DNA, 200 nM of each primer, 200 µM dNTPs in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.4), and 2 units of *Taq* DNA polymerase. PCR reaction conditions were as follows: initial denaturation at 94°C for 1 minute, followed by 32 cycles of denaturation at 94°C for 30 seconds, annealing at 57°C* for 30 seconds, extension at 72°C for 30 seconds, and a final extension of 72°C for 2 minutes. (*Note: certain primer pairs for particular loci required PCR reactions with an annealing temperature anywhere in the range of 50°C to 60°C; the ideal annealing temperature was determined based on a temperature gradient PCR reaction.) Products were run out at 200 mV on a 2% agarose gel stained with 0.1 µg/mL ethidium bromide and loaded next to a 100 base pair ladder. The gels were run for a length of time dependent on the expected size of the products (anywhere from 50 to 70 minutes in this study). The PCR products were then visualized with UV fluorescence.

Species	Common name	Origin	ID	Primate Lineage
Homo sapiens	Human	ATCC	HeLa CCL-2	Old World outgroup species
Pan troglodytes	Common Chimpanzee	IPBIR	NS06006	
Chlorocebus aethiops	African Green monkey	ATCC	CCL70	
Lagothrix lagotricha	Woolly monkey	Coriell	NG05356	Atelidae
Ateles belzebuth	White bellied spider monkey	SDFZ	KB6701	
Alouatta sara	Bolivian red howler monkey	SDFZ	OR749	
Callithrix jacchus	Common marmoset	NERPRC	cj393-99, A02-738	Cebidae
Callithrix pygmea	Pygmy marmoset	SDFZ	OR690	
Saguinus fuscicollis nigrifrons	Geoffroy's saddle-back tamarin	SDFZ	OR621	
Cebus	Capuchin monkey	KP	CA003	
Saimiri s. sciureus	Common squirrel monkey	SDFZ	KB4544	
Aotus trivirgatus	Owl monkey	ATCC	CRL1556	
Pithecia p. pithecia	Northern white-faced saki	SDFZ	OR842	Pitheciidae
Callicebus d. donacophilus	Bolivian gray titi	SDFZ	OR1522	

Table 1: New World Monkey Panel. The above species were used in our panel. The origins of the DNA samples follow. ATCC: American Type Culture Collection cell lines; IPBIR: Integrated Primate Biomaterial and Information Resource; Coriell: Coriell Institute for Medical Research, 403 Hadden Avenue, Camden, NJ; SDFZ: San Diego Frozen Zoo, Conservation and Research for Endangered Species; NERPRC: New England Regional Primate Research Center; KP: Kimberly Phillips, Trinity University.

IV. DNA Sequencing

We recorded presence/absence data of each locus for all species used in the New World monkey panel. This is a relatively simple task since loci on the gel assay that appear to be roughly 300 base pairs longer than orthologous loci in other species are likely to contain an *Alu* insertion, which can be confirmed through DNA sequencing analysis. In this study, the vast majority of loci tested contained a squirrel monkey-specific *Alu* insertion that was absent in all of the orthologous species tested. Whenever PCR products varied from this common result, such as an insertion of similar size in another species, sequencing was performed on both the squirrel monkey-specific amplicon and on the similarly sized amplicon in the other species. We used a standard sequencing reaction with the Applied Biosystems BigDye v3.1 sequencing reagents and Sanger chain termination sequencing

(Sanger *et al.* 1977). Sequencing reactions were carried out using capillary electrophoresis in a 3130X Genetic Analyzer. Sequencing results of both the locus in squirrel monkey and that in the other species were aligned in BioEdit to the original candidate locus. We were then able to determine whether the event resulted from a shared insertion, a near parallel insertion, or a non-repeat mediated insertion.

RESULTS & DISCUSSION

I. General Results

The majority of the squirrel monkey *Alu* elements computationally analyzed in this study were shared with the common marmoset. Since the squirrel monkey and the common marmoset are closely related, having diverged about 17.1 million years ago, this is not an unexpected result (Perez *et al.* 2012). Out of the 1,548 squirrel monkey *Alu* elements computationally analyzed, 1,258 of the insertions were also present in the common marmoset genome. Of the remaining 290 computationally determined to be present in squirrel monkey and absent in common marmoset, 140 loci were found to be suitable for locus-specific PCR analysis. The remaining 150 loci were deemed unsuitable because they did not meet the specified criteria for further analysis (see "Collection of *Alu* Elements and BLAT Querying" in the Materials & Methods section).

We designed working primers for the 140 suitable *Alu* elements and tested them against the New World monkey panel using PCR. The overwhelming majority of these *Alu* elements were found to be squirrel monkey specific, meaning they were absent in all other species tested on the New World monkey panel (Figure 5A). However, 12 of the *Alu* elements analyzed through PCR resulted in the amplification of the *Alu* element in squirrel

monkey as well as a similar sized insertion in another New World monkey (Figure 5B-D). These confounding events were further examined using Sanger chain-termination sequencing (Figure 6, Figure 7).

Of the 12 sequenced loci, three did not yield successful results. Among the 9 successfully sequenced loci, we found one shared *Alu* insertion between capuchin monkey and squirrel monkey to the exclusion of the other New World monkeys on the panel (Figure 5B). This further supports the cebid node, placing squirrel monkey and capuchin monkey as sister taxa to each other. To confirm this shared insertion, we identified shared flanking on both sides of both elements and the presence of the same TSDs on either side of both elements (Figure 6). We found four near parallel insertions of full-length *Alu* elements; these insertions were found in owl monkey (locus 751) (Figure 5C, Figure 7), capuchin monkey (locus 1059), howler monkey (locus 1063), and tamarin (locus 1064) (Figure 5D). We found a near parallel insertion of a truncated *Alu* element in African Green monkey (locus 352). We also identified 3 near parallel insertions of truncated L1 elements in titi (locus 98), woolly monkey (locus 922a), and capuchin monkey (locus 985). To confirm these insertions as near parallel, we looked for shared flanking on both sides of both elements (that is, the squirrel monkey *Alu* element and the *Alu* element in the other species) and the presence of unique TSDs for each element.

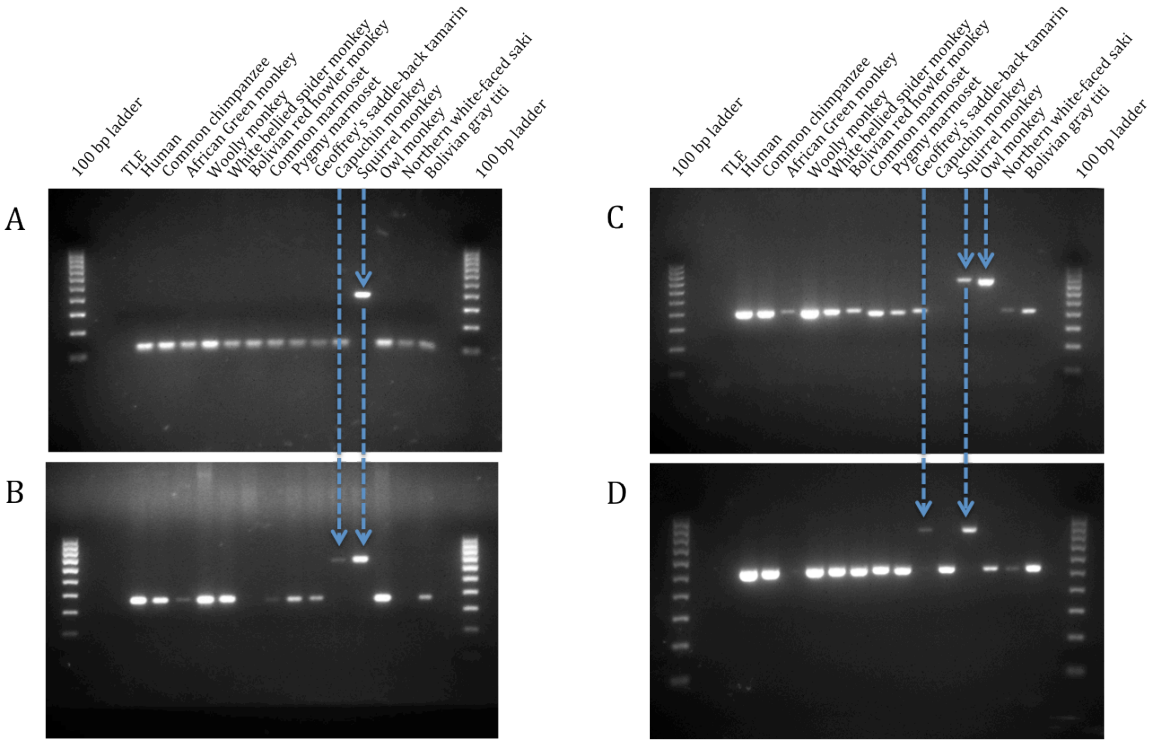

Figure 5: Gel Chromatographs of PCR Results. The gel chromatographs of the panel are flanked with 100 base pair ladders, and human DNA is preceded by a TLE negative control. (A) Locus 460 shows the presence of an insertion (upper band) in squirrel monkey and the absence of an insertion (lower band) in all other species. This was the result seen in the majority of the loci studied. The following three gel pictures show confounding events that were sequenced. (B) Locus 713 is a shared insertion between squirrel monkey and capuchin monkey, confirmed through sequencing. Note: we used redesigned primers to amplify this locus in capuchin monkey and thus confirmed an empty site at this locus in capuchin. (C) Locus 751 shows a squirrel monkey-specific *Alu* element and a near parallel independent insertion of an *Alu* element in owl monkey that was confirmed through sequencing. (D) Locus 1064 shows a squirrel monkey-specific *Alu* element and a near parallel independent insertion in Geoffrey's saddle-back tamarin that was confirmed through sequencing.

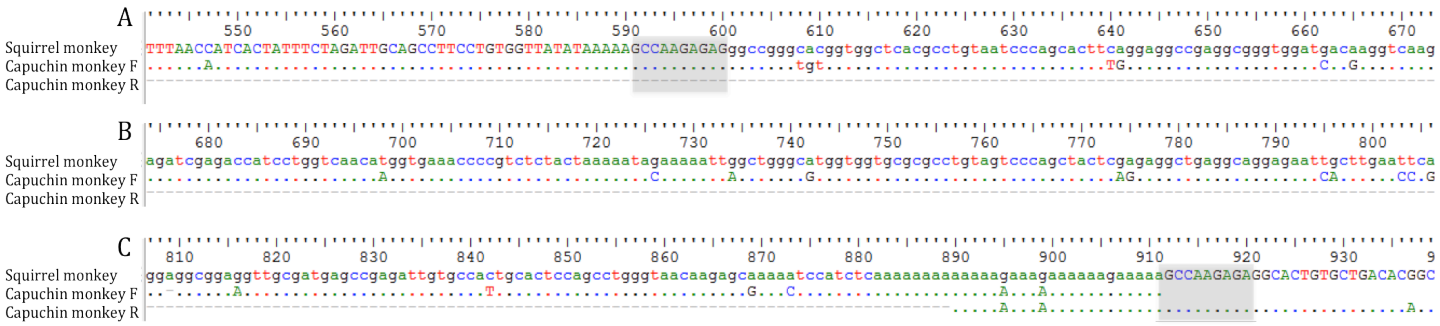

Figure 6: Shared, Cebid-Specific *Alu* Insertion. For locus 713, "Capuchin monkey F" and "Capuchin monkey R" refer to the forward and reverse sequences, respectively. The dots represent shared base pairs with the squirrel monkey sequence. Note the shared flanking sequence just upstream of the 5' end of the element. The TSDs of the *Alu* element are shaded in blue and are present in both species. The reverse sequence shows the polyA-tail followed by the shared flanking sequence downstream of the element. This *Alu* element is specific to the cebids.

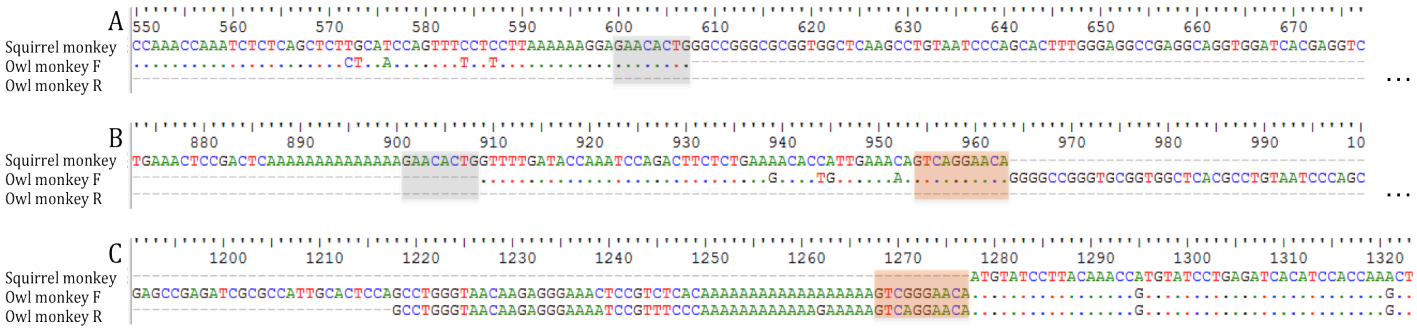

Figure 7: Near Parallel Independent *Alu* Element Insertion. For locus 751, "Owl monkey F" is the forward sequence, and "Owl monkey R" is the reverse sequence. The dots represent shared base pairs, while the dashes represent an absence of base pairs. The ellipses designate the continuation of sequence not shown here. Note the presence of shared flanking sequence in squirrel monkey and owl monkey approaching the beginning of the squirrel monkey-specific *Alu* element in segment A. The TSDs for this element are shaded in grey. After the second TSD in segment B, there is another stretch of shared flanking sequence between squirrel monkey and owl monkey. The owl monkey-specific *Alu* element starts downstream of the squirrel monkey-specific insertion. The TSDs for this element are shaded in orange. The second TSD in segment C is followed by shared flanking sequence between squirrel monkey and owl monkey.

II. Phylogenetic Implications

The specific branching between the callitrichines, cebids, and owl monkey remains unresolved. In this study, we hypothesized that the owl monkey is sister to the cebid node to the exclusion of the callitrichines (see Figure 4). Based on the setup of this experiment, direct support would not be possible for the other scenario. During computational analysis,

we only examined squirrel monkey *Alu*S elements that were not present in common marmoset. Thus, testing these *Alu* elements with PCR on the New World monkey panel could never result in a shared element between squirrel monkey and common marmoset. Finding such an element would begin to suggest the placement of the callitrichines as sister taxon to the cebids, to the exclusion of owl monkey. Although it would be impossible to find an element of that nature in this study, we also did not find an *Alu* element shared between the cebids and owl monkey (i.e. an element to support the hypothesis placing owl monkey and the cebids as sister taxa). However, we did find a shared *Alu* element insertion in squirrel monkey and capuchin monkey, further supporting the cebid node.

FUTURE WORK

The New World monkeys represent a unique group phenotypically, behaviorally, and evolutionarily. From frequent twins and triplets in marmoset litters to a prehensile (grasping) tail in spider monkeys, the platyrrhines are very diverse from both other primates and themselves. The branching order in New World monkeys has received much attention because it can enlighten researchers in reconstructing the evolutionary patterns that led to these unique characteristics.

Once the branching between the genera of the callitrichines, cebids, and owl monkey is fully resolved, the phylogeny of New World monkey species can be further resolved. However, this goal is likely to be difficult to achieve due to the rapid radiation of platyrrhines that occurred only a few million years ago. This challenging phylogeny may be more easily resolved if the genomes of more representatives from the three New World

monkey families (and hopefully, one day, each genus and species) could be fully sequenced with deep coverage and made publicly available.

Increased analysis of squirrel monkey *Alu* elements could unearth key loci to determine the phylogenetic relationships in question. Within the parameters of this study, for example, hypothetical key loci could be shared *Alu* insertions in squirrel monkey, capuchin monkey, and owl monkey. The majority of squirrel monkey *Alu*S elements have not yet been analyzed, and it is probable that key loci remain in this subset of elements. Additionally, the three unsuccessfully sequenced loci should be re-sequenced to see if they yield informative results. Thus, further study of these elements could lead to the discovery of loci that might be able to resolve the phylogenetic relationships in question. Further examination of these elements would also provide the foundation to begin a subfamily analysis of squirrel monkey-specific *Alu* elements.

ACKNOWLEDGEMENTS

I would like to thank my thesis director, Dr. Mark Batzer, whose lab I have worked in for the past four years. His guidance, mentoring, and encouragement have been instrumental in this study as well as in my experiences as an undergraduate. I would also like to thank Dr. Miriam Konkel for her direction, advice, and support in this study. I am grateful to Jerilyn Walker for her continued help and editorial advice throughout this project. I thank Thomas Beckstrom for designing an automation code for computational analysis; it was not extensively used in this study, but it should prove useful in further analysis of the squirrel monkey genome. I am likewise grateful for the support of my peers in this project: Erika Arceneaux, Catherine Fontenot, Megan Ranck, Vallmer Jordan, and Michael Oldenburg.

I thank the following institutions and people for their contributions of DNA samples: American Type Culture Collection; Integrated Primate Biomaterial and Information Resource; Coriell Institute for Medical Research; San Diego Frozen Zoo, Conservation and Research for Endangered Species; New England Regional Primate Research Center; and Kimberly Phillips, Trinity University. This research was supported by National Institutes of Health Grant RO1 GM59290 (M.A.B.) and by a summer research grant from the Howard Hughes Medical Institute through the LSU College of Science undergraduate science education program (M.J.L.).

REFERENCES

Aristide, L, AL Rosenberger, MF Tejedor, and SI Perez. 2013. Modeling lineage and phenotypic diversification in the New World monkey (Platyrrhini, Primates) radiation. Mol Phylogenet Evol 11:008.

Barbulescu, M, G Turner, M Su, R Kim, MI Jensen-Seaman, AS Deinard, KK Kidd, and J Lenz. 2001. A HERV-K provirus in chimpanzees, bonobos and gorillas, but not humans. Curr Biol 11:779-783.

Batzer, MA, and PL Deininger. 2002. *Alu* repeats and human genomic diversity. Nat Rev Genet 3:370-379.

Cordaux, R, and MA Batzer. 2009. The impact of retrotransposons on human genome evolution. Nat Rev Genet 10:691-703.

Cost, GJ, Q Feng, A Jacquier, and JD Boeke. 2002. Human L1 element target-primed reverse transcription in vitro. EMBO J 21:5899-5910.

Craig, NL, R Craigie, M Gellert, and AM Lambowitz. 2002. Mobile DNA II. American Society for Microbiology Press; Washington.

de Koning, AP, W Gu, TA Castoe, MA Batzer, and DD Pollock. 2011. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7(12):e1002384.

Deininger, PL, and GR Daniels. 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet 2:76–80.

Deininger, PL, JV Moran, MA Batzer, and HH Kazazian. 2003. Mobile elements and mammalian genome evolution. Curr Opin Genet Dev 13:651–658.

Dewannieux, M, C Esnault, and T Heidmann. 2003. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35(1):41-48.

Esnault, C, J Maestre, and T Heidmann. 2000. Human LINE retrotransposons generate processed pseudogenes. Nat Gen 24:363-367.

Feng Q, JV Moran, HH Kazazian, and JD Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell 87: 905–916.

Groves, CP. 2005. INFRAORDER SIMIIFORMES. In Wilson, DE; Reeder, DM. Mammal Species of the World (3rd ed.). Baltimore: Johns Hopkins University Press. 128–152.

Hall, TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98.

Harvey, PH, RD Martin, and TH Clutton-Brock. 1987. Life histories in comparative perspective. Eds: BB Smuts, DL Cheney, RM Seyfarth, RW Wrangham, and TT Struhsaker. Primate Societies, University of Chicago Press, Chicago, IL: 181–196.

Hobolth, A, JY Dutheil, J Hawks, MH Schierup, and T Mailund. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res 21:349-356.

Judge, DS, and JR Carey. 2000. Postreproductive life predicted by primate patterns. J Gerontol A Biol Sci Med Sci 55:B201-B209.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc Natl Acad Sci USA 94(5):1872-1877.

Kazazian, HH Jr. 2004. Mobile elements: drivers of genome evolution. Science 303:1626-1632.

Kent, WJ. 2002. BLAT—the BLAST-like alignment tool. Genome Res 12:656-664.

Konkel, MK, JA Walker, and MA Batzer. 2010. LINEs and SINEs of Primate Evolution. Evol Anthr 19:236-49.

Kriegs, JO, G Churakov, et al. 2007. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet 23(4):158-61.

Lander, ES, LM Linton, B Birren, et al. (249 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Li, J, K Han, J Xing, HS Kim, J Rogers, OA Ryder, T Disotell, B Yue, and MA Batzer. 2009. Phylogeny of the macaques (Cercopithecidae: *Macaca*) based on Alu elements. Gene 448: 242-249

Luan, DD, MH Korman, JL Jakubczak, and TH Eickbush. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. Cell 72:595-605.

Mathias, SL, AF Scott, HH Kazazian Jr, JD Boeke, and A Gabriel. 1991. Reverse transcriptase encoded by a human transposable element. Science 254:1808–1810.

McClintock, B. 1950. The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA 36(6):344-55.

McClintock, B. 1956. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol 21:197-216.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Opazo, JC, DE Wildman, T Prychitko, RM Johnson, and M Goodman. 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol Phylogenet Evol 40:274-280.

Osterholz, M, L Walter, and C Roos. 2009. Retropositional events consolidate the branching order among New World monkey genera. Mol Phylogenet Evol 50:507-513.

Pace, JK II, and C Feschotte. 2007. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17(4):422-32.

Perez, SI, J Klaczko, and SF dos Reis. 2012. Species tree estimation for a deep phylogenetic divergence in the New World monkeys (Primates: Platyrrhini). Mol Phylogenet Evol 65:621-30.

Ray, DA, J Xing, DJ Hedges, et al. 2005. *Alu* insertion loci and platyrrhine primate phylogeny. Mol Phylogenet Evol 35:117-126.

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

Rozen, S, and HJ Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, 365-386.

Salem, AH, GE Kilroy, WS Watkins, LB Jorde, and MA Batzer. 2003. Recently integrated *Alu* elements and human genomic diversity. Mol Biol Evol 20(8):1349-1361.

Salem, AH, DA Ray, JC Xing, PA Callinan, JS Myers, DJ Hedges, RK Garber, DJ Witherspoon, LB Jorde, and MA Batzer. 2003. *Alu* elements and hominid phylogenetics. Proc Natl Acad Sci USA 100(22):12787-12791.

Sanger, F, S Nicklen, and AR Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. 74(12):5463–7.

Schneider H, I Sampaio, ML Harada, CML Barroso, MPC Schneider, J Czelusniak, and M Goodman. 1996. Molecular phylogeny of the New World monkeys (Platyrrhini, Primates) based on two unlinked nuclear genes: IRBP intron 1 and e-globin sequences. Am J Phys Anthropol 100:153-179.

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

Singer, MF. 1982. SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell 28(3): 433-434.

Smit AF. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 9(6):657-663.

Smit, AF, R Hubley, and P Green. 1996-2004. RepeatMasker Open-3.0.

Smit, AF, G Toth, et al. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246(3):401-417.

Smith, RJ, and WL Jungers. 1997. Body mass in comparative primatology. J Hum Evol 32:523-559.

Steiper, ME, and NM Young. 2006. Primate molecular divergence dates. Mol Phylogenet Evol 41:384-94

Sussman RW. 2000. Primate ecology and social structure. Volume 2, New world monkeys. Needham Heights (MA): Pearson Custom. p. 207.

Ulla, E, and C Tschudi. 1984. Alu sequences are processed 7SL RNA genes. Nature 312(5990):171-2.

Wildman, DE, NM Jameson, JC Opazo, and SV Yi. 2009. A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). Mol Phylogenet Evol 53:694-702.

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