4-2014

Exploring the Kothi Retrotransposon Through the Common Marmoset Genome

Erika Arceneaux

Exploring the Kothi Retrotransposon Through the Common Marmoset Genome

By Erika Arceneaux

Undergraduate 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 common marmoset (*Callithrix jacchus*) is a New World monkey endemic to Brazil. With its draft genome assembly produced in March 2009, the common marmoset was the first New World monkey to have a sequenced genome. For this reason, the common marmoset genome serves as a valuable tool for examining transposable elements. Here, a novel transposable element identified in the common marmoset genome, called the Kothi element, was the subject of our research. The characteristics of the Kothi element were analyzed, and it was found to possess many traits common to a specific type of transposable element called a retrotransposon.

Following manual analysis of the loci, primer pairs were designed using BioEdit and Primer3. The primers were used in PCR reactions with a panel of New World monkey DNA. Presence/absence data was obtained from the PCR results. From this data, it is clear that Kothi elements have been active retrotransposons in the genomes of New World monkeys. Following phylogenetic analysis, the loci were examined by PCR reactions in a common marmoset population genetics panel. From these PCR results, it was found that some Kothi elements remain polymorphic among marmoset populations, indicative of recent mobilization.

Introduction

The common marmoset (*Callithrix jacchus*) is a New World monkey, a term referring to a monophyletic group of primates indigenous to the tropical areas of South and Central America. Molecular evidence shows that these monkeys, also known as Platyrrhines, diverged from Catarrhines between 40 and 44 million years ago (Steiper and Young, 2006). Between 15 and 19 million years ago, Catarrhines split into Apes and Old World monkeys, which inhabit Asia and Africa and do not originate from South America (Schrago and Russo, 2003). Because New World monkeys evolved in an environment isolated from other primates, they are a very diverse group of species. New World

monkeys are small to medium-sized monkeys that have broad, flat noses (Fleagle, 1999). New World monkey species are in the Infraorder Platyrrhini, which is typically divided into three families: Atelidae (spider, woolly, and howler monkeys), Cebidae (callitrichids, capuchins, squirrel monkeys, and owl monkeys), and Pitheciidae (titi and saki monkeys). However, the branching order of these families has not been resolved with absolute confidence. The common marmoset is a member of the Cebidae family and the Callitrichine subfamily (Ray et al., 2005). Callitrichines diverged from other New World monkeys approximately 18 million years ago (Abbott et al., 2003) and represent the smallest and most distinctive New World monkeys. These monkeys are often brightly colored with little sexual dimorphism and very short snouts.

Specifically, marmosets consist of 22 species belonging to the genera *Callithrix, Cebuella, Callibella, and Mico* and are the smallest New World monkeys. The common marmoset, which is the focus of this study, belongs to the genus *Callithrix* (Groves, 2001). Unlike other platyrrhines, marmosets have some unique adaptations that prove essential for their diet and arboreal lifestyle. Marmosets have enlarged incisors that they use to bite holes in trees to gain access to tree exudates. This adaptation is specifically important for common marmosets. As a result of its seasonal forest habitats in northeastern and central Brazil, the common marmoset depends heavily on tree exudates throughout the year when fruit is scarce (Fleagle, 1999). Additionally, common marmosets have claw-like nails (tegulae) on all toes but the hallux (big toe) which aid them in their locomotion through trees. Common marmosets typically live in family groups comprised of 8-10 individuals (Stevenson and Rylands, 1988; Tardif and Smucny, 2003). Common marmoset social structure revolves around a stable, extended family unit containing a few dominant breeding individuals with flexible mating behavior (Sussman, 2000). Within the family unit, common marmosets practice cooperative breeding in which one couple acts as active breeders. Other members of the family group contribute in alloparental care (Solomon and French, 1997).

Combined with the common marmoset's compact body size, the family-style social grouping makes the common marmoset an ideal, low-maintenance animal for biomedical research (Abbott et al., 2003).

Common marmosets are not endangered in the wild and have been used for biomedical research since the 1960's. However, in the 1970's primate exports from South American were banned, which created a need to establish colonies. In the United States today, marmosets for research or laboratory use come from captive colonies in National Primate Research Centers, academic institutions, pharmaceutical companies, and commercial breeding facilities (Rylands, 1997). The marmosets housed in these colonies have been used in a variety of research inquiries. Unique attributes of marmosets present opportunities for biomedical research. They have been used in neuroscience studies on Alzheimer's (Geula et al., 2002), Parkinson's disease (Jenner et al., 2000), stroke (Bihel et al., 2010; Marshall et al., 2003), multiple sclerosis ('t Hart et al., 1998), Huntington's disease (Kendall et al., 1998), spinal cord injury (Iwanami et al. 2005), as well as post-traumatic stress disorder (Bremner et al., 1995; Bremner et al. 1997; Gurvits et al., 1996). The size of the marmoset's brain relative to its body is similar to humans and has a well-developed cerebral cortex and cerebellum, making the marmoset well-suited for studies in neuroscience. As a result of the marmoset's relatively high reproductive rate, marmoset embryonic stem cells were developed, which helped lead the way for the development of human embryonic stem cells (Thomson et al., 1998, Thomson et al., 1996). However, inconsistencies in marmoset embryonic stem cell development have held up any further development for their use as research tools (Thomson & Marshall, 1998).

Unlike other primates, marmosets and other Callitrichids frequently give birth to two or more offspring per pregnancy (Abbott et al., 2003). For example, in Jaquish et al. (1996), 39 term pregnancies were observed in one common marmoset female. Of these 39 pregnancies, the marmoset gave birth to five single offspring, 19 sets of twins, 13 sets of triplets, and 2 sets of quadruplets. The marmoset

embryos form vascular connections between their placentae which enable them to interchange DNA (Abbott et al., 2003; Moore et al., 1985). As a result, a phenomenon called chimerism occurs. A chimeric organism contains two unique sets of cells. The DNA contained within each cell of the chimeric individual is not identical as it is in non-chimeric individuals (Benirschke et al., 1962; Sweeney et al., 2012). It has been suggested that this chimerism acts as a factor in the development of community infant care among marmosets (Abbott et al., 2003; Haig, 1999). While the differences between marmosets and humans may be more numerable than those between Old World monkeys and humans, these physiological differences offer distinct opportunities for the utilization of marmosets in biomedical research (Abbott et al., 2003).

The use of marmosets as models in biomedical research can be aided by further understanding their genetic makeup. One aspect of marmoset genetics that is interesting in this respect is the makeup of mobile elements in the marmoset genome. Mobile elements are thought to affect the genomes of a wide range of organisms in a variety of ways. The majority of retrotransposon insertions are either neutral or deleterious to the host genome (Cordaux et al., 2006). Mobile element insertions have the potential to cause diseases by altering gene expression in these organisms, including humans (Callinan and Batzer, 2006). Mobile elements can disrupt gene expression by facilitating a deletion of genetic material. This can occur via insertional mutagenesis (Callinan et al., 2005; Gilbert et al., 2002) or mobile element-mediated recombination (Han et al., 2008; Lee et al., 2008; Sen et al., 2006). As mobile elements accumulate in the genome, it becomes more susceptible to nonallelic homologous recombination events due to the high number of similar sequences. These recombination events can result in genome rearrangements, which can result in deletions (Han et al., 2008; Konkel et al., 2010). Additionally, the genetic material of some mobile elements has been shown to become functional via a process called “exonization” (Lev-Maor et al., 2003). Consequently, mobile elements have a significant impact on the structural variation of the genome.

Mobile elements, or transposable elements, are segments of DNA that are mobile within the genome. They were first discovered by Barbara McClintock in *Zea mays* through the observation of multicolored kernel patterns (McClintock, 1956). In fact, mobile elements could make up more than two thirds of the human genome (de Koning et al., 2011). Transposable elements can be broken down into two categories—transposons and retrotransposons—based on the mechanism utilized to move throughout the genome. Transposons move using a "cut and paste" mechanism and a DNA intermediate (Craig et al., 2002). Retrotransposons are "copied and pasted" throughout the genome via an RNA intermediate. Retrotransposons can be further categorized into non-LTR (long terminal repeat) retrotransposons, which move via target primed reverse transcription (TPRT), and LTR retrotransposons, which do not utilize TPRT (Coffin et al., 1997). An LTR retrotransposon such as an endogenous retrovirus possesses long-terminal repeats, which are identical sequences of DNA that can repeat hundreds to thousands of times within an LTR element. However, this study focuses on non-LTR retrotransposons, which employ a mechanism called target primed reverse transcription (TPRT) to move throughout the genome (Cost et al., 2002; Luan et al., 1993). Furthermore, a non-LTR retrotransposon can be autonomous or non-autonomous. Autonomous elements code for the proteins necessary to carry out TPRT (Dombroski et al., 1991). A long interspersed element, or LINE, is an autonomous element because its open reading frames encode the enzymatic machinery used in TPRT. This enzymatic machinery includes a protein with both endonuclease and reverse transcriptase capabilities, as well as a chaperone protein that enhances retrotransposition, but is not required, for retrotransposition (Feng et al., 1996; Khazina et al, 2011). Non-autonomous elements lack the means to produce their own proteins but hijack the enzymatic machinery of the autonomous elements. Short interspersed elements, or SINE's, utilize the proteins of LINE's. For example, the primate specific *Alu* element, a SINE, uses the proteins encoded in the open reading frames of the human LINE-1 (L1) retrotransposon to move throughout the

genome (Dewannieux et al., 2003). The Kothi element was recently discovered. Its origin and characterization are part of this study. It is thought to be a non-autonomous, non-LTR retrotransposon.

Retrotransposons contain three recognizable features that play important roles in their movement via TPRT: an endonuclease cleavage site (endo site), target site duplications (TSDs) and an oligo dA-rich tail (Szak et al., 2002; Cordaux and Batzer, 2009). The first essential step in the retrotransposition mechanism calls for an RNA transcript of the repetitive element to be made. The transcript is produced by an RNA polymerase and will have an oligo dA-rich tail on the 3' end of the element. The RNA polymerase used in the transcription varies according to what retrotransposon is being copied. A nick is made at the target locus on one strand of the genomic DNA by an L1 endonuclease at an integration site (Jurka, 1997; Cordaux and Batzer, 2009). This is where the cleavage site comes into play. Endonuclease enzymes make nicks at sequences called cleavage sites, which can vary in length from 4 to 8 nucleotides. The cleavage site for L1 endonuclease typically has the sequence 5'-TTTT/AA-3'. However, it is not essential that the integration site contains this exact sequence (Repanas et al., 2009). The L1 endonuclease makes the cut between the four T's and the 2 A's, making the integration site available for the element. The oligo dA-rich tail of the RNA transcript then anneals to the TTTT sequence made available after the cut. Following this binding, the L1 reverse transcriptase carries out reverse transcription of the element. Following the reverse transcription, a second nick is made on the complementary strand of the target DNA and the element is fully integrated into the target site. The mechanism behind the second nick is currently not well understood (Dewannieux et al., 2003). Movement via TPRT results in staggered DNA cuts in the genome, and DNA polymerase carries out the extension of complementary DNA to fill in these cuts. After this occurs, the element is flanked by short sequences of identical host DNA called Target Site Duplications. TSDs are the primary evidence that TPRT has occurred (Szak et al., 2002; Cordaux and Batzer, 2009). They are one of the mechanism's principle characteristics and thus, a defining characteristic of retrotransposons.

As the result of their “copy and paste” mechanism, retrotransposons accumulate in hierarchical manner. Moving in this “identical by descent” fashion means that the ancestral state of a retrotransposon insertion at any locus is the absence of that insertion. It also indicates that a shared insertion of a retrotransposon with identical TSDs at a given locus signifies a common ancestor (Salem et al., 2005; Ray et al., 2006). Namely, if a retrotransposon insertion is made in a primate genome, it is not the result of convergent evolution. Consequently, any two primates that share an insertion at the same locus must have shared a common ancestor. It is unlikely that all traces of a retrotransposon at a locus would be removed, as precise deletions are rare (Ray et al., 2005; Ray et al., 2006). However, major deletions that occur as the result of transposable element mediated recombination are more common. Deletions of this type contribute to genome instability because recombination between two mobile elements could result in the deletion of a functional region that contains multiple genes, as well as a deletions of part of the mobile element. It is also possible for an element to be truncated. In fact, most LINE-1 elements are significantly truncated upon insertion into the human genome (Vincent et al., 2003). This unidirectional means of moving in the genome is what allows researchers to use retrotransposons to elucidate phylogenetic relationships without making needless assumptions. While the use of other types of genetic markers may yield unclear results because they are “identical by state,” retrotransposons allow for a powerful, nearly homoplasy-free method of identifying phylogenetic relationships because they are “identical by descent” (Konkel et al., 2010; Ray et al, 2006; Xing et al., 2007). In this study, a novel retrotransposon called the Kothi element was analyzed and used to examine the phylogenetic relationships between New World monkeys. Through this analysis, more information can be obtained regarding the genetics of marmosets and New World monkeys. Learning more about the genetics of marmosets can benefit the scientists using marmosets for biomedical research.

Materials and Methods

Manual Analysis of Kothi Loci:

A consensus sequence for the Kothi element was constructed and used in a local installation of RepeatMasker (Smit et al., 1996). The common marmoset reference genome (CalJac3.2) was run through the local installation of RepeatMasker using a custom library, which included the Kothi element consensus sequence, along with the *Alu* consensus sequence. These RepeatMasker (Smit et al., 1996) results were used with a local installation of BLAT with custom scripts (Kent, 2002) to obtain all Kothi element sequences in the common marmoset reference genome (CalJac3.2). Additionally, each element was retrieved with 500 bp of flanking sequence one each side of the element from the marmoset genome (CalJac3.2) using a local installation of BLAT (Kent, 2002). These sequences were put into a BioEdit file (Hall 1999). For purposes of this study, 1,125 Kothi elements were aligned and analyzed. New World monkey specificity for the Kothi element was confirmed by using BLAT to compare the common marmoset (CalJac3.2) and human (hg19) reference genomes. Following the verification, target site duplications, endonuclease cleavage site, and oligo dA-rich tail were identified and recorded for each locus in an Excel spreadsheet. Excel sorting functions were used to obtain data used to create graphs regarding TSD length distribution and variation in oligo dA-rich tail length. It is important to note that when recording the length of oligo dA-rich tails, any A's that were present in the TSD were not included in the length of the oligo dA-rich tail. The endonuclease cleavage site sequences for each locus were plugged into a web-based application called WebLogo which generated a graphical figure illustrating the most common bases in each position of the cleavage site (Crooks et al., 2004).

Primer Design:

For locus-specific primer design, the youngest appearing candidate loci (i.e. loci with the fewest number of random mutations) were selected from each Kothi subfamily. In addition, Kothi elements had

to be close to full-length. For each candidate locus, orthologous sequences from the human (hg19), chimpanzee (panTro4), orangutan (ponAbe2), rhesus macaque (rheMac3), and squirrel monkey (saiBol1) genome assemblies were obtained from BLAT. The sequences were aligned using BioEdit (Hall 1999). Primers were designed by locating the best conserved regions of flanking sequence and subsequently using Primer3 (Rozen and Skaletsky 2000). Default settings for Primer3 were used with the following exceptions: a minimum annealing temperature of 57°C, a maximum temperature of 61°C, and a maximum of 4 consecutive identical nucleotides. Results from Primer3 were chosen according to the level of apparent conservation in the orthologous sequences where the primer was located. In other words, the primer with the least number of mutations in the orthologous sequences was selected. Additionally, primers with mutations within the 3' end were avoided. The chosen primer was then run through BLAT to verify its uniqueness in the marmoset genome. Following this verification, the primers were run through the UCSC in-silico PCR program to be certain that the primers resulted in only one PCR product and to record the expected amplicon sizes of the empty sites which lack the Kothi insertion and filled sites which contain the insertion (Kent 2002). The Sigma-Aldrich (Woodlands, TX, USA) stock primers were then resuspended to a concentration of 100 µM and diluted with TLE buffer (10mM Tris-HCl / 0.1mM EDTA) to 2 µM. Once this was complete, each primer pair was tested using PCR with HeLa DNA (human cell line HeLa-CCL-2 from ATCC) (Table 1).

Polymerase Chain Reaction:

Primers were tested with PCR using the following protocol: initial denaturation at 95°C for one minute, 32 cycles of denaturation at 95°C for 30 seconds, annealing at 57°C (if not otherwise indicated) for 30 seconds, and extension at 72°C for 30 seconds, the protocol ended with additional extension at 72°C for two minutes and a holding temperature of 25°C. If the primer pair amplified successfully with the annealing temperature of 57 °C, the chosen loci were run on a phylogenetic panel containing the

DNA of 11 different New World monkeys with TLE as a negative control, and HeLa, chimpanzee, and African green monkey as outgroups (Table 1). If the primer pair did not amplify successfully in the test, PCR was run with HeLa DNA and common marmoset DNA (NE 8) using a protocol with a 50°C -60°C annealing temperature gradient in order to determine an appropriate annealing temperature. Once the appropriate annealing temperature was determined for the locus, it was run on the New World monkey phylogenetic panel. The thermal cycler protocol for the phylogenetic panel was consistent with that of the aforementioned primer test with adjustments of the annealing temperature made when necessary. The reactions were run using a Perkin Elmer GeneAmp 9700 or a BioRad i-cycler thermal cycler and MicroAmp Optical 96-well Reaction Plates with a volume of 25 µl used for each PCR reaction. Each PCR reaction contained the following ingredients: 15-25 ng of template DNA, 200 nM of each oligonucleotide primer, 1.5-2 mM $MgCl_2$, 1X PCR buffer (50 mM KCl; 10 mM TrisHCl, pH 8.3), 0.2 mM dNTPs; and 1-2 U *Taq* DNA polymerase. Following the PCR, the PCR products were size fractioned using gel electrophoresis at 200V for 50 -70 minutes on 2% agarose gels with 0.1 µg/mL ethidium bromide. Following electrophoresis, the loci were visualized using UV light. Presence/absence data for each locus was then recorded.

If a locus amplified only in common marmoset in the phylogenetic panel, PCR using a common marmoset population panel was performed. The population genetic panel (Table 2) contained DNA from 24 common marmoset individuals from three different colonies. HeLa was used as an empty site control amplifying the pre-insertion site, and pygmy marmoset was used to verify that the locus was common marmoset-specific, also amplifying the pre-insertion site. The thermal cycler protocol was identical to that used for the phylogenetic panel. Again, PCR products were size fractioned using gel electrophoresis. Presence/absence data for each locus was recorded following visualization with UV light.

DNA Sequencing:

If a fragment of unexpected size amplified in a phylogenetic species panel, meaning an aberrant band amplified unexpectedly, DNA sequencing was performed. The aberrant fragments of PCR reactions were excised from the agarose gel using a straight razor blade. PCR products Gel slices were purified using Wizard SV Gel and PCR Purification System from Promega (Promega Corporation, Madison, WI, USA, catalog A9282). The standard protocol was used for all but one step. We performed two final elution steps with 50 µl Nuclease-Free Water instead of the one listed in the Wizard System's standard protocol. This was done to maximize the yield of DNA sample from the purification process to be used in the next step: sequencing PCR. Once purified, the DNA for each locus was used for sequencing PCR with the following thermal cycler protocol: hot start at 95° C, initial denaturation at 95°C for two minutes, 40 cycles of denaturation at 95°C for 10 seconds, annealing at 50°C for 5 seconds, and extension at 60°C for 4 minutes, and a holding temperature of 4°C. Following the sequencing PCR, the PCR products were precipitated using EDTA, 100% ethanol, and 70% ethanol. The PCR products were sequenced using Sanger sequencing on an ABI 3130xl genome analyzer. This was done in order to confirm the presence or absence of a Kothi element. Once sequence was obtained, the next step was to analyze and align it with the original locus containing a Kothi element marmoset using DNASTAR Version 5.0 and BioEdit. From here, it was determined whether the Kothi insertion was in fact shared, or if the site appeared to be filled due to the presence of a different repetitive element.

Table 1. New World monkey Phylogenetic Panel

#	Species	Common name	Origin	ID
1	*Homo sapiens*	Human	ATCC	HeLa CCL-2
2	*Pan troglodytes*	Common Chimpanzee	IPBIR	NS06006
3	*Chlorocebus aethiops*	African Green monkey	ATCC	CCL70
4	*Lagothrix lagotricha*	Woolly monkey	Coriell	NG05356
5	*Ateles belzebuth*	White bellied spider monkey	SDFZ	KB6701
6	*Alouatta sara*	Bolivian red howler monkey	SDFZ	OR749
7	*Callithrix jacchus*	Common marmoset	NERPRC	cj393-99, A02-738
8	*Callithrix pygmea*	Pygmy marmoset	SDFZ	OR690
9	*Saguinus fuscicollis nigrifrons*	Geoffroys saddle-back tamarin	SDFZ	OR621
10	*Cebus*	Capuchin monkey	KP	CA003
11	*Saimiri s. sciureus*	Squirrel monkey	SDFZ	KB4544
12	*Aotus trivirgatus*	Owl monkey	ATCC	CRL1556
13	*Pithecia p. pithecia*	Northern white-faced saki	SDFZ	OR842
14	*Callicebus d. donacophilus*	Bolivian gray titi	SDFZ	OR1522

ATCC: From cell lines provided by the American Type Culture Collection
IPBIR: Integrated Primate Biomaterials and Information Resource
Coriell: Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ
SDFZ: San Diego Frozen Zoo, Conservation and Research for Endangered Species (CRES)
NERPRC: New England Regional Primate Research Center
KP: Kimberly Phillips, Trinity University

Table 2. Common Marmoset Population Panel

#	Sex	Litter Makeup	PRC Origin	ID	Alternate ID	Other ID
1	F	Unknown	NERPRC	cj 337-00, A03,168	Cm_NE-10	cj 337-00
2	F	Unknown	NERPRC	cj 15-98, A02-446	Cm_NE-6	cj 15-98
3	F	Unknown	NERPRC	cj 65-98, A03-145	Cm_NE-9	cj 65-98
4	M	Unknown	NERPRC	cj 99-01, A02-485	Cm_NE-3	cj 99-01
5	M	Unknown	NERPRC	cj 389-97, A02-387	Cm_NE-1	cj 389-97
6	F	Unknown	NERPRC	cj 12-97, A02-737	Cm_NE-7	cj 12-97
7	M	Unknown	NERPRC	cj 215-99, A02-677	Cm_NE-5	cj 215-99
8	F	Unknown	NERPRC	32782	Cm_1-2008	1-2008
9	M	Unknown	NERPRC	cj 501-98, A02-498	Cm_NE-2	cj 501-98
10	M	Unknown	NERPRC	cj 72-98, A02-584	Cm_NE-4	cj 72-98
11	F	Unknown	NERPRC	cj 393-99, A02-738	Cm_NE-8	cj 393-99
12	F	Unknown	NERPRC	Cm_SW-27537	Cm_SW-27537	27537
13	M	3 males	SWNPRC	Cm_SW-19478	Cm_SW-19478	19478
14	F	Unknown	SWNPRC	Cm_SW-27552	Cm_SW-27552	27552
15	M	2m., 1f	SWNPRC	Cm_SW-17461	Cm_SW-17461	17461
16	M	2 males	SWNPRC	Cm_SW-17953	Cm_SW-17953	17953
17	M	3 males	SWNPRC	M32784	Cm_272	272
18	M	1 male	SWNPRC	Cm_SW-19570	Cm_SW-19570	19570
19	F	2 females	SWNPRC	Cm_SW-25556	Cm_SW-25556	25556
20	?	Unknown	WNPRC	Cm_34040	Cm_34040	CJ0910
21	?	Unknown	WNPRC	Cm_34044	Cm_34044	CJ1544
22	?	Unknown	WNPRC	33426	Cm_CJ1590	CJ1590
23	?	Unknown	WNPRC	Cm_34042	Cm_34042	CJ1660
24	?	Unknown	WNPRC	Cm_34043	Cm_34043	CJ1580

NERPRC: New England Regional Primate Research Center
SWNPRC: Southwest National Primate Research Center
WNPRC: Wisconsin National Primate Research Center

Results and Discussion

Manual Analysis of Kothi elements:

Data obtained through the analysis of 1,125 Kothi elements suggests that the Kothi element is a non-autonomous retrotransposon. The Kothi element possesses three characteristics indicative of a non-LTR retrotransposon: an endonuclease cleavage site, target site duplications, and an oligo dA-rich tail, that are shown in Figure 1. The element is typically ~120 base pairs in length including the oligo dA-rich tail at the 3' end of the element. Deviation in total element length occurs due to variation in the length of the oligo dA-rich tail. Most oligo dA-rich tails are between 11 and 18 base pairs in length with an overall range of 11-40+ base pairs (Figure 1). A portion of the Kothi elements analyzed contained oligo dA-rich tails shorter than 11 base pairs. However, this only occurred when a significant portion of the oligo dA-rich tail was included in the TSD. The length of the element's TSDs also varies considerably. Like the oligo dA-rich tail, most of the TSD's are between 11 and 18 base pairs in length. TSD length in the Kothi element ranges from 4 to 40 base pairs (Figure 2). For the L1 endonuclease, a consensus sequence of six nucleotides (5'-TTTT/AA-3') is recognized as the cleavage site. Of the 1,125 Kothi elements, the most common base in each position of the site is represented in Figure 3. The consensus sequence in the figure (5'-TTTT/AA-3') is in agreement with the accepted consensus sequence for the endonuclease cleavage site of L1 endonuclease. However, the relationship between L1 endonuclease and this consensus sequence is not strict. Out of the 1,125 Kothi elements analyzed, 209 possessed this exact sequence at the endonuclease cleavage site.

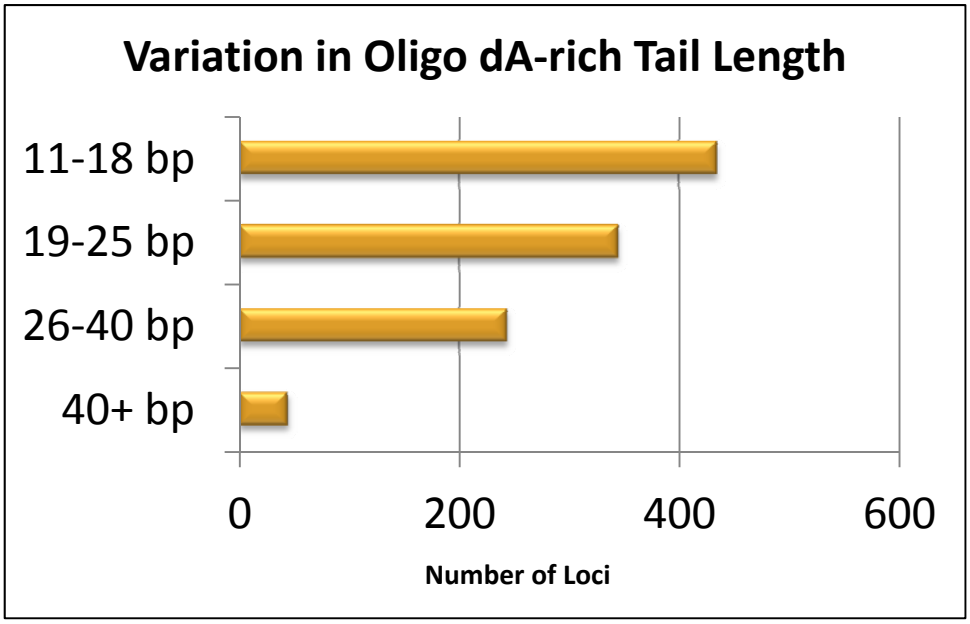


Figure 1. Oligo dA-rich Tail Length Distribution among Kothi loci. A length between 11 and 18 base pairs was most common.

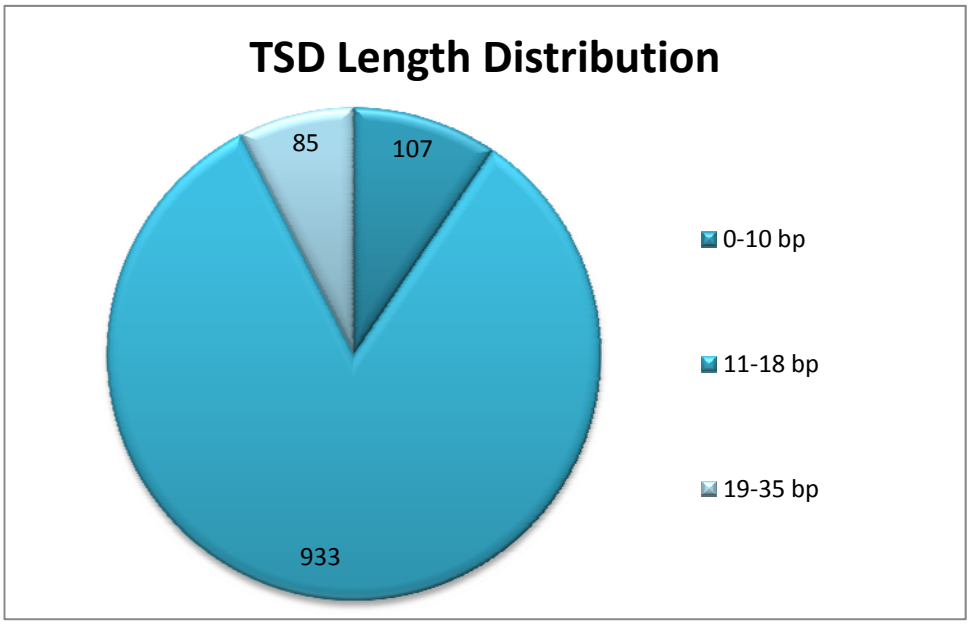


Figure 2. TSD Length Distribution among Kothi loci. The majority of the 1,125 Kothi loci analyzed had TSD's 11-18 base pairs in length. The chart shows the number of loci possessing TSDs of the specified length.

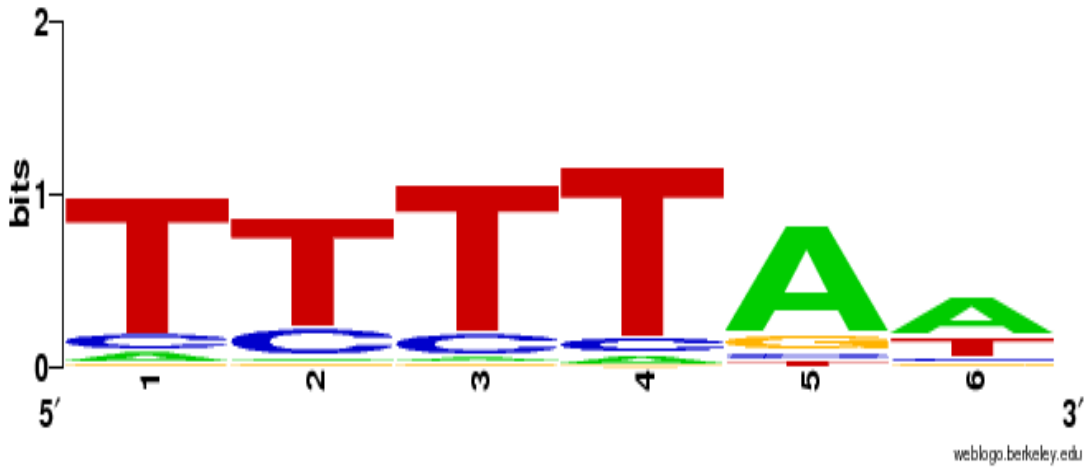


Figure 3. Distribution of Endonuclease Cleavage Site Sequences. From the analysis of 1,125 Kothi element loci, the most common bases at each position are shown. The height of each stack of residues is measured in bits (shown on y-axis). The horizontal axis indicates the position in the cleavage site.

Phylogenetic and Population Analysis:

Of the 1,125 Kothi elements analyzed, oligonucleotide primer pairs were designed for 152 loci and tested for amplification. Forty-three of these primer pairs did not amplify initially and needed to be redesigned. With these additional primer pairs, a total of 195 loci were tested for amplification. Out of the total designed primer pairs, 156 were considered usable for phylogenetic analysis. The reliability of the panel results was based on how many of the panel species amplified. No more than two amplification failures were allowed for a panel to be considered successful. Out of 156 phylogenetic panel results, 81 loci amplified in enough species of the panel to be used in this study. PCR results from the 81 successful phylogenetic panels, supported the results for the manual analysis of 1,125 Kothi elements: the Kothi element is a New World monkey specific mobile element. The PCR results also show that there are polymorphic loci present in the common marmoset. Phylogenetic panel results from successful PCR generated four outcomes. The first option was that the locus was present in all New

World monkeys on the panel (7 of 79). The element sequences at these loci were usually the most mutated, indicating that they were older than the other elements analyzed. The second was that the Kothi element insertion was present in common marmoset, pygmy marmoset, and tamarin (24 of 79). The third was for the locus to be present in both common and pygmy marmoset (38 of 79). The fourth was that the locus was present in only common marmoset (12 of 79). Examples of loci for these results are shown in Figure 4A-D with the percent distribution shown in Figure 4E. Any loci that did not originally fall within these four categories were resolved with DNA sequencing.

Sanger sequencing was performed on a total of 14 loci with nine yielding successful results. Four of the remaining five loci did not yield reliable DNA sequences following Sanger sequencing. One of the five unsuccessful loci did not amplify in sequencing PCR. A locus was chosen to be sequenced if the PCR amplicon was longer than expected. This would appear as a larger than expected band on an agarose gel following UV visualization. Two of the 14 selected loci possessed amplicons that appeared to be the same size as the Kothi insertion. The remaining 12 loci had fragments that were significantly larger in size than a Kothi insertion, and they were sequenced in order to determine if the large amplicon contained a Kothi element within it. Three of the nine successfully sequenced loci showed the presence of a shared Kothi insertion between the species sequences and the common marmoset. The species of these three loci were the capuchin monkey, squirrel monkey, and pygmy marmoset. The remaining six successfully sequenced loci did not show shared Kothi insertions. Unique upstream and downstream sequences were able to be aligned with common marmoset, indicating shared unique sequence. In four of the six loci, the downstream alignment was less than 50 base pairs. In these alignments, it was apparent that the sequenced loci did not contain the Kothi element insertion at their respective loci. When run through RepeatMasker (Smit et al., 1996), a sequence from owl monkey was recognized as an *Alu* insertion. For the remaining five loci, it was determined that the extra sequence in the amplicon did

not contain a Kothi element. For each atypical locus, the sequencing results lead to the locus belonging to one of the four pre-defined categories shown in Figure 4.

The Kothi element loci specific to the common marmoset are the youngest. In agreement with the common-marmoset specific elements being the youngest, these loci were the most similar to each other and had the least number of random mutations. A Kothi insertion specific to common marmoset could be polymorphic within the common marmoset population. To test this, Kothi elements specific to common marmoset were tested on a population panel that contained 24 common marmosets from three Primate Centers. Out of 12 common marmoset-specific loci run on a common marmoset population panel, 10 were fixed present for the insertion in all 24 common marmoset individuals, while 2 loci were polymorphic (variable presence / absence pattern) within the common marmoset populations. The PCR results from a polymorphic locus determine whether the individual is homozygous or heterozygous for the presence of the Kothi element (Figure 5). The individual can have a homozygous filled site, meaning the marmoset possesses two alleles with the Kothi insertion, or a homozygous empty site indicating two alleles without the insertion. A possible third option is that the individual is heterozygous for the allele, meaning that the individual contains one filled allele and one empty allele. A polymorphic element could still be active in the common marmoset lineage, indicating that it is not yet fixed at that locus within the population. Ongoing retrotransposition generates both inter- and intra-population diversity.

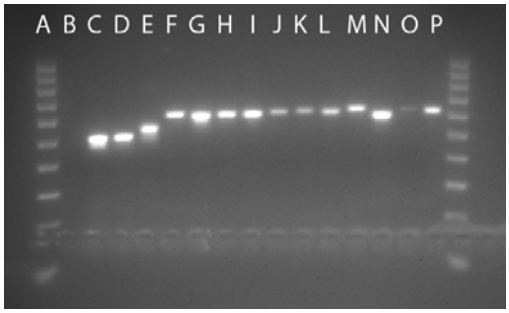
Because of the retrotransposon character of the Kothi element, it can be used as a phylogenetic marker. Non-LTR retrotransposons are considered to be “identical by descent,” meaning that a shared insertion at a locus indicates a common ancestor. As a result of the “copy and paste” mechanism used by non-LTR retrotransposons, using the Kothi retrotransposon as a phylogenetic marker is a nearly homoplasy-free method. The results of the phylogenetic analysis were consistent with, and added

support to, the current view of New World monkey phylogeny and did not further resolve any additional lineage branches. Analysis using the common marmoset population panel revealed that some Kothi elements are not fixed in the common marmoset population. These polymorphic loci could still be actively carrying out retrotransposition in the common marmoset genome. It is also possible that retrotransposition may have ceased very recently. It is important to note that active retrotransposition is not exclusive to polymorphic loci. While the activity level of polymorphic loci is more clearly visible using laboratory methods, fixed loci can still have the ability to propagate throughout the genome. Even if a locus is fixed within a population, it can still be an active retrotransposon.

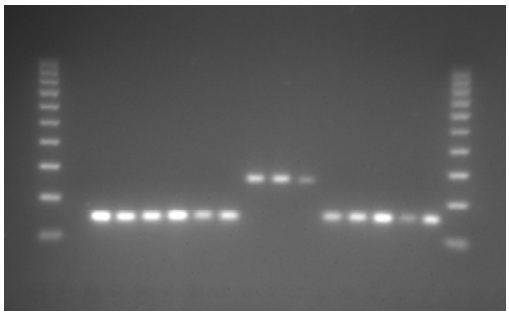
Conclusions:

The Kothi element is a non-autonomous retrotransposon that is specific to New World monkeys. When used in New World monkey phylogenetic analysis, the presence and absence data of the Kothi element confirmed and supported the current view of New World monkey phylogeny. Although the vast majority of Kothi element loci appeared fixed present in the New World monkey phylogenetic panel, two of the 79 loci were found to be polymorphic. The presence of these loci indicates recent or possibly ongoing retrotranspositional activity of the Kothi element in the common marmoset genome, which demonstrates continued diversity between the three common marmoset populations analyzed. This study illustrates and analyzes the key features of the Kothi retrotransposon in the common marmoset genome, as well as indicates the Kothi element's value in phylogenetic and population genetics studies.

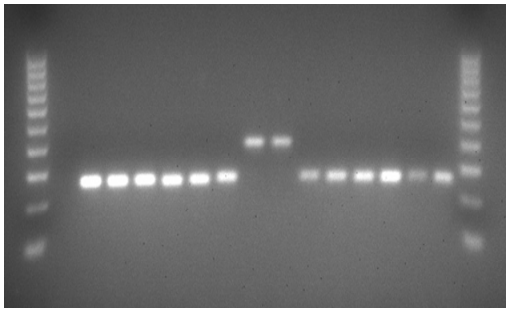
A)



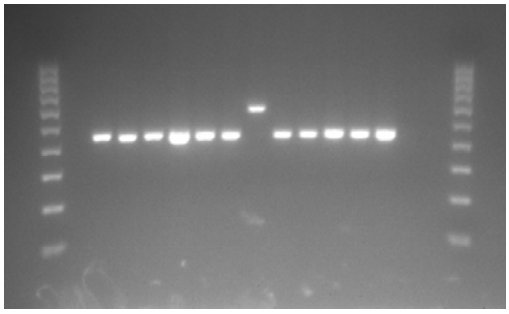
B)



C)



D)



Label	Sample in New World monkey Panel
A	100 bp ladder
B	TLE (negative control)
C	Human
D	Common chimpanzee
E	African green monkey
F	Woolly monkey
G	Spider monkey
H	Howler monkey
I	Common marmoset
J	Pygmy marmoset
K	Geoffroy's saddle-back tamarin
L	Capuchin monkey
M	Squirrel monkey
N	Owl monkey
O	Northern white-faced saki
P	Bolivian gray titi

E)

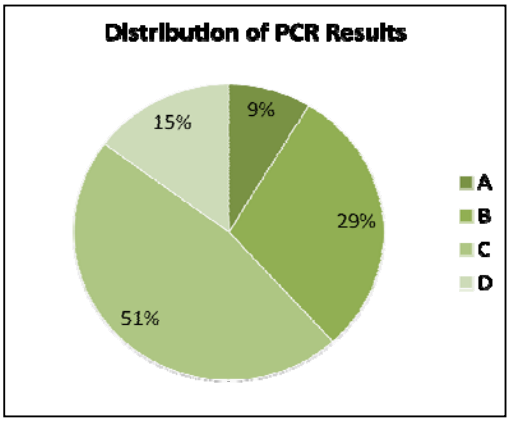


Figure 4. Agarose gel pictures of Kothi element insertions. A) All New World monkeys in the panel contain the element. B) The element is present in common marmoset, pygmy marmoset, and tamarin (higher bands). C) Common and pygmy marmoset contain element. D) Common marmoset only contains element. E) Graph representing distribution of PCR results. Segments A-D represent the four pictures labeled A-D. The most common result was Figure 4C.

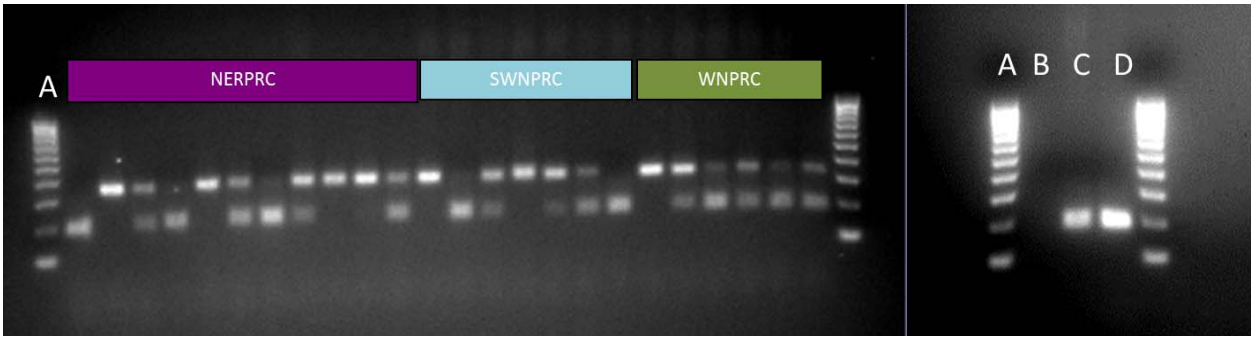


Figure 5. Agarose gel picture of a polymorphic locus. Common marmoset individuals from the New England, Southwest, and Wisconsin colonies were analyzed. It is clearly visible that an individual can possess one of three different genotypes: homozygous empty (bottom band only), homozygous filled (top band only), or heterozygous sites (both bottom and top bands). [A: 100 bp ladder; B: TLE; C: Human; D: Pygmy marmoset; NERPRC: New England Regional Primate Research Center; SWNPRC: Southwest National Primate Research Center; WNPRC: Wisconsin National Primate Research Center]

Future Work:

In order to enhance the scope of this study, continued phylogenetic and population genetic studies are necessary. Further the analysis of the Kothi element could aid in determining where exactly the Kothi element is derives from. Because it is newly discovered, the Kothi element does not have the same background and history available that other retrotransposons have, such as *Alu* and LINE-1. Further study with a more fine tuned analysis would be beneficial in this respect. Additionally, further population genetic studies would aid in preserving the diversity between the colonies of common marmosets. Prevention of extreme homogenization or extreme diversification is essential to continued health of the population, as well as continued use in biomedical research. Study of the Kothi element could also be broadened to the genomes of other New World monkeys, such as the squirrel monkey.

Additionally, the growing field of personalized medicine creates the need for more comparative genomic research. DNA sequencing and genome assembly are becoming less expensive, and the opportunity to directly compare genomes will become more available to the medical field. Because

mobile elements have been shown to cause disease in humans, mobile element research can enhance the field of medicine. As a future physician, the possibility treating someone used an individualized treatment based on their genetics excites me. I am certain than future work in this field will greatly impact the world of medicine in my lifetime.

Acknowledgements:

First and foremost, I would like the thank Dr. Mark Batzer, my thesis director, for allowing me to join his lab three years ago. He has been a source of essential guidance, advice, and opportunity throughout my time here. My experiences in Dr. Batzer's lab have taught me valuable lessons that I will carry with me throughout my life. I would also like to thank the members of the Batzer Lab for answering my *many* questions and guiding me in the right direction: Jerilyn Walker, Vallmer Jordan, Michelle Limbaugh, Catherine Fontenot, and Sreeja Sanampudi. A special thank you belongs to Dr. Miriam Konkel for providing the direction for this project and encouraging me to think for myself throughout the process. I thank Dr. Mark Batzer, Dr. Miriam Konkel, and Dr. John Protevi for serving on my Thesis Committee.

I thank the New England, Southwest, and Wisconsin Primate Centers for their contributions of DNA samples, as well as the American Type Culture Collection, Integrated Primate Biomaterials and information Resource, the Coriell Institute for Medical Research, Kimberly Phillips of Trinity University, and the Sand Diego Frozen Zoo for their contributions. This research was supported by National Institutes of Health R01 GM59290 (M.A.B.). I would also like to thank Howard Hughes Medical Institute for summer research funding (E.L.A.)

References

't Hart, B. A., J. Bauer, H. J. Muller, B. Melchers, K. Nicolay, H. Brok, R.E. Bontrop, H. Lassmann, and L. Massacesi. 1998. Histopathological characterization of magnetic resonance imaging-detectable brain white matter lesions in a primate model of multiple sclerosis: a correlative study in the experimental autoimmune encephalomyelitis model in common marmosets (Callithrix jacchus). *American Journal of Pathology* 153:649-663.

Abbott, D.H., D.K. Barnett, R.J. Colman, M.E. Yamamoto, and N.J. Schultz-Darken. 2003. Aspects of common marmoset basic biology and life history import for biomedical research. *Comparative Medicine* 53:339-50.

Benirschke, K., J.M. Anderson, and L.E. Brownhill. 1962. Marrow chimerism in Marmosets. *Science* 138: 513-515.

Bihel, E., P. Pro-Sistiaga, A. Letourneur, J. Toutain, R. Saulnier, R. Insausti, M. Bernaudin, S. Roussel, and O. Touzani. 2010. Permanent or transient chronic ischemic stroke in the non-human primate: behavioral, neuroimaging, histological, and immunohistochemical investigations. *Journal of Cerebral Blood Flow and Metabolism* 30:273-285.

Bremner, J.D., P. Randall, E. Vermetten, L. Staib, R.A. Bronen, C. Mazure, S. Capellli, G. McCarthy, R.B. Innis, and D.S. Charney. 1997. Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse—a preliminary report. *Biological Psychiatry* 41:23-32.

Bremner, J.D., P. Randall, T.M. Scott, R.A. Bronen, J.P. Seibyl, S.M. Southwick, R.C. Delaney, G. McCarthy, R.B. Innis, and D.S. Charney. 1995. MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *American Journal of Psychiatry* 152:973-981.

Callinan P.A., J. Wang, S.W. Herke, R.K. Garber, P. Liang, and M.A. Batzer. 2005. Alu retrotransposition-mediated deletion. J Mol Biol 2005;348:791–800.

Callinan, P. A. and M. A. Batzer. 2006. Retrotransposable elements and human disease. *Genome Dynamics* 1:104-115 .

Coffin, J.M, S.H. Hughes, and H.E. Varmus. 1997. *Retroviruses.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Cordaux, R. and M. A. Batzer. 2009. The impact of retrotransponsons on human genome evolution. *Nature Reviews Genetics* 10: 691-703.

Cordaux, R., J. Lee, L. Dinoso and M. A. Batzer. 2006. Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene* 373:138-144.

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

Craig, N. L., R. Craigie, M. Gellert, and A.M. Lambowitz. 2002. *Mobile DNA II*. Washington, D.C.: American Society for Microbiology.

Crooks, G.E., G. Hon, J.M. Chandonia, S.E. Brenner. 2004. WebLogo: A sequence logo generator. *Genome Research* 14:1188-1190.

de Koning, A. P. J., W. Gu, T. A. Castoe, M. A. Batzer, and D. D. Pollock. 2011. Repetitive elements may comprise over two-thirds of the human genome*. PLoS Genetics* 12:e1002384.

Dewannieux, M., C. Esnault, and T. Heidmann. 2003. LINE-mediated retrotransposition of marked Alu sequences*. Nature Genetics* 35:41–48.

Dombroski, B.A., S.L. Mathias, E. Nanthakumar, A.F. Scott, and H.H. Kazazian Jr. 1991. Isolation of an active human transposable element. *Science* 254:1805-1808.

Feng, Q., J.V. Moran, H. Kazazian Jr.,and J.D. Boeke. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87:905–916.

Fleagle, J.G. 1999, *Primate Adaptation and Evolution*. 2nd ed. San Diego: Academic Press.

Geula, C., N. Nagykery, and C.K. We. 2002. Amyloid-beta deposits in the cerebral cortex of the aged common marmoset (*Callithrix jacchus*): incidence and chemical composition. *Acta Neuropathologica* 103:48-58.

Gilbert, N., S. Lutz-Prigge, J.V. Moran. 2002. Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110:315–25.

Groves, C. 2001. *Primate Taxonomy*. Washington D.C.: Smithsonian Institution Press.

Gurvits, T.V., M.E. Shenton, H. Hokama, H. Ohta, N.B. Lasko, M.W. Gilbertson, S.P. Orr, R. Kikinis, F.A. Joesz, R.W. McCarley, and R.K.Pitman. 1996. Magnetic resonance imaging study of hippocampal volume in chronic combat-related posttraumatic stress disorder. *Biological Psychiatry* 40:1091-1099.

Haig, D. 1999. What is a marmoset? *American Journal of Primatology* 49**:**285-296.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acid Symposium Series* 41:95-98.

Han, K., J. Lee, T.J. Meyer, P. Remedios, L. Goodwin, and M.A. Batzer. 2008. L1 recombination-assoicated deletions generate human genomic variation. *Proceedings of the National Academy of Sciences*, USA 105:19366-19371.

Iwanami, A., J. Yamane, H. Katoh, M. Nakamura, S. Momoshima, H. Ishii, Y. Takoika, N. Tamaoki, T. Nomura, Y. Toyama, and H. Okano. 2005. Establishment of graded spinal cord injury model in a nonhuman primate: the common marmoset. *Journal of Neuroscience Research* 80:172-181.

Jaquish, C.E., S.D. Tardif, R.L. Toal and R.L. Carson. 1996. Patterns of prenatal survival in the common marmoset (Callithrix jacchus). *Journal of Medical Primatology* 25:57- 63.

Jenner, P. L. Smith, M. Jackson, M. Hansard, C. Bonhomme, and C. Chezaubernard. 2000. Piribedil induces low levels of dyskinesia in MPTP-treated common marmosets. *Neurology* 54:A53.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proceedings of the National Academy of* Sciences, USA 94:1872–1877.

Kendall A.L., F.D. Rayment, E.M. Torress, H.F. Baker, R.M. Ridley, and S. B. Dunnett. 1998. Functional integration of striatal allografts in primate model of Huntington's disease. *Nature Medicine* 4: 727-729.

Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome Research* 12:996-1006.

Kent, W.J.. BLAT - the BLAST-like alignment tool. 2002. *Genome Research* 12:656-64.

Khazina E., V. Truffault, R. Buttner, S. Schmidt, M. Coles, & O. Weichenrieder. 2011. Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nature Structural & Molecular Biology* 18:1006-1014.

Konkel, M. K., J. A. Walker and M. A. Batzer**.** 2010. LINEs and SINEs of primate evolution. *Evolutionary Anthropology* 19:236-249.

Lee, J., K. Han, T.J. Meyer, H.S. Kim, and M.A. Batzer. 2008. Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS One* 3:e4047.

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

Luan, D.D., M.H. Korman, J.L. Jakubczak, and T.H. Eickbush. 1993. Reverse transcription of R2BmRNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595-605.

McClintock, B. 1956. Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* 21:197-216.

Moore, H.D., S. Gems, and J. P. Hearn. 1985. Early implantation stages in the marmoset monkey (Callithrix jacchus). *American Journal of Anatomy* 172:265-278.

Ray, D. A., J. Xing, A.H. Salem and M. A. Batzer. 2006. SINEs of a nearly perfect character. *Systematic Biology* 55: 928-935.

Ray, D. A., J. Xing, D. J. Hedges, M. Hall, M. E. Laborde, B. A. Anders, B. R. White, N. Stoilova, J. D. Fowlkes, L. Chemnick, O. A. Ryder and M. A. Batzer. 2005. Alu insertion loci and platyrrhine primate phylogeny. *Molecular Phylogenetics and Evolution* 35:117-126.

Repanas, K., G. Fuentes, S.X. Cohen, A.M.J.J. Bonvin, and K. Perrakis. 2009. Insights into the DNA cleavage mechanism of human LINE-1 retrotransposon endonuclease. *Proteins* 74:917-928.

Rylands, A.B. 1997. The callitrichidae: a biological overview. In: Pryce C, Scott L, Schnell C, editors. *Marmosets and tamarins in biological and biomedical research: proceedings of a workshop*. Salisbury, United Kingdom: DSSD Imagery.

Salem, A.H., D.A. Ray, and M.A. Batzer. 2005. Identity by descent and DNA sequence variation of human SINE and LINE elements. *Cytogenetic and Genome Research* 108:63-72.

Schrago, C.G. and C.A.M. Russo. 2003. Timing the Origin of New World Monkeys. *Molecular Biology and Ecolution* 20: 1620- 1625.

Sen. S.K., J. Wang, J. Lee, H. Wang, P.A. Callinan, M.Dyer, R. Cordaux, P. Liang, and M.A. Batzer. 2006. Human genomic deletions mediated by recombination between Alu elements. *American Journal of Human Genetics* 79:41–53.

Smit, A.F., R. Hubley and P. Green. RepeatMasker Open-3.0. 1996-2010 [http://www.repeatmasker.org].

Solomon, N.G. and J.A. French. 1997. The study of mammalian cooperative breeding. In: *Cooperative Breeding in Mammals*(Ed. by N. G. Solomon & J. A. French). Cambridge: Cambridge University Press.

Steiper, M.E., and N.M. Young. 2006. Primate molecular divergence dates. *Molecular Phylogenetics and Evolution* 41:384-394.

Stevenson, M.F. and A.B. Rylands. 1988. The marmosets, genus *Callithrix*. In: *Ecology and behavior of neotropical primates, Volume 2*. (Ed. By R.A. Mittermeier, A.B. Rylands, A.F. Coimbra-Filho, and G.A.B. da Fonseca). Washington DC: World Wildlife Fund. p 131-222.

Sussman, R.W. 2000. Primate ecology and social structure. *New world monkeys*, Volume 2. Needham Heights, Massachusetts: Pearson Custom.

Sweeney, C.G., E. Curran, S.V. Westmoreland, K.G. Mansfield, and E.J. Vallender. 2012. Quantitative molecular assessment of chimerism across tissues in marmosets and tamarins. *BMC Genomics* 13:98.

Szak, S.T., O.K. Pickeral, W. Makalowski, M.S. Boguski, D. Landsman, and J.D. Boeke. 2002. Molecular archeology of L1 insertions in the human genome. *Genome Biology* 3: research0052-research0052.18.

Tardif, S.D. and D.A. Smucny. 2003. Reproduction of captive common marmosets (*Callithrix jacchus*). *Comparative Medicine* 53:364-368.

Thomson, J. A., J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall, and J. M. Jones. 1998. Embryonic stem cell lines derived from human blastocysts. *Science* 282:1145-1147.

Thomson, J. A., J. Kalishman, T. G. Golos, M. Durning, C. P. Harris, and J. P. Hearn. 1996. Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts. *Biology of Reproduction* 55:254-259.

Thomson, J.A. and V.S. Marshall. 1998. Primate embryonic stem cells. *Current topics in development biology* 38: 133-165.

Vincent, B.J., J.S. Meyers, H.J. Ho, G.E. Kilroy, J.A. Walker, W.S. Watkins, L.B. Jorde, and M.A. Batzer. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sties. *Molecular Biology and Evolution* 20:1338-1248.

Xing, J., D. J. Witherspoon, D. A. Ray, M. A. Batzer and L. B. Jorde**.** 2007. Mobile elements and primate evolution*. Yearbook of Physical Anthropology* 50: 2-19.