5-2014

Reconstructing New World monkey phylogeny: Alu elements in the common marmoset genome

Catherine Fontenot

Reconstructing New World monkey phylogeny: *Alu* elements in the common marmoset genome

by

Catherine Fontenot

Undergraduate Thesis under the direction of

Dr. Mark A. Batzer

Department of Biological Sciences

Submitted to the Honors College in partial fulfillment of

the Upper Division Honors Program

May 2014

Louisiana State University

& Agricultural and Mechanical College

Baton Rouge, Louisiana

ABSTRACT

New World monkeys (NWM, Platyrrhini) diverged from Catarrhini primates approximately 40 million years ago and have since successfully propagated in Central and South America. The branching order among the three NWM families and their 14-16 genera has long been studied and debated using various morphological and sequence analyses. It has been especially challenging to resolve the branching of the genera *Cebus*, *Saimiri*, and *Aotus*, likely due to rapid speciation. The completion of the common marmoset (*Callithrix jacchus*) draft genome allows for a more thorough investigation of NWM relationships. Short Interspersed Elements (SINEs), including primate-specific *Alu* elements, are retrotransposons that are particularly useful genetic markers because a) the absence of an insertion at a locus is known to be the ancestral state, and b) insertions are nearly-homoplasy-free phylogenetic characters. *Alu* elements identified in the *C. jacchus* genome can therefore be used to search for shared insertions (and hence shared common ancestry) among platyrrhines.

The common marmoset draft genome (CalJac3.2) was scanned for full-length *Alu* insertions using a local RepeatMasker installation with a custom library. These were then sorted by subfamily and by percent divergence from each respective consensus sequence. Each element, including flanking sequence, was queried against the human, chimpanzee, orangutan, and rhesus macaque genomes using BLAT. Orthologous sequences from all species were aligned using BioEdit, and primers were designed with Primer3. Polymerase Chain Reactions (PCRs) were performed on a panel of 21 individuals from 17 genera, including outgroups, and analyzed for presence/absence of an insertion by agarose gel electrophoresis. PCR amplicons displaying unexpected fragment sizes or contradictory

phylogenetic relationships were subjected to gel purification and chain termination DNA sequencing analysis.

Using the *C. jacchus* genome assembly, we reconstructed the evolutionary relationships within the NWM lineage using *Alu* elements as markers. Our phylogenetic analyses support the expected close relationship of common and pygmy marmosets. Moreover, we confirm the monophyly of Callithrichinae. We examined a range of percent divergences from several subfamilies to determine periods of retrotransposon activity, allowing us to focus our search on elements most likely to resolve the branch order within Family Cebidae.

BACKGROUND

New World Monkeys and Marmosets

New World monkeys (parvorder Platyrrhini, NWMs) are the group of arboreal anthropoid primates that inhabit regions in Central and South America ranging from southern Mexico to southern Brazil (Figure 1). There are 16 extant genera encompassing up to 75 species (Rosenberger and Hartwig 2001). It is estimated that the platyrrhines evolved separately from Catarrhini (Old World monkeys and apes) after an island hopping event from Africa to South America. Fossil evidence suggests that primates first arrived in South America 35-40 million years ago (Orlosky and Swindler 1975). A more recent study using mitochondrial and nuclear DNA, fossil evidence, and accepted branching patterns estimated that the last common ancestor of the platyrrhines existed closer to 20-30 million years ago (Perez et al. 2013). Despite the similar common naming style of Old World monkeys (subfamily Cercopithedoidea, OWMs) and NWMs, these two geographically-separated groups of higher primates (simians) are hypothesized to share only a prosimian, or non-monkey (i.e. lemurs, tarsiers), common ancestor. Many of their similarities are therefore contributed to parallel evolution from this non-monkey ancestor (Orlosky and Swindler 1975).

NWMs differ from OWMs in several morphological ways. First, NWMs have flat noses with side-facing nostrils (from which their name is derived), as compared to the narrow noses with downward facing nostrils of Old World primates (including humans). Second, NWMs have prehensile tails while OWMs and apes do not. Dental patterns are also different between the two groups. Platyrrhines have three premolar teeth in each row in contrast to two premolars in catarrhines (Rosenberger and Hartwig 2001).

Callitrichid primates (marmosets and tamarins, the morphologically smallest subfamily) are a unique group of NWMs. These species usually give birth to fraternal twins (Hill 1926). Twin fetuses develop within a single placenta and share blood supply through anastomoses (Wislocki 1939). This allows cells from both twins to settle in the bone marrow of each, causing hematopoietic chimerism (Benirschke et al. 1962). This definite chimerism is studied using XX and XY karyotes and has been observed in the cells from bone marrow, blood, lymph nodes, and spleen (Gengozian et al. 1964). Chimerism has also been observed in other tissues from the pancreas and small intestine. However, because of greatly varying levels of chimerism it is believed that this is a result of invasion of blood and lymphocytes, rather than true chimerism within those tissues (Sweeney et al. 2012).

The common marmoset *Callithrix jacchus* inhabits the rainforests of southeastern Brazil. Its body length ranges from 12-15 cm, and its tail length ranges from 29.5-35 cm. Morphologically, *C. jacchus* is distinguished by its white ear tufts and white forehead patch (Parker 1990) (Figure 2). This species has been extensively studied and is increasingly important to biomedical research. There is the potential for the *C. jacchus* to replace the Old World rhesus macaque (*Macaca mulatta*) as the main non-human primate research specimen in Europe and North America. Their small body size, short life cycle, and short reproduction time make them easy to maintain in captivity and inexpensive to experiment with (reviewed in Layne and Power 2003, Tardif et al. 2003, Abbott et al. 2003). Marmoset biology is important to the study of development, behavior, reproduction, neurobiology, immunology, infectious disease, endocrine signaling, obesity, aging, toxicology, and drug development (reviewed in Abbott et al. 2003, Ludlage and Mansfield 2003).

The whole-genome sequencing of *C. jacchus* has allowed the common marmoset to

be examined more closely and has been useful in the study of many other NWM species. It is important to the advancement of research of human disease, disease treatments, phylogeny, and comparative genomics.

Mobile Elements in Genomes, Phylogenetic Analyses

Mobile elements, or "jumping genes," are segments of DNA that can move around in genomes by either a cut-and-paste or a copy-and-paste mechanism. First discovered by Barbara McClintock in *Zea mays* (McClintock 1950), these DNA segments are present in many different organisms and contribute greatly to genome evolution and expansion (reviewed in Batzer and Deininger 2002, Kazazian 2004, Cordaux and Batzer 2009, Konkel et al. 2010). Mobile elements can be separated into two broad categories. Transposons are a type of mobile DNA that excise themselves out of one location and then insert themselves into another segment (Smit 1996). Retrotransposons, in contrast, are copied into an RNA intermediate before being inserted into a new locus by reverse transcription (Smit et al. 1995). These elements are duplicated within genomes, and this copy-and-paste process has played a great role in genome expansion.

It is estimated that the human genome consists of at least 42% retrotransposon sequences (Cordaux and Batzer 2009). Within primate genomes, *Alu* elements are one type of short interspersed element (SINE). These mobile elements are classified into subfamilies based on diagnostic mutations and the master gene theory (Shen et al. 1991; Deininger et al. 1992; Batzer et al. 1996; Han et al. 2005). A full-length *Alu* retrotransposon is approximately 300 base pairs (bp) long, and it is derived from 7SL RNA (Ullu and Tschundi 1984). The structure of an *Alu* is distinct (Figure 3). The left monomer and right monomer are joined by an A-rich region. The left monomer contains A and B boxes that code for an

RNA polymerase III promoter. The sequence of the monomers is GC-rich. An oligo(dA)-rich tail of varying length is usually present at the end of an *Alu* sequence. Target site duplications, a result of the reinsertion mechanism, are present on either end of the insertion (Batzer and Deininger 2002; Cordaux and Batzer 2009).

There are ~1.1 million *Alu* elements in the human genome, which make up more than 10% of DNA sequences (Lander et al. 2001). Current estimates support that a new *Alu* retrotransposition event happens one in every ~20 births (Cordaux et al. 2006; Xing et al. 2009), which is a much higher rate than previous estimates of one in approximately every 200 births (Deininger and Batzer 1999). *Alu* insertions play a great role in structural variation and mutations in the human genome including insertion mutations, recombination events, double strand breaks, deletion mutations, and mutations that affect cancer and other disease genes (Deininger and Batzer 1999; Batzer and Deininger 2002; Cordaux and Batzer 2009; Xing et al. 2009; Zhang et al. 2011).

Alu elements are non-autonomous sequences, meaning that they do not code for the proteins needed for their own mobilization. Instead, it is supported that they hijack the machinery of LINE-1 elements (Long Interspersed Element) (Dewannieux et al. 2003). L1s are autonomous and code for two proteins: one that binds to RNA (Kolosha and Martin 1997), and another that acts as both endonuclease (Feng et al. 1996) and reverse transcriptase (Mathias et al. 1991). Both of these proteins are essential in the process of target-primed reverse transcription (TPRT), by which both L1s and Alu elements move (Luan et al. 1993; Cost et al. 2002).

TPRT has the following mechanism. Because of the RNA polymerase III promoter, the *Alu* element is transcribed into an RNA intermediate strand. This strand continues into

the flanking sequence until the polymerase reaches a stop codon of four or more consecutive thymidines. The L1 endonuclease cuts the one strand of the target site, which is T-rich, allowing for hybridization of the tail of the retrotransposon RNA intermediate. Then, the mobile element is reverse transcribed by the L1 reverse transcriptase. After reverse transcription is complete, the other DNA strand is then also cut, and the newly synthesized DNA copy of the element is incorporated into the second strand. DNA synthesis follows to build the complementary strand according to the element template sequence. Target site duplications are found on both the 5' and 3' end of the insertion, and they indicate the original staggered cut site sequence.

This insertion mechanism is part of the reason that *Alu* elements, among other SINEs, are useful in phylogenetic and other evolutionary studies. Norihiro Okada (Nikaido et al. 1999) first used SINEs to determine the family relationships of cetartiodactyls. It is thought that the insertion of a SINE at a locus is an essentially irreversible event (Nikaido et al. 1999), that shared insertions are inherited from a common ancestor (Batzer and Deininger 1991; Batzer and Deininger 2002), and that the absence of an insertion at a locus is the ancestral state (Batzer et al. 1994). These characteristics render *Alu* elements nearly-homoplasy-free because precise parallel insertions and deletions are rare (Ray et al. 2006). Additionally, when the insertion of a DNA segment seemingly indicates a shared insertion, despite the fact that it is actually the result of convergent evolution or some other sequence of DNA, the contradiction is easily clarified using PCR and DNA sequencing. An examination of insertions and insertion sites may reveal a near-parallel independent insertion or the insertion of a different *Alu* element at the same locus as another element. Mobile element insertions, particular *Alu*, have proven useful in the phylogeny and population studies of

many primate taxa (reviewed in Konkel et al. 2010) including the human-chimpanzee-gorilla trichotomy (Salem et al. 2003), orangutans (Locke et al. 2011), gibbons (Meyer et al. 2012), and NWMs (Ray et al. 2005).

Despite the utilization of various research techniques, the complete phylogeny of NWMs has proven difficult to determine across the 16 extant genera (reviewed in Schneider and Sampaio 2013). Over time, several studies have attempted to resolve the branching of the genera and their further groupings into higher clades. Research has also focused on the relationships among the higher groupings. In three molecular studies from the same researchers, results supported both the grouping of NWMs into two families, Cebidae and Atelidae (Schneider et al. 1993, 1996), and the grouping of NWMs into three families, Cebidae, Pitheciidae, and Atelidae (Harada et al. 1995), though the branching of these families remained unresolved. Further studies agreed that a three-family grouping of NWMs is the most likely (Ray et al. 2005; Opazo et al. 2006; Wildman et al. 2009; Osterholz et al. 2009; Aristide et al. 2013) and supported the sister grouping of families Cebidae and Atelidae to the exclusion of Pitheciidae (Steiper and Ruvolo 2003; Ray et al. 2005; Wildman et al. 2009; Osterholz et al. 2009; Aristide et al. 2013). The monophyly of all platyrrhines is also supported (Singer et al. 2003; Ray et al. 2005; Osterholz et al. 2009).

Even with these advancements in the phylogenetic research of NWMs, not all questions have been answered. The genera *Cebus*, *Aotus*, and *Saimiri* have posed a particular challenge. Different kinds of morphological and molecular analyses have yielded different results. Recently, studies have supported a *Cebus*-*Saimiri* clade that is a sister taxon to *Aotus*, though these results are not definitive (Opazo et al. 2006; Osterholz et al. 2009). Because of the certainty of results utilizing *Alu* elements as phylogenetic markers,

our research focused on confirming the order of the NWM families and resolving the long-standing questions in the NWM genus-level branching.

Figure 1: Habitats of NWMs and Common Marmosets
New World monkeys inhabit Central and South American Ranging from southern Mexico to northern Brazil. Common marmoset habitats are concentrated in the Atlantic coastal forests of Brazil.

Figure 2: **The *common marmoset,* Callithrix jacchus**
The common marmoset is a small NWM with a body length ranging from 12 to 15 cm. It is distinguishable by its white ear tufts and white forehead patch. (Image courtesy of the Southwest Foundation for Biomedical Research and the National Human Genome Research Institute)

Figure 3:* Alu *Element Structure
The DNA sequence of an *Alu* element is characterized by several different segments. The target site duplications (TSDs) are created at the cut site when the element is inserted into a new locus in the genome. The beginning of the element is GC-rich and usually contains the exact sequence shown above. The A and B boxes are promoter sites for RNA polymerase III. The center of the *Alu* element is A-rich, and the end is characterized by a poly-A tail.

MATERIALS AND METHODS

Sequence Identification and Primer Design

To identify marmoset-specific *Alu* elements, we downloaded the genome (calJac3.2) from the Baylor College of Medicine Genome Sequencing Center and searched it for *Alu* elements using a local RepeatMasker installation (Smit et al. 1996) with a custom library containing *Alu* subfamilies specific to the lineage leading to the common marmoset. After the identification of the candidate loci, we used the University of California Santa Cruz's (USCS) BLAT (Kent 2002) – locally installed – to extract an additional 500 bp of flanking sequence on each side of the elements. For our phylogenetic analyses we selected *Alu* elements from a range of subfamilies. We included *Alu* elements from subfamilies with evidence of recent mobilization (e.g. sf4) as well as older *Alu* subfamilies (e.g. sf39, sf63). Homologous sequences of each locus from human (hg19), chimpanzee (panTro2), orangutan (ponAbe2), and rhesus macaque (rheMac2) were downloaded from the UCSC BLAT genome browser (Kent 2002) and were aligned with the marmoset sequences using BioEdit (Hall 1999).

Forward and reverse oligonucleotide primers were designed within conserved, homologous regions of the flanking sequences of five primate species mentioned above

using Primer3 (Rozen & Skaletsky 1999). Specifications for the primer selection included a length of 20-27 bp, an annealing temperature ranging from 57-61 degrees Celsius, and the exclusion of four or more mononucleotide repeats. Online primer tests were conducted with the UCSC *in-silico* polymerase chain reaction (PCR) (Kent 2002). Previous to wet bench experiments, this confirmed that the primers yielded one PCR product in each genome and also provided the estimated base pair lengths of sites where the *Alu* insertions were present in marmoset and absent from other species' genomes. If the *in-silico* PCR yielded two PCR products, the base pair length of each was recorded, and these exceptions were noted for their subsequent appearance in wet bench analyses. If more than two PCR products resulted, the primers were redesigned. Primer sequences were additionally tested for specificity using BLAT. Primers were ordered from Sigma-Aldrich and were re-suspended and diluted using TLE [10mM Tris/0.1mM EDTA (ethylenediaminetetraacetic acid)] buffer. Each primer pair was tested by wet-bench PCR for binding accuracy and amplification using HeLa DNA (From Human Cell Line ATCC HeLa-CCL-2). If primers did not amplify during the PCR primer test, a temperature gradient PCR was conducted to test for optimal annealing temperatures. This again was conducted with HeLa DNA, and the temperatures tested ranged from 50-60°C. If primer failed to amplify during the temperature gradients, they were re-designed.

Polymerase Chain Reaction and Sequencing

Once primers were successfully tested, PCRs were conducted on a phylogenetic panel that included genomic DNA from human, chimpanzee, African green monkey, owl monkey, Bolivian red howler monkey, white-bellied spider monkey, woolly monkey, pygmy marmoset, common marmoset (three individuals), Geoffroy's saddle-back tamarin,

capuchin monkey (three individuals), squirrel monkey, Bolivian gray titi, northern white-faced saki, tarsier, and ring-tailed lemur (Table 1). Some loci were additionally tested in Goeldi's marmoset once they were identified as present in common and pygmy marmosets (these additional panels not included in figures). The panels also included TLE (10mM Tris/ 0.1mM EDTA) as a negative control.

Each 25-µl reaction contained 15-25 ng of DNA, 200 nM of forward and reverse oligonucleotide primers, 1X cresol red loading dye (34% sucrose, 0.05% cresol red), 1.5-2 mM MgCl2, 1x PCR buffer (50 mM KCl and 10 mM TrisHCL, pH 8.3), 0.2 mM dNTPs, 1-2 units of Taq polymerase, and was filled to volume with water. The reactions were carried out in 96-well Optical Reaction Plates from Applied Biosystems. BIO-RAD iCycler and Applied Biosystems GeneAmp® PCR System 9700 thermocyclers were used with the following PCR protocol: 94°C for 60 s (initial denaturation); 32 cycles of 94°C for 30 s (denaturation), annealing temperature of the primers for 30 s, 72°C for 30 s (extension); 72°C for 2 min. (final extension and termination); 25°C hold, if necessary. PCR products and a 100-bp DNA ladder were then run on a 2% agarose gel [98% TBE (Tris base, boric acid, EDTA)]; also containing 0.1 µg/mL ethidium bromide] using BIO-RAD Power Pac 3000 electrophoresis machines for 40 to 60 min. DNA fragments were visualized using a BIO-RAD UV camera.

Gel Purification, Cloning, and DNA Sequencing

If any unusual or unexpected banding patterns were observed, the DNA amplicons in question were sequenced. This method allows for the exact DNA sequences of amplicons to be examined, which is a useful technique in clarifying seemingly confounding events. DNA segments were first cut out of the agarose gel using a straight razor blade. The gel was

then digested using a Promega Wizard SV Gel and PCR Clean-up System according to the manufacturer's directions, with the following amendments. DNA was eluted twice with 50 µl of nuclease-free water then dehydrated with a Speed Vac for 60 min. Then, the DNA was resuspended in a final volume of 10-25 µl of nuclease free water. This approach was an attempt to elute as much DNA as possible, as well as create a more highly-concentrated DNA solution for use in sequencing.

Sequencing reactions were conducted in 96-well plates in 10-µl volumes that included 4 µl of concentrated DNA solution described above, Applied Biosystems® BigDye® Terminator v3.1 Cycle Sequencing solution, 320 nM either forward or reverse primer, 2.5X sequencing buffer, and water. BIO-RAD iCycler and Applied Biosystems GeneAmp® PCR System 9700 thermocyclers were used with the following PCR sequencing protocol: 95°C initial denaturation for 2 min.; 25 cycles (40 cycles used for difficult sequences, when maximum amplification was needed) of 95°C for 10 s, 50°C for 5 s, 60°C for 4 min.; hold at 4°C.

Cycle sequencing reactions were then precipitated by adding 5 µl of 70 mM EDTA and 30 µl of 100% ethanol (EtOH) to each sample. The plate was shaken and then left to incubate at room temperature for 15 min. After incubation, the plate was spun in a centrifuge at 2500 RCF for 15 to 30 min. The liquid was then removed by inverting the plate and spinning at 150 RCF for <1 min. Then 30 µl of 70% EtOH was added to each reaction, and the plate was again centrifuged at 2500 RCF for 5 to 10 min. Again, the ethanol was removed by inverting the plate and spinning as described previously. After the traces of EtOH were removed by evaporation, the sequencing products were resuspended in 15 µl of ABI Hi-Di formamide. The reactions were then analyzed using Applied

Biosystems 3130XL Genetic Analyzers, located at the LSU College of Science Genomics Facility. Output was visualized and examined using BioEdit, and alignments with consensus sequences and homologous sequences of various species were used for comparisons, when necessary.

For certain loci where accurate amplification and sequencing was particularly difficult, DNA amplicons were cloned into vectors using the Invitrogen TOPO TA cloning kit according to the provided protocol. Cells were spread onto lysogeny broth (LB) + ampicillin agar plates and incubated overnight at 37°C. A minimum of two colonies from each locus were randomly selected for colony PCR, with the following protocol: 94°C for 5 min.; 25 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 90 s; 72°C for 10 min., and a hold at 25°C if necessary. Each 25-µl reaction contained the same ingredients as previously stated with the exception of the M13 primers provided in the cloning kit (annealing temperature 55°C) and the DNA that was added from picking the colonies (<2 µl). Colony PCR was used as a test to determine whether or not colonies contained the insertion of interest. The positive colonies were picked and grown overnight in an LB + ampicillin broth at 37°C. The DNA was extracted from these clones using a 5PRIME FastPlasmid Mini Kit, following the manufacturer's protocol. These cloned loci were then subjected to the sequencing protocol described above.

Phylogenetic Analysis

Loci that successfully amplified on the phylogenetic panel and those that were sequenced were included in the phylogenetic analysis. Each locus was scored according to the following rubric: a "1" was assigned to an individual on the panel with an *Alu* insertion, a "0" was assigned to an individual with no *Alu* insertion, and a "?" was assigned to an

individual with no amplification or otherwise unknown results. These scores were put into the Mesquite program (http://mesquiteproject.org/) and were subjected to a Dollo parsimony analysis (Felsenstein 1979) (Dollo's law of irreversibility—it is difficult if not impossible to lose an evolved/inherited trait, such as an *Alu* insertion at a specific site). Additionally, the insertion patterns were analyzed with 10,000 bootstrap replicates using PAUP* 4.0b10 (Swofford 2003). Acceptable branches included node support of greater than 50% from bootstrap replications, and the presence of three of more *Alu* insertions was considered statistically significant support at each internal node (Waddell, Kishino et al. 2001).

	Species Name	Common Name	Origin*	ID
1	*Homo sapiens*	Human	ATCC	HeLa-CCL-2
2	*Pan troglodytes*	Chimpanzee	IPBIR	NS06006
3	*Chlorocebus aethiops*	African green Monkey	ATCC	CCL70
4	*Aotus trivirgatus*	Owl Monkey	ATCC	CRL1556
5	*Alouatta sara*	Bolivian Red Howler Monkey	SDFZ	OR749
6	*Ateles belzebuth*	White-Bellied Spider Monkey	SDFZ	KB7601
7	*Lagothrix lagotricha*	Woolly Monkey	Coriell	NG05356
8	*Callithrix pygmaea*	Pygmy Marmoset	SDFZ	OR690
9	*Callithrix jacchus*	Common Marmoset	NEPRC	1-2008
10	*Callithrix jacchus*	Common Marmoset	SWNPRC	272
11	*Callithrix jacchus*	Common Marmoset	WNPRC	CJ1590
12	*Saguinus fuscicollis nigrifons*	Geoffroy's Saddle-Back Tamarin	SDFZ	OR621
13	*Cebus apella*	Capuchin Monkey	KP	CA003
14	*Cebus apella*	Capuchin Monkey	KP	30156
15	*Cebus apella*	Capuchin Monkey	KP	30157
16	*Saimiri s. sciureus*	Squirrel Monkey	SDFZ	KB4544
17	*Callicebus donacophilus*	Bolivian Gray Titi	SDFZ	OR1522
18	*Pithecia p. pithecia*	Northern White-Faced Saki	SDFZ	OR842
19	*Tarsius syrichta*	Tarsier	SDFZ	OR2831
20	*Lemur catta*	Ringtailed Lemur	Coriell	NG07099
21	*Callimico goeldii*	Goeldi's Marmoset		

Table 1: Species Included in Phylogenetic Panel

A total of 21 species from 16 different genera were utilized in phylogenetic panels. Each individual is listed above with the species name, common name, location of origin, and identification code. *C. goeldii* was utilized only for select loci and is, therefore, not included in all figures within this text.

**NEPRC (New England Primate Research Center), WNPRC (Wisconsin National Primate Research Center), SWNPRC (Southwest National Primate Research Center), KP (Kimberly Phillips, Trinity University), ATCC (American Tissue Culture Center), IPBIR (Integrated Primate Biomaterials and Information Resource), Coriell (Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ), SDFZ (San Diego Frozen Zoo), CRES (Conservation and Research for Endangered Species)*

RESULTS AND DISCUSSION

PCR and Sequencing Analyses

The first step in utilizing *Alu* element insertions in phylogenetic analyses is examining PCR amplification results across a phylogenetic panel. The 21-member species panel was subjected to PCR and then visually analyzed for presence and absence patterns. On an agarose gel, the high bands (longer amplicons) include *Alu* insertions at the locus of

interest, and the low bands (shorter amplicons) lack the insertion. The four most commonly observed results were *Alu* insertions shared among the three common marmoset samples or "common-marmoset-specific insertions," insertions among all marmosets [common, pygmy, and Goeldi's (not included in figures because analyzed separately)], insertions in subfamily Callitrichinae, and insertions in all NWMs (Figure 4).

Some banding patterns gave conflicting results that supported a shared insertion in species that are thought not to be closely related. These loci were subjected to chain termination sequencing, and the DNA sequences were examined to determine the insertion sequence and the insertion site. Often, the sequencing revealed near-parallel independent insertions (examples in Figure 5), which occurs when a mobile element or other piece of DNA is inserted near the same insertion site as in another species. Sequencing proved that these apparent shared insertions were actually different elements that inserted at a nearby, but not exact, insertion site to that of the marmoset *Alu*. A list of sequenced loci and their results are in Table 2. In our data set, 23 of the 231 loci were lineage-specific near-parallel independent *Alu* insertions that were confirmed using chain termination sequencing. This is an equivalent rate of ~0.10 events/insertions. In a previous study using *Alu* elements in NWM phylogeny, the rate among their data was ~0.6 events/insertion (Ray et al. 2005). One study analyzed several thousand *Alu* insertions compiled from several different publications and found an overall of near-parallel independent *Alu* insertions to be ~.0004 (Ray et al. 2006). Our higher rate is most probably the result of a relatively small sample size, rapid radiation of the species in question, varying rates of retrotransposition among organisms, and the researchers' selection of confirmed loci that would be informative for a phylogenetic analysis.

Phylogenetic Analysis

In total, 231 informative loci were included in our tree-building phylogenetic analysis. The following loci unambiguously supported branching and genus separations. Fifty *Alu* insertions were common marmoset specific. Twenty-five supported the sister relationship between common and pygmy marmosets. There were 52 *Alu* elements found to group all marmosets (common, pygmy, and Goeldi's). Forty-nine insertions grouped together Subfamily Callithrichinae. Four *Alu* elements supported the monophyly of Family Cebidae. A single *Alu* insertion supported the sister relationship between Family Cebidae and Family Atelidae. There were 19 elements that supported the monophyly of NWMs, to the exclusion of OWMs, apes, and prosimians, our outgroups.

The resulting phylogenetic tree (Figure 6) had a Consistency Index (CI) of 0.9914. A CI of 1 indicates that no aspect of the tree is conflicting and that the data is homoplasy free. Our high value indicates strong support for the calculated branching pattern. Alternatively, the Homoplasy Index (HI) was 0.0086. Each node is also supported by bootstrap values. This is a percentage that represents the number of times out of 10,000 replicates that a branching pattern resulted from the data. As displayed in Figure 6, most branches had support of 99% or 100%. The nodes of Family Cebidae and the sister grouping of Family Cebidae with Family Atelidae had bootstrap values of under 70%, which are not statistically significant on their own. Further support for the tree was given by the Waddell likelihood test (based on Dollo parsimony). Most nodes had p-values of <0.001, indicating that the number of informative loci left a 0.1% chance of error. The node indicating the monophyly of Familly Cebidae had a p-value of <0.05. The node linking Cebidae and Atelidae was not strongly supported by Dollo parsimony.

Our data supports the findings of other studies that Family Cebidae is a sister taxon to Family Atelidae, to the exclusion of Family Pitheciidae (Steiper and Ruvolo 2003; Ray et al. 2005; Wildman et al. 2009; Osterholz et al. 2009; Aristide et al. 2013). Additionally, we strongly support the inclusion of *Callithrix*, *Callimico*, and *Saguinus* in Subfamily Callithrichinae. Within this subfamily, the sister relationship between pygmy marmoset and common marmoset was supported to the exclusion of other species. Our data also indicated that the sister taxon to *Callithrix* is *Callimico*.

The monophyly of Family Cebidae demonstrated the value in using multiple statistical analyses to evaluate data. According to the bootstrap replicates, this node only had 69% support, which does not indicate definitive results. However, there were four *Alu* insertions that supported this node, and Waddell likelihood requires three *Alu* insertions to statistically support a node. Therefore, the p-value for this sister grouping was <0.05, and our results are, in fact, statistically significant and do lend strong support for this node.

The 50 common-marmoset-specific *Alu* insertions will be useful in further studies of *C. jacchus* (i.e. population studies). They and other species-specific *Alu* loci can also be used in PCR analyses to identify unknown organisms. As stated previously, 24 species-specific *Alu* insertions were identified by Sanger sequencing after apparent conflicting results from PCR.

Alu *Subfamily Divergences*

The loci utilized in the phylogenetic tree were also selected based on their percent divergences from the subfamily consensus sequences (Figure 7). A low percent divergence indicates a younger element, as it has had a relatively short time to mutate. A higher percent divergence indicates an older element that has mutated for a longer period of

evolutionary time. *Alu*sf4 and *Alu*sf63 were the subfamilies with the most individual loci within our dataset. *Alu*sf4 elements have percent divergences ranging from 0.3% to 9.7%. Our study utilized insertions with divergences ranging from 0.3% to 6.9%. Most *Alu*sf4 insertions, despite their percent divergence, were shared between common and pygmy marmosets. There were 31 common-marmoset-specific insertions of *Alu*sf4 with divergences ranging from 0.3% to 4.2%. Twenty-four *Alu*sf4 insertions (0.3-7.0%) were found to be present in both common and pygmy marmosets. Three *Alu*sf4 insertions (3.1-6.7%) were specific to marmosets and capuchin monkeys. One element, respectively, was found to be specific to Subfamily Callithrichinae (6.2% divergence) and NWMs (6.9% divergence).

*Alu*sf63 is an older subfamily with divergences ranging from 0.7% to 10.6% (4.0-10.1% present in analysis). This subfamily was present in all species in Subfamily Callithrichinae (7 insertions, 5.5-10.1%). Three *Alu*sf63 insertions were present only in marmosets and *Callimico* (5.5-7.5%). Three *Alu*sf63 insertions were specific to marmosets (4.0-8.0%). Only one (9.7%) was common marmoset specific, and one (8.5%) was present in all NWMs.

We further grouped loci from several other subfamilies into Low Divergence (0.0-3.0%), Mid Divergence (3.1-6.0%), and High Divergence (6.1-10.0%) sets. None of our Low Divergence loci were present in all NWMs. Instead, they were concentrated within Subfamily Callithrichinae. Twenty three (0.0-2.8%) were common-marmoset specific. Thirty-nine (0.3-2.8%) were present in common and pygmy marmosets. One Low Divergence insertion (2.8%) was present in marmosets and capuchin monkey. Nineteen insertions (0.7-2.9%) were present in all members of Callithrichinae. Mid Divergence loci

showed more variation. Ten *Alu* insertions ranging from 4.4% to 6.0% divergent were identified in all NWMs. Seven (3.1-5.1%) were present in Subfamily Callithrichinae. One insertion (3.5%) was specific to marmosets and *Callimico.* There were 11 Mid Divergence loci (3.1-4.8%) shared by common and pygmy marmosets, and one (3.1%) was specific to common marmosets. The three High Divergence insertions (6.7-7.0%) were only found to be NWM specific.

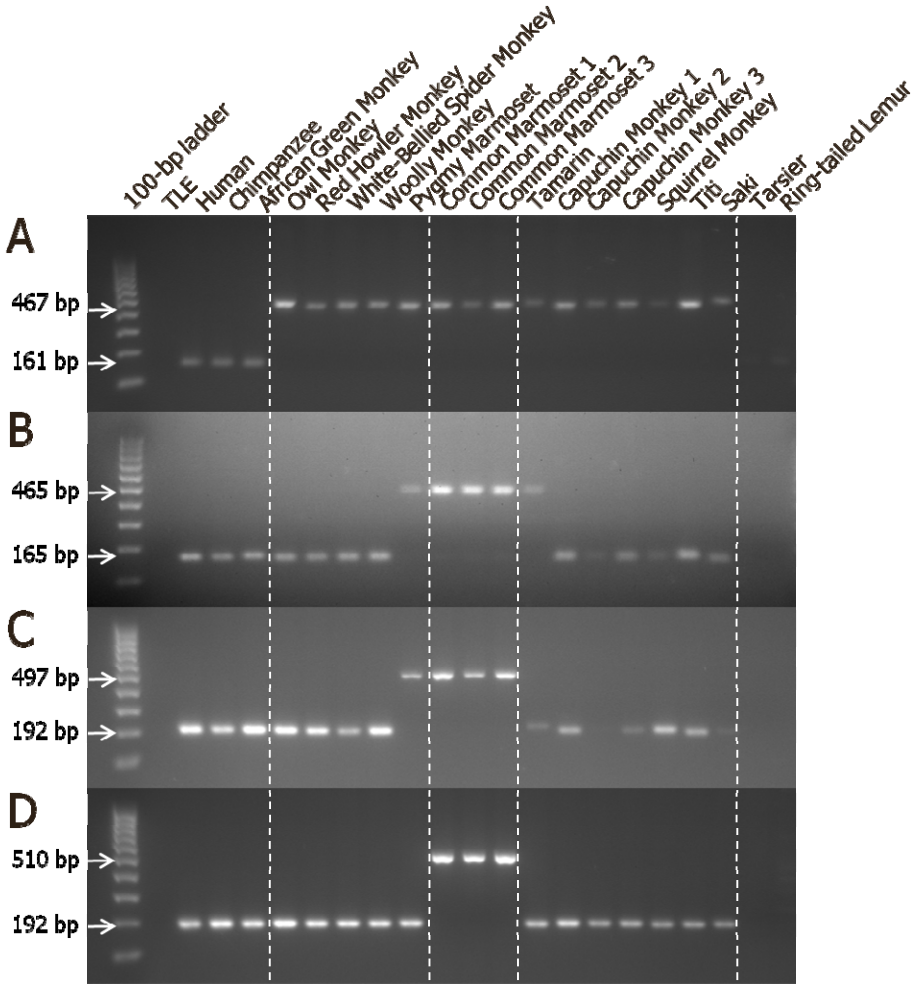

Figure 4: Common Insertions

The pictures (A-D) show the 21-sample panel, including a negative control (TLE) and 5 outgroups. Individuals are labeled across the top. In these pictures, a high band is a DNA amplicon that includes an *Alu* element. The lower bands indicate the same fragment of DNA without the *Alu* insertion. The four most common insertions we identified with PCR analysis were A) NWM specific, B) Subfamily Callithrichinae specific (shared between marmosets and tamarin), C) marmoset specific [shared among pygmy, common, and Goeldi's (not pictured) marmosets], and D) common marmoset specific.

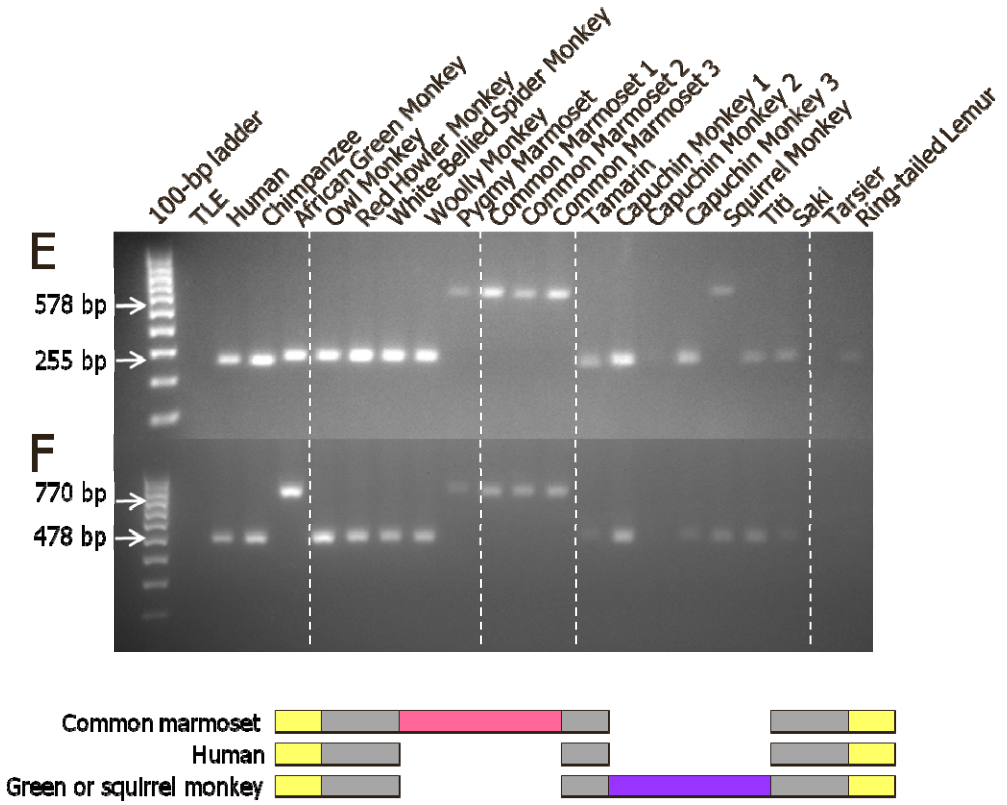

Figure 5: Examples of Near-Parallel Independent Insertions

The gel pictures above are examples of PCR results that required further Sanger sequencing analysis. The bands gave conflicting results because they indicate a shared insertion among species that are not considered closely related. Sequencing results showed lineage-specific, near-parallel independent insertions of *Alu* elements in squirrel monkey (E, Locus5), and African green monkey (F, Locus75). The graphic below the gel pictures illustrates a near parallel-independent insertion. The top strand represents the common marmoset genome, the middle strand human, and the bottom strand squirrel monkey or African green monkey. Primers within the flanking sequence are shown as yellow boxes. The gray boxes are flanking sequences that are shared among all species. The pink box represents the *Alu* insertion in the marmoset, and the purple box shows the relative location of the *Alu* insertion in the other species. During a PCR reaction, the *Alu* insertions that fall between the primers are copied and appear similar on the gel image.

	Locus	Species	Sequencing Results*
1	Ta7_2	Squirrel Monkey	NPI insertion, FL *Alu*
2	Ta7_15_sf74_8a	Several**	Shared truncated *Alu*, marmoset-specific FL *Alu*
3	Ta7_15	Northern White-Faced Saki	NPI insertion, FL *Alu*
4	RMsf39_1	African Green Monkey	NPI insertion, FL *Alu*
5	RMsf39_1	Capuchin Monkey	NPI insertion, truncated *Alu*
6	RMsf39_5	Owl Monkey	NPI insertion, FL *Alu*
7	RMsf39_18	Owl Monkey	NPI insertion, FL *Alu*
8	James_RMsf62_7	African Green Monkey	NPI insertion, FL *Alu*
9	RM4_248	Owl Monkey	NPI insertion, FL *Alu*
10	RMsf47_1	Capuchin Monkey	NPI insertion, FL *Alu*
11	sf60_7	Bolivian Red Howler Monkey	NPI insertion, FL *Alu*
12	sf60_9	Squirrel Monkey	NPI insertion, FL *Alu*
13	sf74_7	Capuchin Monkey	NPI insertion, FL *Alu*
14	sf74_10_Ta7	Owl Monkey	NPI insertion, FL *Alu*
15	sf10_3	Capuchin Monkey	NPI insertion, FL *Alu*
16	sf10_12_sf39_10	Bolivian Red Howler Monkey	NPI insertion, FL *Alu*
17	sf69_10	Squirrel Monkey	NPI insertion, FL *Alu*
18	sf69_15_sf10	Capuchin Monkey	NPI insertion, FL *Alu*
19	sf69_15_sf10	Squirrel Monkey	NPI insertion, FL *Alu*
20	sf_4_26	Owl Monkey	NPI insertion, FL *Alu*
21	sf43_6	Bolivian Gray Titi	NPI insertion, truncated *Alu*
22	Locus5	Squirrel Monkey	NPI insertion, FL *Alu*
23	Locus72	Squirrel Monkey	NPI insertion, FL *Alu*
24	Locus75	African Green Monkey	NPI insertion, FL *Alu*

Table 2: Sequenced Loci and Their Results

Twenty-four (24) of the 231 *Alu* insertions used for the phylogenetic analysis were derived from sequencing results of confounding PCR results. While the majority were near-parallel independent insertions of full-length *Alu* elements (NPI, FL*), some were truncated. Locus Ta_7_15_sf74_8a yielded interesting results.** A truncated *Alu* was shared by marmosets, tamarin, squirrel monkey, and capuchin monkey. Additionally at this locus, a marmoset-specific *Alu* was inserted into the truncated *Alu.*

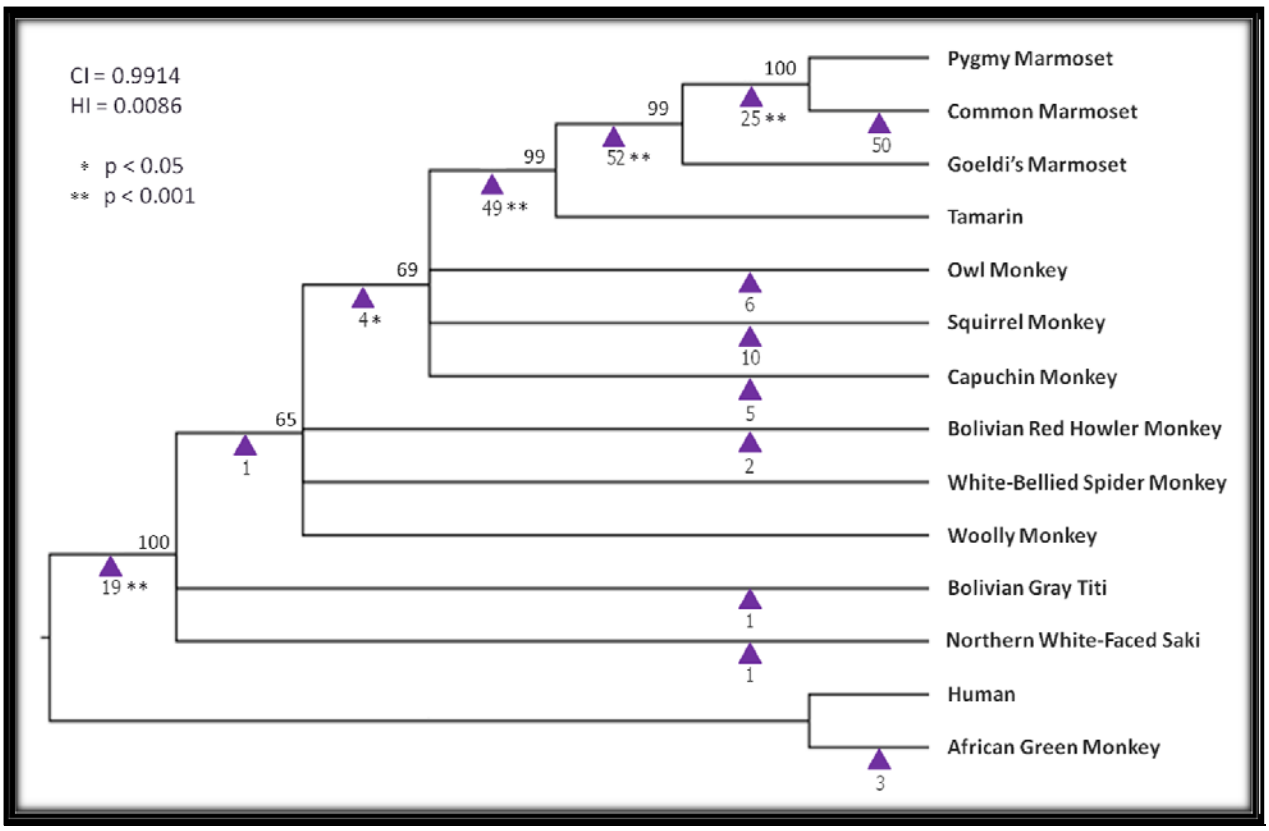

Figure 6: Cladogram of Panel Species

Two hundred thirty-one informative *Alu* insertions were used to construct a phylogenetic tree with a Consistency Index (CI) of 0.9914 and a Homoplasy Index (HI) of 0.0086 using the PAUP* software. Bootstrap percentages are displayed above each node. Numbers below the purple arrows indicate the number of non-homoplasious *Alu* element insertions along each branch. Asterisks indicate statistical support for a node from Dollo parsimony analysis.

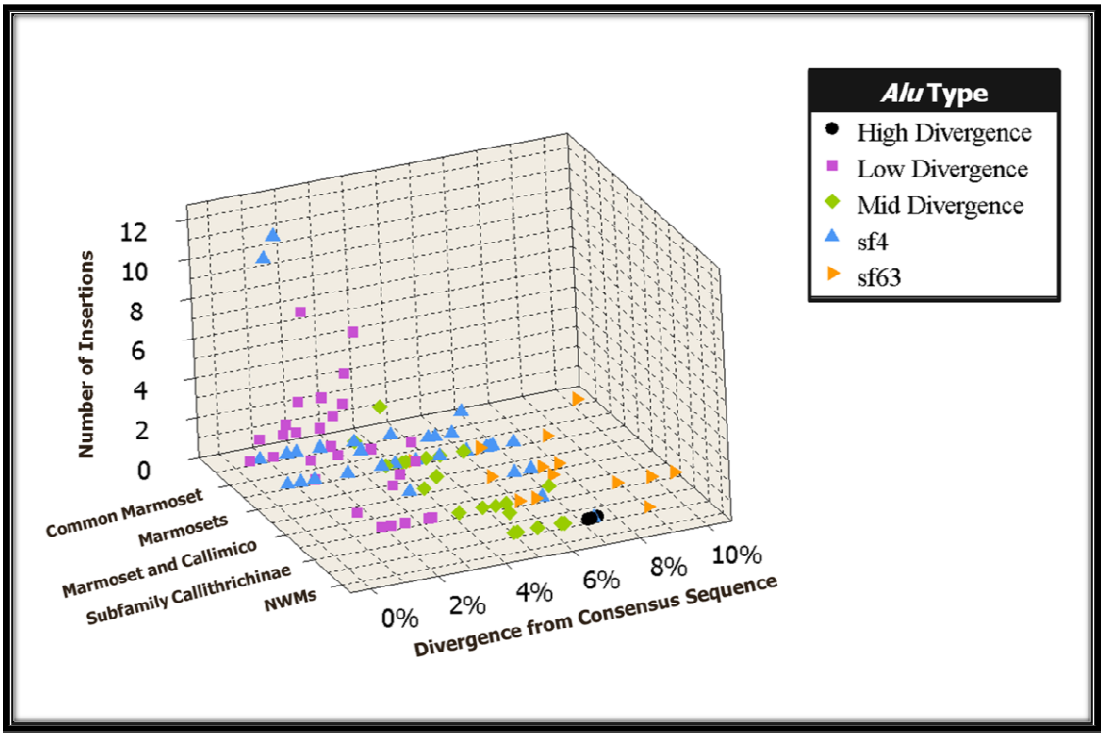

Figure 7:* Alu *Insertions by Percent Divergence
This plot shows the distribution of *Alu* insertions across different species groups and percent divergence from respective consensus sequences. The two triangle symbols represent the two subfamilies that we examined most closely. The other symbols represent groups of several subfamilies separated by percent divergences: low (0.0-3.0%), medium (mid; 3.1-6.0%), and high (6.1-10.0%). The different percent divergences were selected to determine time periods of retrotransposon activity with respect to the branching order of NWMs. Most *Alu*sf4 insertions, across a wide range of divergence, are only shared in common and pygmy marmosets.

CONCLUSION

Our analysis of over 200 *Alu* insertions is one of the largest studies of its kind within NWMs (Ray et al. 2005, Osterholz et al. 2009). Our results further support the monophyly of NWMs, Callithrichinae, and Cebidae. We also identified several species-specific *Alu* insertions that will be useful in species identification. Though our main goal at the start of this project was to elucidate the branching order of *Cebus*, *Saimiri*, and *Aotus*, we were unable to find data to definitively group two as sister taxa to each other while excluding the

third. We examined *Alu* elements with a range of percent divergences from their respective subfamilies as a rough estimate of when the *Alu* insertions may have occurred, allowing us to focus our search on elements most likely to resolve the branching order within Cebidae. The lack of insertions that resolve the branching among *Aotus*, *Cebus*, and *Saimiri* is most likely due to a rapid speciation event, where mobile elements did not have the generational time needed to become fixed in different genomes, therefore identifying sister taxa with identity-by-descent. Though there were no conflicting *Alu* insertions of this type among our data, examples can be found in previous studies (Osterholz et al. 2009, Ray et al. 2005).

Future work will be needed to continue this analysis of the *Callithrix jacchus* genome and NWMs. It is approximated that the *Cebus-Saimiri-Aotus* divergence began between 15 and 20 million years ago (Schrago et al. 2013). It would be useful to further examine the percent divergences of various subfamilies and utilize different ranges in PCR analyses on a phylogenetic panel, searching for windows of time where retrotransposition was concurrent with the speciation of *Cebus, Aotus,* and *Saimiri.* However, without full-genome sequences of *Cebus, Aotus,* and *Saimiri*, it will be difficult to thoroughly compare the *Alu* insertions present in each genus.

Further work in genome sequencing of NWMs is needed to further scientific knowledge, to resolve long-standing phylogenetic questions, and to aid in future research, particularly if other NWM species become important model organisms. The common marmoset is becoming increasingly useful in biomedical research, and a full understanding of the genome will only aid in the understanding of other biological functions and processes. The search for *Alu* elements that resolve the trichotomy within Family Cebidae and those that aid in population studies will continue.

ACKNOWLEDGEMENTS

First, I would like to thank my thesis director, Dr. Mark Batzer, for his continued support throughout the last five years. I think we both distinctly remember the first day my Basilian accent and I walked into his office, and I hope I have grown since then. I would also like to thank Jerilyn Walker, Dr. Granger Babcock, and Cindy Seghers for their years of advice, teaching, and encouragement. More thanks goes out to every member of the Batzer Laboratory, especially Megan Ranck, my long-time teammate; Josh Meyer, who never hesitates to assist; and Erika Arceneaux, Michelle Limbaugh, Will Gensler, Glenn Carman, Vallmer Jordan, and Emmy Bennett, my coworkers and friends. I will forever be indebted to Dr. Miriam Konkel for her listening, her understanding, her compassion, and her mentorship, in addition to her guidance and patience throughout this project and others. I thank Dr. Batzer, Dr. Babcock, and Dr. Konkel for serving on my thesis committee.

I would also like to take this opportunity to thank my family for their love and support: my mom, whose unconditional love has never wavered; my dad, who taught me to appreciate and deliver hard work and a job well-done; my brothers Lindsey and Robert, who set the example and toughened me up; my sister Jennifer, who will always be fiercer than I; and my sister Thea, the baby who takes more care of me than I do of her. I would not have made it this far without them.

I thank the many research centers who contributed DNA samples. The common marmoset genome project is funded by the National Human Genome Research Institute (NHGRI) and by the National Institutes of Health (NIH). This research was supported by National Institutes of Health R01 GM59290 (MAB). Parts of this project were also supported by the Howard Hughes Medical Institute's summer research funding (CCF).

Works Cited

Abbott, D. H., D. K. Barnett, et al. (2003). "Aspects of common marmoset basic biology and life history important for biomedical research." Comp Med 53(4): 339-50.

Aristide, L., A. L. Rosenberger, et al. (2013). "Modeling lineage and phenotypic diversification in the New World monkey (Platyrrhini, Primates) radiation." Mol Phylogenet Evol. In Press.

Batzer, M. A. and P. L. Deininger (1991). "A human-specific subfamily of *Alu* sequences." *Genomics* 9:481-7.

---. (2002). "Alu repeats and human genomic diversity." Nat Rev Genet 3(5): 370-9.

Batzer, M., P. L. Deininger, et al. (1996). "Standardized nomenclature for Alu repeats." J Mol Evol 42(1): 3-6.

Batzer, M. A., M. Stoneking, et al. (1994). "African origin of human-specific polymorphic Alu insertions." P Natl Acad Sci USA 91(25): 12288-192.

Benirschke, K., J. M. Anderson, et al. (1962). "Marrow Chimerism in Marmosets." Science 138(3539): 513-5.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet 10(10): 691-703.

Cordaux R., D. J. Hedges, et al. (2006) "Estimating the retrotransposition rate of human Alu elements." Gene 373: 134-7.

Cost, G. J., Q. Feng, et al. (2002). "Human L1 element target-primed reverse transcription in vitro." EMBO J 21(21): 5899-910.

Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab 67(3): 183-93.

Deininger, P. L., M. A. Batzer, et al. (1992). "Master genes in mammalian repetitive DNA amplification." Trends Genet 8:307-11.

Dewannieux, M., C. Esnault, et al. (2003). "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet 35(1): 41-8.

Felsenstein, J. (1979). "Alternative methods of phylogenetic inference and their interrelationship." Syst Biol 28(1): 49-62.

Feng, Q., J. V. Moran, et al. (1996). "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell 87(5): 905-16.

Gengozian, N., J. S. Batson, et al. (1964). "Hematologic and Cytogenetic Evidence for Hematopoietic Chimerism in the Marmoset, Tamarinus nigricollis." Cytogenet Genome Res 3(6): 384-93.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series.

Han, K., J. Xing, et al. (2005). "Under the genomic radar: the stealth model of Alu amplification." Genome Res 15(5): 655-64.

Harada, M. L., H. Schneider, et al. (1995). "DNA evidence on the phylogenetic systematics of New World monkeys: support for the sister-grouping of Cebus and Saimiri from two unlinked nuclear genes." Mol Phylogenet Evol 4(3): 331-49.

Hill, J. P. (1926). “Demonstration of the embryologia varia (development of Hapalejacchus).” J Anatomy 160: 486-7.

Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." Science 303(5664): 1626-32.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res 12(4): 656-64.

Kolosha, V. O., and S. L. Martin (1997). “In vitro properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition.” Proc Natl Acad Sci USA 94(19): 10155-60.

Konkel, M. K., J. A. Walker, et al. (2010). "LINEs and SINEs of primate evolution." Evolutionary Anthropology: Issues, News, and Reviews 19(6): 236-49.

Lander, E. S., L. M. Linton, et al. (2001). "Initial sequencing and analysis of the human genome." Nature 409(6822): 860-921.

Layne, D. G., and R. A. Power. (2003). “Husbandry, handling, and nutrition for marmosets.” Comp Med 53(4): 351-9.

Locke, D. P., L. W. Hillier, et al. (2011). "Comparative and demographic analysis of orang-utan genomes." Nature 469(7331): 529-33.

Luan, D. D., M. H. Korman, et al. (1993). "Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition." Cell 72(4): 595-605.

Ludlage, E. and K. Mansfield (2003). "Clinical care and diseases of the common marmoset (Callithrix jacchus)." Comp Med 53(4): 369-82.

Mathias, S. L., A. F. Scott, et al. (1991). “Reverse transcriptase encoded by a human transposable element.” Science 254(5039): 1808-10.

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

Meyer, T. J., A. T. McLain, et al. (2012). "An Alu-based phylogeny of gibbons (hylobatidae)." Mol Biol Evol 29(11): 3441-50.

Nikaido, M., A. P. Rooney, et al. (1999). "Phylogenetic relationships among cetartiodactyls based on insertions of short and long interpersed elements: hippopotamuses are the closest extant relatives of whales." Proc Natl Acad Sci U S A 96(18): 10261-6.

Opazo, J. C., D. E. Wildman, et al. (2006). "Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates)." Mol Phylogenet Evol 40(1): 274-80.

Orlosky, F. J. and D. R. Swindler (1975). "Origins of New World monkeys." Journal of Human Evolution 4(2): 77-83.

Osterholz, M., L. Walter, et al. (2009). "Retropositional events consolidate the branching order among New World monkey genera." Mol Phylogenet Evol 50(3): 507-13.

Parker, S. (1990). Grzimek's Encyclopedia of Mammals Vol 2. New York: McGraw-Hill Inc.

Perez, S. I., M. F. Tejedor, et al. (2013). "Divergence Times and the Evolutionary Radiation of New World Monkeys (Platyrrhini, Primates): An Analysis of Fossil and Molecular Data." PLoS One 8(6): e68029.

Ray, D. A., J. Xing, et al. (2005). "Alu insertion loci and platyrrhine primate phylogeny." Mol Phylogenet Evol 35(1): 117-26.

Ray, D. A., J. Xing, et al. (2006). "SINEs of a nearly perfect character." Syst Biol 55(6): 928-35.

Rosenberger, A. L. and W. C. Hartwig (2001). New World Monkeys. eLS, John Wiley & Sons, Ltd.

Rozen, S. and H. Skaletsky. (1999). “Primer3 on the WWW for general users and for biologist programmers.” In Bioinformatics methods and protocols (pp. 365-386). Humana Press.

Salem, A. H., D. A. Ray, et al. (2003). "Alu elements and hominid phylogenetics." Proc Natl Acad Sci U S A 100(22): 12787-91.

Schneider, H. and I. Sampaio. (2013). "The systematics and evolution of New World primates - A review." Mol Phylogenet Evol. In Press.

Schneider, H., I. Sampaio, et al. (1996). "Molecular phylogeny of the New World monkeys (Platyrrhini, primates) based on two unlinked nuclear genes: IRBP intron 1 and epsilon-globin sequences." Am J Phys Anthropol 100(2): 153-79.

Schneider, H., M. P. Schneider, et al. (1993). "Molecular phylogeny of the New World monkeys (Platyrrhini, primates)." Mol Phylogenet Evol 2(3): 225-42.

Schrago, C. G., B. Mello, et al. (2013). "Combining fossil and molecular data to date the diversification of New World Primates." J Evol Biol 26(11): 2438-46.

Shen, M. R., M.A. Batzer, and P.L. Deininger. (1991). “Evolution of the master Alu gene (s).” Journal of Molecular Evolution 33(4): 311-320.

Singer, S. S., J. Schmitz, et al. (2003). "Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates)." Mol Phylogenet Evol 26(3): 490-501.

Smit, A. F. A. (1996). "The origin of interspersed repeats in the human genome." Current opinion in genetics & development 6(6): 743-8.

Smit, A. F. A., R. Hubley, et al. (1996). RepeatMasker Open-3.0.

Smit, A. F., G. Tóth, et al. (1995). “Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences.” Journal of molecular biology 246(3): 401-17.

Steiper, M. E. and M. Ruvolo (2003). "New World monkey phylogeny based on X-linked G6PD DNA sequences." Mol Phylogenet Evol 27(1): 121-30.

Sweeney, C. G., E. Curran, et al. (2012). "Quantitative molecular assessment of chimerism across tissues in marmosets and tamarins." BMC Genomics 13: 98-104.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tardif, S. D., D. A. Smucny, et al. (2003). “Reproduction in captive common marmosets (Callithrix jacchus).” Comparative Medicine 53(4): 364-8.

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

Waddell, P. J., H. Kishino, et al. (2001). "A phylogenetic foundation for comparative mammalian genomics." Genome Inform 12: 141-54.

Wildman, D. E., N. M. Jameson, et al. (2009). "A fully resolved genus level phylogeny of neotropical primates (Platyrrhini)." Mol Phylogenet Evol 53(3): 694-702.

Wislocki, G. B. (1939). "Observations on twinning in marmosets." American Journal of Anatomy 64(3): 445-83.

Xing, J., Y. Zhang, et al. (2009). "Mobile elements create structural variation: analysis of a complete human genome." Genome Res 19(9): 1516-26.

Zhang, W., A. Edwards, et al. (2011). "Alu distribution and mutation types of cancer genes." BMC Genomics 12: 157-74.