Louisiana State University
LSU Scholarly Repository

Honors Theses

Ogden Honors College

4-2017

Secrets of SINES: the Story of Platy-1 Propagation in the Saimiri Lineage

Sarah Brantley

Follow this and additional works at: https://repository.lsu.edu/honors_etd

Part of the Biology Commons

Recommended Citation
Brantley, Sarah, "Secrets of SINES: the Story of Platy-1 Propagation in the Saimiri Lineage" (2017). *Honors Theses*. 226.
https://repository.lsu.edu/honors_etd/226

This Thesis is brought to you for free and open access by the Ogden Honors College at LSU Scholarly Repository. It has been accepted for inclusion in Honors Theses by an authorized administrator of LSU Scholarly Repository. For more information, please contact ir@lsu.edu.

Secrets of SINES: the Story of Platy-1 Propagation in the *Saimiri* Lineage

by

Sarah Brantley

Undergraduate honors thesis under the direction of

Dr. Mark A. Batzer

Department of Biological Sciences

Submitted to the LSU Roger Hadfield Ogden Honors College in partial fulfillment

of

the Upper Division Honors Program

April, 2017

Louisiana State University

& Agricultural and Mechanical College

Baton Rouge, Louisiana

Introduction

I. *Saimiri* Background

Squirrel monkeys are a diverse group of New World Monkeys from the Cebidae family of Platyrrhines native to Central and South America (Hershkovitz 1984). Even today the phylogenetics of this species and its relationship to the other platyrrhini clades are not well understood. This is due to several factors including the rapid divergence of squirrel monkeys from other platyrrhines and a high degree of hybridization among the different squirrel monkey species (Wildman et al. 2009; Carneiro et al. 2016). Between one and seven species of squirrel monkeys have been proposed by numerous studies, with up to 17 subspecies (Ruiz-Garcia et al. 2015). Some of the more commonly accepted species are *Saimiri sciureus, S. oerstedii, S. collinsi, S. ustus, S. boliviensis, S. cassiquiarensis,* and *S. vanzolinii* (Carneiro et al. 2016). Of these, *S. oesterdii* is classified as endangered and *S. vanzolinii* is classified as vulnerable (Rogers et al. 2015). Furthermore, squirrel monkeys have become a useful biological model for studying human diseases such as several types of herpes viruses and polyomaviruses due to the similarities between squirrel monkey and human immune systems (Rogers et al. 2015).

One of only four primate species located throughout the Amazonian region, the majority of squirrel monkey species inhabit South America, with the exception of *S. oerstedi citrinellus* and *S. o. oerstedii* which inhabit Central America (Hershkovitz 1984; Lynch Alfaro et al. 2015). The remaining squirrel monkey species live in the Amazon basin and Guianas (Hershkovitz 1984; Carneiro et al. 2016). The numerous squirrel monkey species are mainly found in certain areas of South America, however, many studies have found that zones of hybridization are common, resulting in intermingling of the species (Hershkovitz 1984; Ruiz-Garcia et al. 2015;

Carneiro et al. 2016). This understandably complicates any attempt to elucidate an accurate squirrel monkey and new world monkey evolutionary tree, the ultimate objective for this study.

Morphological studies have been conducted to further understand the differences among the many species of squirrel monkeys (Hershkovitz 1984; Merces et al. 2015). Hershkovitz originally segregated squirrel monkeys into two groups: gothic and roman arched. Both groups refer to the white markings above the eyes of the monkeys, but Gothic arches are much more pronounced than Roman arches (Hershkovitz 1984). The Roman arched species include *S. boliviensis peruviensis* and *S. b. boliviensis*, and are found south of the Rio Maranon-Amazonas (Hershkovitz 1984). The Gothic arched type, *S. sciureus*, inhabits the area north of Rio Maranon-Solimoes-Amazonas and east of Rio Negro (Hershkovitz 1984). More recently, Merces et. al. (2015) described two geographically separated color states of the crown at the top of their heads: a yellower cap south of the Amazon and on the Marajo Archipelago and a grayer cap north of the same river. Furthermore, they found that the color difference also extended to the hip and saddle areas of the monkeys. These findings corroborated the existing speciation of *S. scuireus* north of the Amazon and *S. collinsi* south of the river (Merces et al. 2015).

Squirrel monkeys are classified as Cebidea within the broader parvorder Platyrrhini (Steiper and Young 2006; Wildman et al. 2009). Platyrrhines are New World Monkeys, and they diverged from Old World Monkeys, Catarrhines, between 37.3-52.4 million years ago (mya) (Steiper and Young 2006). Within Platyrrhines, the Cebidae family includes three clades: *Aotinae, Cebinae,* and *Callitrichinae* (Wildman et al. 2009). The relationship between these three clades is still being studied, but Steiper et. al. estimated the divergence between the squirrel monkey and capuchin (*Cebinae*) and marmoset (*Callitrichinae*) clades to be 17.1 mya. The exact relationship between the three groups is difficult to characterize morphologically due to the

extremely rapid divergence of so many New World Monkey species and their frequent hybridization. Genetic studies have emerged as a more accurate method of determining evolutionary relationships such as this, with studies on mitochondrial DNA, *Alu* insertions, non-coding DNA markers, and genomic data contributing theoretical phylogenetic trees (Steiper and Young 2006; Wildman et al. 2009; Merces et al. 2015; Ruiz-Garcia et al. 2015).

II. Mobile Element Background

First described by Barbara McClintock in 1950, mobile elements have become an increasingly popular target for evolutionary studies of all eukaryotes, but especially primates (McClintock 1950). Mobile elements, also known as "jumping genes" or "transposable elements", are repetitive stretches of nucleotides found in a genome. They can be a simple repeat of a few bases like a microsatellite, a string of nucleotides thousands of bases long found in multiple areas of the genome, and several sizes in between (Konkel et al. 2010). In a 2011 study, it was estimated that over two-thirds of the human genome is composed of transposable elements (de Koning et al. 2011). Repetitive elements are either classified as tandemly arrayed or interspersed (Batzer and Deininger 2002). Interspersed elements are further classified based on size. These categories are Short Interspersed Elements (SINEs) of less than 500 bp in length and Long Interspersed Elements (LINEs) of greater than 500 bp, often reaching kilobases (Batzer and Deininger 2002). Mobile elements can also be characterized by their method of propagation, either through an RNA or a DNA intermediate. Those that utilize an RNA intermediate replicate themselves while preserving the previous insertions, sometimes referred to as a "copy and paste" mechanism, are termed retrotransposons and considered Class I elements (Konkel et al. 2010). Conversely transposons, Class II elements, utilize a DNA intermediate and do not preserve the previous insertions, known as a "cut and paste" propagation method (Konkel et al. 2010).

Within the broad retrotransposon category, elements either have long terminal repeats (LTR) or they are non-LTR, and within the non-LTR retrotransposons, they either contain the necessary components to replicate themselves (autonomous) or they do not (non-autonomous) (Kazazian 2004).

Retrotransposable elements utilize an RNA intermediate to essentially copy themselves from one area of the genome and paste themselves into another part. This is accomplished through a method called target primed reverse transcription (TPRT) initially described in 1993 (Luan et al. 1993; Batzer and Deininger 2002). TPRT involves cutting the bottom strand of DNA at a target site, usually 5'-TTTT/AA-3', by the L1 endonuclease. Then, the retrotransposon RNA binds at the DNA cleavage site and the L1 reverse transcriptase performs reverse transcription on the element's RNA. Finally, the second strand is nicked and the complementary nucleotides are added by a mechanism not well understood (Fig 1). The target side duplications(TSD) produced at both ends of the inserted element is a hallmark of TPRT retrotransposition (Konkel et al. 2010). L1 elements, within the LINEs category, are the only currently active autonomous retrotransposon, which means they encode their own enzymatic machinery for replication. They also demonstrate a *cis*-preference, indicating that RNA recruits its own translated proteins (Konkel et al. 2010). *Alu* elements, a primate specific SINE, must use other element's proteins because they have no open reading frame of their own, so no proteins can be made from its sequence (Batzer and Deininger 2002). The open reading frame (ORF) is the portion of DNA that begins with a start codon and ends in a stop codon, with no premature stop codons between them. ORFs typically code for proteins or polypeptides (Konkel et al. 2010). Additionally, the poly-A (or A-rich) tails that are a standard feature of both LINEs and *Alu*s allow the *Alu* elements to compete for the LINEs' enzymatic proteins necessary for *Alu*

amplification. This theory of *Alu* proliferation is supported by the fact that both LINEs and *Alu*s have become exceedingly more common in primate genomes in the last 150 million years, so it

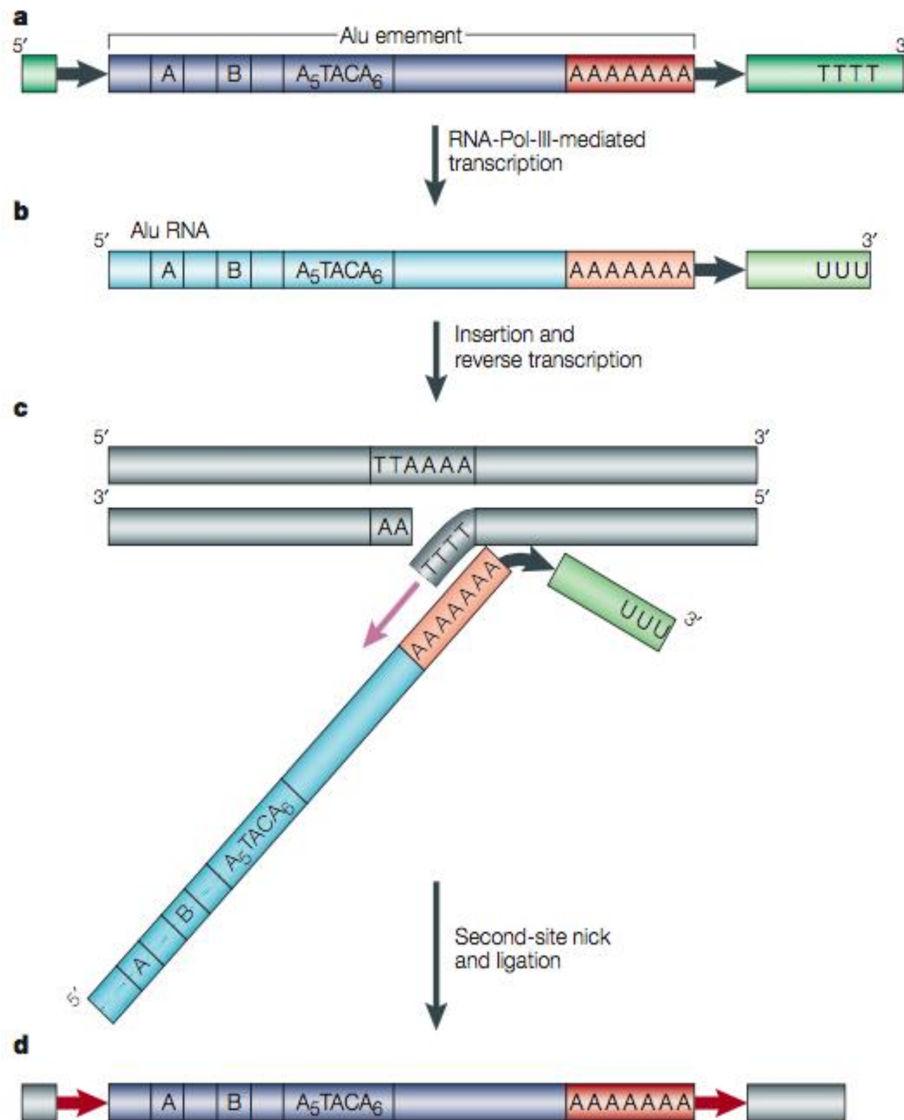


Fig 1. Retrotransposition of an *Alu* element (Taken with permission from Batzer and Deininger, 2002)
The standard *Alu* structure includes TSDs shown by the flanking black arrows, A and B boxes for the RNA polymerase III promoter signal, a middle A-rich region, and a poly-A tail (a). Reverse transcription occurs to create an RNA sequence with a 3’ end that extends into the flanking sequence until it reaches the UUU complement RNA polymerase III termination signal (b). The A tail at the end of the element anneals directly to the DNA at the insertion site for TPRT (c). The nick at the insertion site is made by the L1 endonuclease at the TTAAAA consensus site. A new set of TSDs, shown by the red arrows, is made during the insertion event (d).

is reasonable to suppose that LINEs are essential for *Alu* amplification (Batzer and Deininger 2002).

III. Advantages of Mobile Element Phylogenetic Studies

Mobile element propagation, in particular *Alu* element propagation, have been studied to determine divergence of several primate species including but not limited to humans, orangutans, baboons, gibbons, lemur, marmoset, macaque and chimpanzees (Wang et al. 2006; Liu et al. 2009; Meyer et al. 2012). The retrotransposons that are seemingly frozen in place in the genome due to their identical by descent TPRT replication method are great genetic markers for researchers (Batzer and Deininger 2002). *Alu* elements tend to be passed on to offspring with very few radical mutations. Additionally, the presence of a shared element with matching TSDs at a given locus indicates a common ancestor between species (Walker et al. 2012). In this way, a researcher can theoretically analyze the hierarchical *Alu* elements to construct a homoplasy free phylogenetic tree for a primate species.

Homoplasy is the shared genetic state not inherited from a common ancestor but is created due to independent evolution, and using homoplasious genetic markers may produce an incorrect genetic tree (Konkel et al. 2010; Meyer et al. 2012). *Alu* elements and other SINEs are nearly homoplasy free because the absence of the element in an organism's genome is the ancestral state (Batzer et al. 1994; Deininger and Batzer 1999; Ray et al. 2006; Meyer et al. 2012). There are four situations in which SINEs can indicate misleading results: lineage sorting, parallel insertions, SINE deletions, and paralogous insertions. However, these situations can be resolved via Sanger sequencing or further polymerase chain reaction (PCR) amplification (Ray et al. 2006).

III. Platy-1 Element Background: A New World Monkey Specific Element

The Platy-1 repetitive element was originally reported in 2016 by Konkel et. al. in the *Callithrix jacchus* (common marmoset) reference genome [calJac3.2]. The short, ~100 bp, elements are unique to platyrrhines, but have thus far only been studied in *C. jacchus*. Platy-1 elements have typical features of SINE retrotransposons including target site duplications (TSDs) of > 7 bp, an endonuclease cleavage site, and a homopolymeric tail of adenosine nucleotides (poly-A) tail (Konkel et al. 2016). Within the marmoset reference genome, 62 subfamilies and a total of 2268 full length Platy-1 elements were discovered, with 10% of the elements matching their respective subfamily consensus sequence perfectly (Konkel et al. 2016). Platy-1 elements with no or very few mutations, among other factors including a long poly-A tail, could potentially be source elements and provide a trampoline for these "jumping genes", pending further testing (Walker et al. 2012). A source element for any retrotransposon is a nearly perfect or perfect element that contains all of the necessary "parts" for TPRT (Walker et al. 2012).

The origin of the Platy-1 repeat family seems to be similar to that of *Alu* elements. Konkel et. al. (2016) determined Platy-1 elements to be derived from 7SL RNA. The most likely mode of replication is via target-primed reverse transcription (TPRT) with L1 enzymatic machinery due to the presence of TSDs, an endonuclease cleavage site, and a poly-A tail without a prior polyadenylation signal, like *Alu* elements (Konkel et al. 2016). This theory requires further testing to confirm, though.

Platy-1 elements seemed to have increased propagation in marmoset following a slower beginning. This almost exponential growth could be due to the increasing number of potential source elements as the repeat family became more established (Konkel et al. 2016). This

expansion coincided with the rapid speciation of New World Monkeys, indicating that these repeats, although present in all platyrrhines, may have expanded independently in other NWM species lineages, too.

Materials and Methods

I. Squirrel Monkey Specificity and Primer Design

Using methods outlined in Konkel et al. (2016), the 62 Platy-1 consensus sequences were compared to the squirrel monkey, *Saimiri boliviensis*, reference genome [Broad/saiBol1] in UCSC's BLAT search engine and RepeatMasker to identify candidate loci (Smit 1996-2010; Kent 2002). Full length candidate Platy-1 elements with 500 bp of flanking sequence were compared to the common marmoset reference genome [WUGSC 3.2/calJac 3] in BLAT. As of the time of the experiment, the common marmoset was the closest New World Monkey relative to the squirrel monkey with a published reference genome. Any locus that did not have a matching squirrel monkey potential Platy-1 element was noted as possibly unique to squirrel monkeys. Those noted elements were then computationally examined in BioEdit (Hall 1999) for Platy-1 markers: target site duplications (TSDs), endonuclease cleavage sites, start site, and poly-Adenosine tail. An element must contain all four markers to be considered a full length Platy-1 element.

The loci determined to contain Platy-1 elements unique to squirrel monkeys were put into individual BioEdit files also containing the homologous sequences from the reference genomes of humans [GRCh38/hg38], rhesus macaque monkeys [BCM Mmul_8.0.1/rheMac 8], orangutans [WUGSC 2.0.2/ponAbe2], and chimpanzees [CSAC 2.1.4/panTro4], obtained from BLAT. Forward and reverse oligonucleotide primers were created for 15 loci of the 22 that were identified (Appendix Table 1). The oligonucleotide primers were created using Primer3 (v.

0.4.0) with the following adjustments to the standard settings: max Tm difference of 2.0, Max Tm of 62.0, Max Poly X of 3, and minimum GC% of 40.0 (Untergasser et al. 2012). The resulting oligonucleotide primer pairs were screened via Blat's InSilico PCR to ensure they could computationally function properly in squirrel monkeys and the common marmoset, if not all six species. The oligonucleotide primer sequences were ordered from Sigma Aldrich (Woodlands, TX, USA).

II. Polymerase Chain Reaction and Gel Electrophoresis

Each oligonucleotide primer pair was tested on human (HeLa) cell DNA with polymerase chain reaction (PCR) at 57° C annealing temperature and gel electrophoresis to confirm amplification. For those oligonucleotide primer pairs that were unable to amplify at 57° C, a temperature gradient was performed with a TLE negative, HeLa, and KB 4544 squirrel monkey sample, to find optimal PCR annealing temperatures. Each pair was then tested against a New World Monkey panel that contained 19 samples and a squirrel monkey panel containing 32 *Saimiri* samples plus HeLa and a common marmoset outgroup samples (Appendix Tables 2 and 3). Included in the NWM panel samples were one *Saimiri s. sciureus* (squirrel monkey), three *Cebus* (capuchin monkey), one *Callithrix jacchus* (common marmoset), and one *Aotus trivirgatus* (owl monkey) samples. Some of the oligonucleotide primer pairs did not amplify well, so they were redesigned and retested. Of the redesigned primer sets, some still failed to amplify the target DNA sequence.

The PCR protocol was as follows. For the amplification 25 ng DNA of the species of interest were added to each well of a PCR plate. The master mix contained 10X PCR buffer (50 mM KCl; 10 mM TrisHCl; pH 8.4), 200 nM of each oligonucleotide primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, and 1-2 U *Taq* DNA Polymerase. The initial denaturation was done at 94 °C for

60 s. Then 32 cycles of denaturation were performed at 94 °C for 30 s, the optimum annealing temperature for 30 s, and an extension at 72 °C for 30 s. A final extension was done at 72 °C for 2 minutes before the reaction was terminated.

Aliquots of 20 µl of PCR product were inserted into a horizontal gel chamber containing a 2% agarose gel with 0.2 µg/ml ethidium bromide. The agarose gel was run for 40 minutes at 200 V. The DNA fragments were analyzed using UV fluorescent light and images were saved with a BioRad Chemidoc XRS imaging system (Hercules, CA).

III. Sanger Sequencing

Some loci were further evaluated via Sanger Sequencing procedures to confirm shared elements between squirrel monkey and capuchin (Sanger et al. 1977). Normal PCR reactions for each loci were performed with quadruple replicates to be purified with the Wizard SV gel purification kit (Promega Corporation, Madison, WI, catalog A9282). The protocol was followed with one modification: the 50 µl elution was performed twice to produce 100 µl, which was then dried in a SpeedVac (ThermoSavant SPD 111V). The DNA was reconstituted in 30 µl TVLE (Tris Very Low Ethylenediaminetetraacetic acid [EDTA], 10 mM Tris/0.05 mM EDTA). 4 µl of the reconstituted DNA was used for chain termination cycle sequencing using Big Dye Terminator v3.1. The cycle sequencing was then performed with conditions of 95 °C initial denaturation for 2 minutes, then 40 cycles of 95 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 4 minutes, ending in a hold at 4 °C. The sequencing reactions were cleaned with a standard ethanol precipitation to remove the unincorporated dye terminators. They were then stabilized in 15 µl Hi-Di Formamide (Life Technologies, Inc.). The cleaned sequencing reactions were run on capillary electrophoresis on an ABI 3130x1 Genetic Analyzer (Applied Biosystems, Inc., Foster City, CA). The sequence quality was evaluated using ABI software

Sequence Scanner v1.0. Each locus was then examined for the presence or absence of a Platy-1 element.

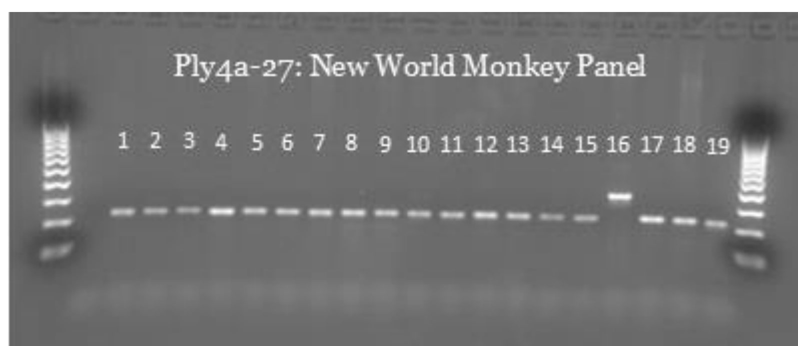
Results and Discussion

A total of 208 Platy-1 loci were retrieved by Dr. Miriam Konkel from the squirrel monkey reference genome [saiBol1] using the same methods as previously described for the marmoset genome [calJac3] (Konkel et al. 2016). Compared to the 2275 Platy-1 elements reported in the 2016 study from marmoset, it is clear that the squirrel monkey lineage has a much lower number of lineage-specific Platy-1 insertions. Of the 208 candidate loci, 54 were confirmed to contain full-length Platy-1 sequences possessing the four typical structural features: target site duplications, endonuclease cleavage site, GGGG or GGAG start site, and an adenosine-rich tail, or A-tail. The 54 sequences were then computationally compared to the common marmoset reference genome [calJac3] to identify those unique to the squirrel monkey reference genome, resulting in 22 lineage-specific Platy-1 insertions events.

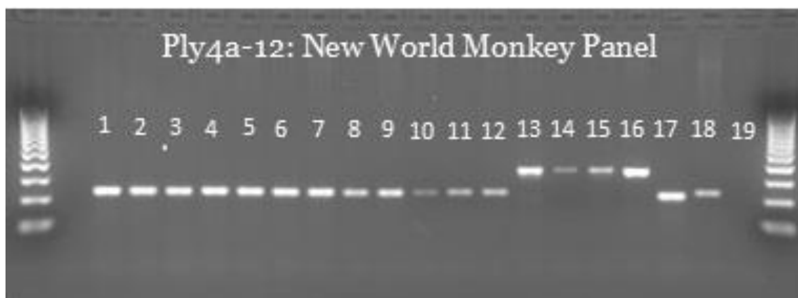
Oligonucleotide primer pairs for PCR were created for all of the loci containing Platy-1 elements unique to the squirrel monkey reference genome with adequate flanking sequence (15 of 22). A list of the primer sequences and corresponding loci is in the appendix as Table 1. The primers were analyzed by PCR on a DNA panel containing 19 samples from a spectrum of New World Monkeys and then on a second DNA panel containing 32 individual squirrel monkey samples from multiple *Saimiri* species. A complete list of these DNA samples with species names and ID numbers are shown in Appendix Table 2 and 3. DNA samples from human (HeLa cell line) and common marmoset were included as outgroup samples. Because each oligonucleotide primer set was tested against two DNA panels, each locus has two corresponding gel images (Fig 2 and 3). When examining a gel electrophoresis, the DNA bands are compared

to a DNA ladder at either end of the row to determine an approximate size of the amplicon. In this study, each ladder was 100 bp, so each "rung" increased by 100 bases. The expected amplicon size is known when the oligonucleotide primers are designed, so identifying filled and empty site bands is a confirmation step. Any DNA band that is the same size as a predicted filled site band theoretically contains the Platy-1 element, while the bands ~100 bp lower on the gel indicate empty sites that do not contain the element.

A.



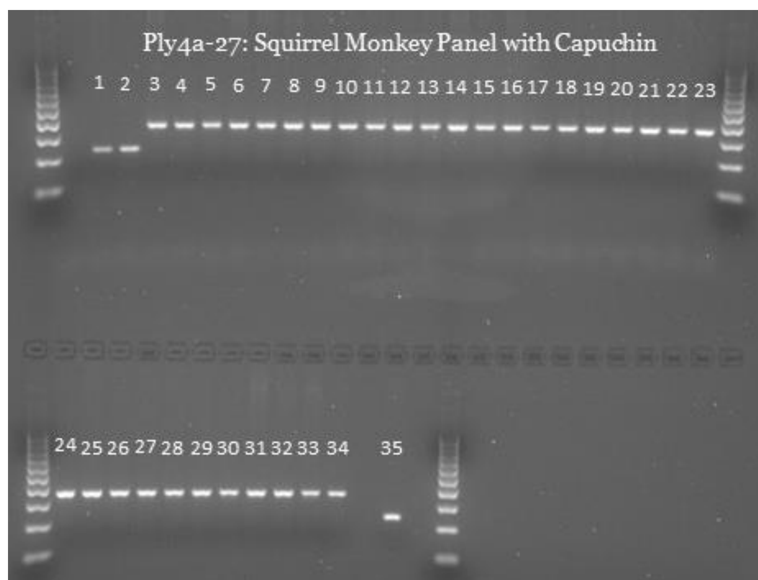
B.



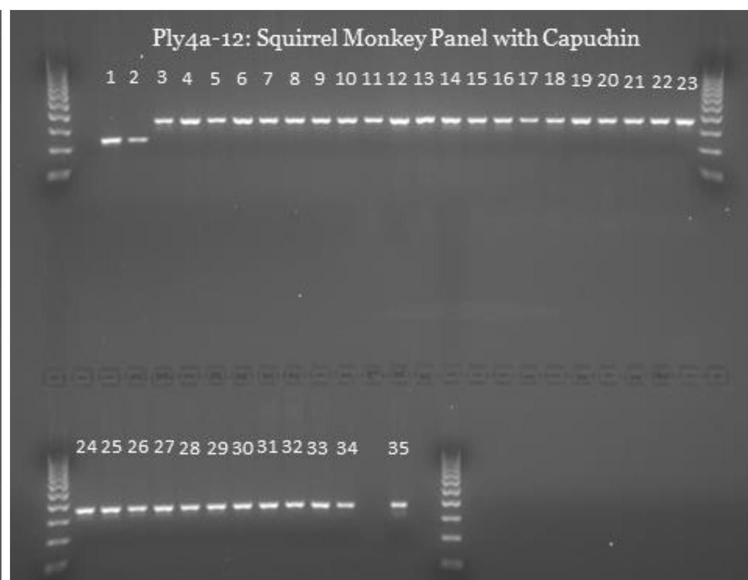
Lane Number	Species
1	Human
2	Common Chimpanzee
3	African Green monkey
4	Woolly monkey
5	White bellied spider monkey
6	black-handed spider monkey
7	Bolivian red howler monkey
8	Common marmoset
9	Pygmy marmoset
10	Goeldi's marmoset
11	red-chested mustached tamarin
12	Geoffroys saddle-back tamarin
13	Capuchin monkey
14	Capuchin monkey
15	Capuchin monkey
16	Squirrel monkey
17	Owl monkey
18	Northern white-faced saki
19	Bolivian gray titi

Fig 2. Agarose gel pictures of PCR amplification of Platy-1 insertions on the New World Monkey DNA panel. Species used in the panel are listed to the right. On either end of each row is a 100 bp DNA ladder for amplicon size reference. A) Locus 4a-27: The element was homozygous present only in the squirrel monkey sample, and absent from the other new world monkeys and humans. B) Locus 4a-12: The element was homozygous present in the squirrel monkey sample and all three capuchin samples, indicating the element is shared among those species. It was absent from the remaining new world monkey and human samples.

A.



B.



Lane Number	Species	Lane Number	Species
1	Human	18	Bolivian squirrel monkey
2	Common marmoset	19	Bolivian squirrel monkey
3	common Squirrel monkey	20	Bolivian squirrel monkey
4	common squirrel monkey	21	Bolivian squirrel monkey
5	common squirrel monkey	22	Bolivian squirrel monkey
6	common squirrel monkey	23	Bolivian squirrel monkey
7	common squirrel monkey	24	Bolivian squirrel monkey
8	common squirrel monkey	25	Bolivian squirrel monkey
9	common squirrel monkey	26	Bolivian squirrel monkey
10	common squirrel monkey	27	Bolivian squirrel monkey
11	common squirrel monkey	28	Bolivian squirrel monkey
12	common squirrel monkey	29	Puruvian squirrel monkey
13	common squirrel monkey	30	Puruvian squirrel monkey
14	common squirrel monkey	31	Peruvian Black-Capped Squirrel Monkey
15	Bolivian squirrel monkey	32	Panamanian Red-Backed Squirell Monkey
16	Bolivian squirrel monkey	33	Ecuadorian Squirrel Monkey
17	Bolivian squirrel monkey	34	Squirrel monkey
		35	Capuchin

Fig 3. Agarose gel pictures of PCR amplification of Platy-1 insertions on the Squirrel Monkey panel with capuchin added. Species used in the panel are listed above. On either end of each row is a 100 bp DNA ladder for amplicon size reference. A) Locus 4a-27: The element was homozygous present in all squirrel monkey samples, and absent from the human, common marmoset, and capuchin samples. B) Locus 4a-12: The element was homozygous present in all of the squirrel monkey samples and the capuchin sample. It was absent from the human and common marmoset samples.

With squirrel monkey DNA samples from both panels, 15 loci were confirmed to contain the Platy-1 element. One primer set did not amplify well in the human and other old world monkey outgroups, and therefore gave no empty site, and another primer set was nonfunctional due to the Platy-1 element inserted into an existing *Alu* element. This data is not shown.

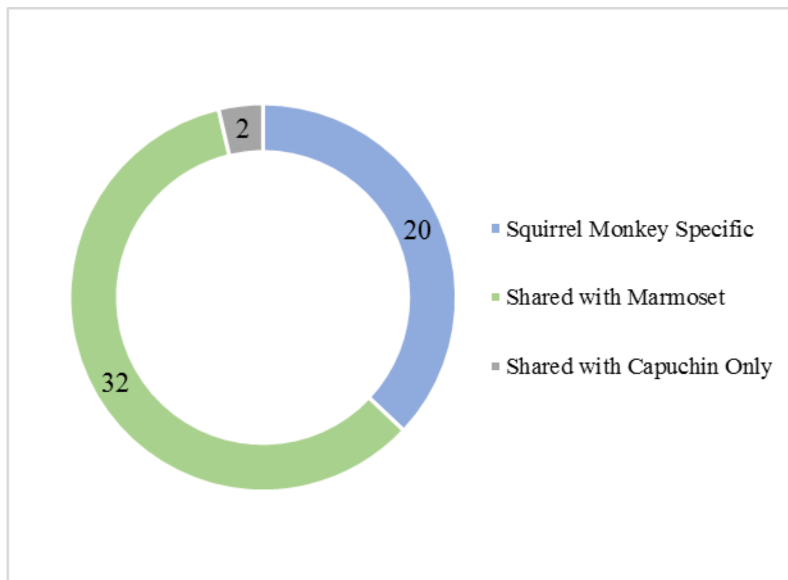


Fig 4. Graphical representation of the Platy-1 element distribution between squirrel monkeys, marmosets, and capuchins. The majority of the 54 full length Platy-1 elements present in squirrel monkeys are shared with the common marmoset. Only two loci identified in this study are shared between squirrel monkeys and capuchins.

Following PCR analysis, two loci required Sanger sequencing (Sanger et al. 1977) because they showed a filled site amplification in capuchin in addition to squirrel monkey. Additionally, another loci was Sanger sequenced due to potential heterozygosity in the KB4544 squirrel monkey sample of the New World Monkey panel. The sequencing was necessary to confirm that either the Platy-1 element was shared in the capuchin DNA, and not another insertion of a similar size, or that the secondary band of the squirrel monkey sample was a true empty site, indicating a polymorphic insertion. The two loci sequenced to confirm the capuchin Platy-1 PCR amplicon band were successful. The capuchin sequence matched the squirrel monkey sequence, indicating that the Platy-1 element is shared in both species at those two loci. This was also confirmed by matching target sight duplications. Thus, two of the loci thought to be squirrel monkey specific were shared with capuchin (Fig 4). The third locus, however, was

not confirmed to be heterozygous. The "empty" site sequence was not clear, and upon further PCR on the squirrel monkey panel, the second band that initially indicated a putative heterozygote was present in all 32 squirrel monkey DNA samples. This could be a paralogous amplification, meaning the Platy-1 element inserted into a previously duplicated sequence of DNA (Batzer et al. 1991). So, the oligonucleotide primers amplified both the loci containing the Platy-1 element and the duplicated site, resulting in two distinct bands on the agarose gel, data not shown.

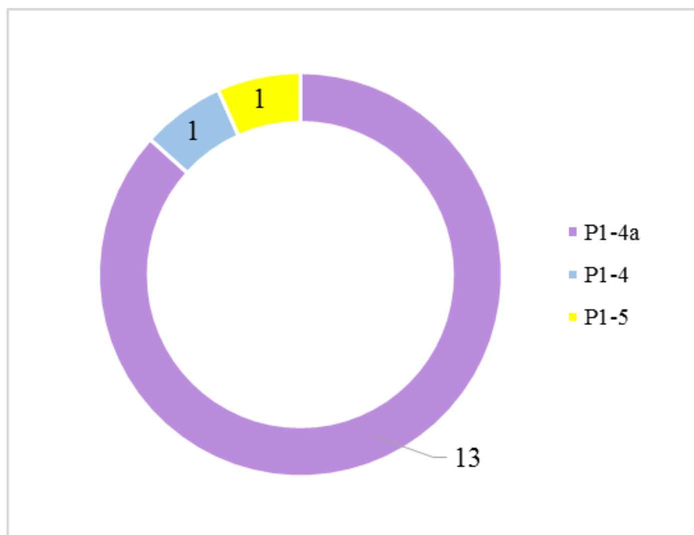


Fig 5. A graphical representation of the subfamilies of Platy-1 elements unique to squirrel monkeys. All but two Platy-1 elements unique to squirrel monkeys are from the 4a subfamily. One element was in the 4 and the 5 subfamilies each. This indicates that the 4a subfamily was much more successful in propagating in the species.

The majority of the 15 squirrel monkey specific Platy-1 insertions analyzed by PCR were from subfamily 4a, previously characterized by Konkel, et al (2016) (Fig 5). This subfamily is one of the oldest Platy-1 subfamilies that were originally discovered in the marmoset genome (Fig 6). In the common marmoset, the Platy-1 propagation was rapid and widespread. It contains 62 subfamilies, with some newer ones creating bush-like nodes on the evolutionary tree. Some of the loci identified by Konkel, et al (2016) are polymorphic within common marmosets and therefore likely still propagating today in that lineage. This is in sharp contrast to the findings of this study, where only 15 loci were confirmed to contain full-length squirrel monkey specific Platy-1 elements and none were polymorphic insertions across a diverse panel of 32 different squirrel monkey individuals. This

suggests that while the Platy-1 element continued to propagate rapidly in the common marmoset, it slowed to a stop, or near complete cessation, in the squirrel monkey lineage.

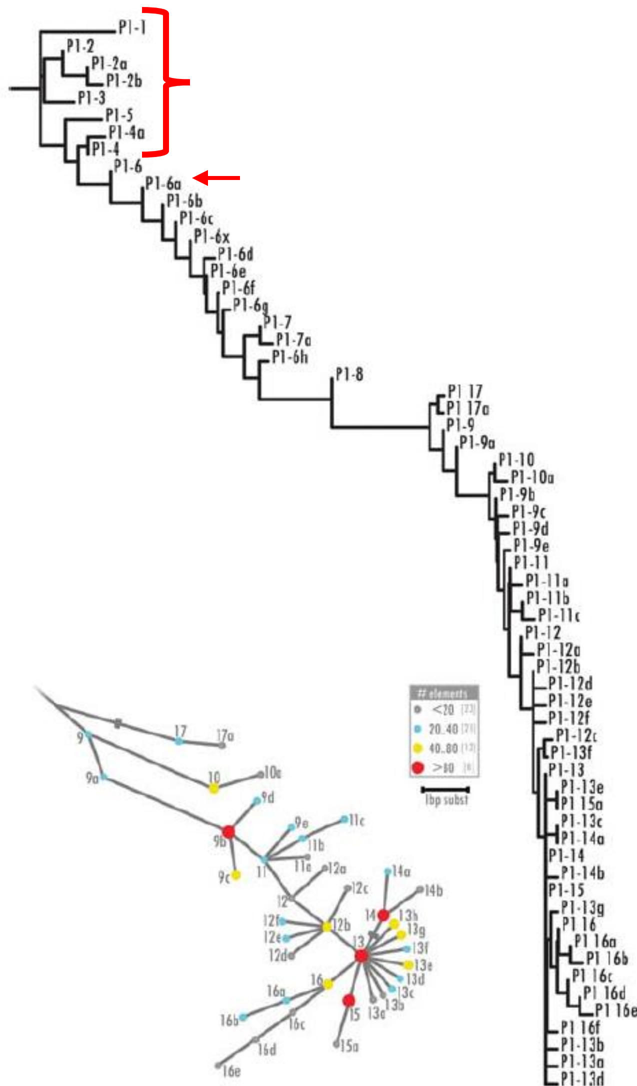


Fig 6. The evolutionary tree of Platy-1 subfamilies identified in the common marmoset genome with the squirrel monkey subfamilies highlighted in red. All of the Platy-1 elements identified in the squirrel monkey reference genome were from subfamilies 1, 2, 2a, 2b, 3, 4, 4a, 5, and 6a. These subfamilies are from the older end of the tree, indicating that the split between squirrel monkeys and common marmosets was early in the propagation of Platy-1 elements. This figure was adapted from Konkel et al 2016 to highlight the few subfamilies present in the squirrel monkey genome.

Conclusions and Future Work

Although this study did not recover thousands of Platy-1 repeats in the *Saimiri* (squirrel monkey) genome as in the *Callithrix jacchus* (common marmoset) genome, it is interesting to note the approximate 10-fold difference in propagation rates between the two species. The rapid and diverse growth of Platy-1 elements in the common marmoset genome as demonstrated by Konkel et. al. (2016) was not seen in the squirrel monkey genome. This may have been caused

by the initial elements inserted into the squirrel monkey genome being too old (subject to mutation and decay) to continue propagation at a substantive rate, but further characterization is needed to confirm the undelying molecular basis for the difference. The slow propagation is not due to a lack of available enzymatic machinery because it has been shown that L1, the element that provides the enzymes necessary for TPRT, has recently amplified in *Saimiri* among other NWM species (Feng et al. 1996; Boissinot et al. 2004).

In terms of creating a better phylogentic tree of NWMs and the place *Saimiri* has in it, these Platy-1 repeat elements do support a tree where *Saimiri* and *Cebus* are more closely related than *Saimiri* and *Callithrix*. While none of the Platy-1 repeats originally identified as unique to squirrel monkeys were polymorphic among *Saimiri* species, it was found that two of the repeats were shared in capuchin monkeys.

In the future, Platy-1 elements can be used to elucidate a more accurate NWM phylogenetic tree by examining more platyrrhine species. Particularly, the Platy-1 elements in capuchin monkeys should be examined to confirm the relationship they have with squirrel monkeys as suggested by these experimental findings. However, as more NWM reference genomes become published, such as the capuchin and owl monkey genomes currently in scaffold format, this Platy-1 element research can continue. As the only platyrrhine-specific SINE reported thus far, Platy-1 elements represent a novel tool for the future of NWM genomics research.

Acknowledgements

I would like to thank Dr. Mark A. Batzer for his mentorship in both research and school, and for allowing me to work in his Laboratory of Comparative Genomics since August 2013. I would also like to thank Jerilyn Walker for her endless patience and guidance as this project was

conceived and executed. Additionally, thank you to the past and present members of Dr. Batzer's lab for their support and friendship, including but not limited to Jasmine Brown Baker, Cody Steely, Sreeja Sanampudi, Michael Denham, Paulina Gonzales Quiroga, and Jackson Mierl. I would also like to thank Dr. Mark A. Batzer, Dr. Kevin Smith, and Dr. Leaf Boswell for serving on my thesis committee.

Furthermore, I would like to give a special thanks to Dr. John A. Vanchiere, M.D., Ph.D. Chief, Pediatric Infectious Diseases, LSUHSC-Shreveport, LA, for providing very much needed squirrel monkey DNA samples. I would also like to thank the American Type Culture Collection, the New England Regional Primate Research Center, the San Diego Frozen Zoo Conservation and Research for Endangered Species, the LSU Museum of Natural Science Collection of Genetic Resources, the Burke Museum of Natural History and Culture, the Coriell Institute for Medical Research, the Yale Peabody Museum of Natural History, the Museum of Vertebrate Zoology, the Michale E. Keeling Center for Comparative Medicine and Research, the San Diego Zoo Global Biomaterials Review Group San Diego Zoo Institute for Conservation Research, the Integrated Primate Biomaterials and Information Resource, and Kimberly Phillips, Trinity University for providing DNA samples used in this research. This research was supported by the National Institutes of Health R01 GM59290 (M. A. B.). Additionally, this research was supported in part by the LSU President's Future Leaders in Research Scholarship (S. A. B.).

References

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. *Nature reviews Genetics* **3**(5): 370-379.

Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL. 1991. Amplification dynamics of human-specific (HS) Alu family members. *Nucleic acids research* **19**(13): 3619-3623.

Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ et al. 1994. African origin of human-specific polymorphic Alu insertions. *Proceedings of the National Academy of Sciences of the United States of America* **91**(25): 12288-12292.

Boissinot S, Roos C, Furano AV. 2004. Different rates of LINE-1 (L1) retrotransposon amplification and evolution in New World monkeys. *Journal of molecular evolution* **58**(1): 122-130.

Carneiro J, Rodrigues-Filho LF, Schneider H, Sampaio I. 2016. Molecular data highlight hybridization in squirrel monkeys (Saimiri, Cebidae). *Genetics and molecular biology* **39**(4): 539-546.

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

Deininger PL, Batzer MA. 1999. Alu repeats and human disease. *Molecular genetics and metabolism* **67**(3): 183-193.

Feng Q, Moran JV, Kazazian HH, Jr., Boeke JD. 1996. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**(5): 905-916.

Hall TA. 1999. Bioedit: a user-friendly biological sequence alignment editor and analysis program from Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**: 95-98.

Hershkovitz P. 1984. Taxonomy of squirrel monkeys genus Saimiri (cebidae, platyrrhini): A preliminary report with description of a hitherto unnamed form. *American Journal of Primatology* **6**(4): 257-312.

Kazazian HH, Jr. 2004. Mobile elements: drivers of genome evolution. *Science (New York, NY)* **303**(5664): 1626-1632.

Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome research* **12**(4): 656-664.

Konkel MK, Ullmer B, Arceneaux EL, Sanampudi S, Brantley SA, Hubley R, Smit AF, Batzer MA. 2016. Discovery of a new repeat family in the Callithrix jacchus genome. *Genome research* **26**(5): 649-659.

Konkel MK, Walker JA, Batzer MA. 2010. LINEs and SINEs of primate evolution. *Evolutionary anthropology* **19**(6): 236-249.

Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE. 2009. Comparative analysis of Alu repeats in primate genomes. *Genome research* **19**(5): 876-885.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 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.

Lynch Alfaro JW, Boubli JP, Paim FP, Ribas CC, Silva MN, Messias MR, Rohe F, Merces MP, Silva Junior JS, Silva CR et al. 2015. Biogeography of squirrel monkeys (genus Saimiri): South-central Amazon origin and rapid pan-Amazonian diversification of a lowland primate. *Molecular phylogenetics and evolution* **82 Pt B**: 436-454.

McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* **36**(6): 344-355.

Merces MP, Lynch Alfaro JW, Ferreira WA, Harada ML, Silva Junior JS. 2015. Morphology and mitochondrial phylogenetics reveal that the Amazon River separates two eastern squirrel monkey species: Saimiri sciureus and S. collinsi. *Molecular phylogenetics and evolution* **82 Pt B**: 426-435.

Meyer TJ, McLain AT, Oldenburg JM, Faulk C, Bourgeois MG, Conlin EM, Mootnick AR, de Jong PJ, Roos C, Carbone L et al. 2012. An Alu-based phylogeny of gibbons (hylobatidae). *Molecular biology and evolution* **29**(11): 3441-3450.

Ray DA, Xing J, Salem AH, Batzer MA. 2006. SINEs of a nearly perfect character. *Systematic biology* **55**(6): 928-935.

Rogers DL, McClure GB, Ruiz JC, Abee CR, Vanchiere JA. 2015. Endemic Viruses of Squirrel Monkeys (Saimiri spp.). *Comparative medicine* **65**(3): 232-240.

Ruiz-Garcia M, Luengas-Villamil K, Leguizamon N, de Thoisy B, Galvez H. 2015. Molecular phylogenetics and phylogeography of all the Saimiri taxa (Cebidae, Primates) inferred from mt COI and COII gene sequences. *Primates; journal of primatology* **56**(2): 145-161.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**(12): 5463-5467.

Smit AFA, Hubley, R., and Green, P. 1996-2010. RepeatMasker Open-3.0. http://repeatmasker.org.

Steiper ME, Young NM. 2006. Primate molecular divergence dates. *Molecular phylogenetics and evolution* **41**(2): 384-394.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. *Nucleic acids research* **40**(15): e115.

Walker JA, Konkel MK, Ullmer B, Monceaux CP, Ryder OA, Hubley R, Smit AF, Batzer MA. 2012. Orangutan Alu quiescence reveals possible source element: support for ancient backseat drivers. *Mobile DNA* **3**: 8.

Wang J, Song L, Gonder MK, Azrak S, Ray DA, Batzer MA, Tishkoff SA, Liang P. 2006. Whole genome computational comparative genomics: A fruitful approach for ascertaining Alu insertion polymorphisms. *Gene* **365**: 11-20.

Wildman DE, Jameson NM, Opazo JC, Yi SV. 2009. A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). *Molecular phylogenetics and evolution* **53**(3): 694-702.

Appendix

Table 1. Oligonucleotide primer pairs designed for Platy-1 loci not shared between squirrel monkeys and common marmosets. The temperature listed is the annealing temperature used in this study.

Locus	Annealing Temp	Forward Sequence	Reverse Sequence
Ply4a-7	58	CCTCTCTCCTCCCTTTCCAA	GCAGCATTCCAAGTTCAAGA
Ply4a-8	58	GTGGCTTACATGAGTGGAAG	GCCAGTGCTTTGGTGTTGTA
Ply4a-10	58	CTTTCCCAGTCCTCTGGCTA	GTCCTGCCTTGGTCTTTCTG
Ply4a-12	58	CCCTGAGCACTTAAATTAGCC	GGGAGAACATGACCTAAGCAG
Ply4a-16	58	AGTTTGAGGCTGCAGTAGGC	CCAGGATGCCCATAATAAGC
Ply4a-17	61	TGAGCCCTGCTCAGTTCC	GGGAGCAGTGATTTCTGAGC
Ply4a-21	58	TCCGTTGAGATGTCTGTTGC	CGTTCCCAAACTTGAGAAAGAT
Ply4a-25	60	CCACCATGGCATGTGTGTAC	AACTGACTTGCTCGCTGTTT
Ply4a-27	58	GGGTCCATCAACTGGAAGAC	ATGCCGTTACTGCCACATCT
Ply4a-30	58	TATGACTGGGCCATGTTTCC	GGAGCTCATTCATGTCATTGC
Ply4a-34	56	GCTTTGCACCTCAGCTCCTA	AGCAATTTACAAAGGAATATATGCAG
Ply4a-37	58	CACCTGCTGCTTGGGTAGAT	GCTTAGTGTGATGCTTGGCATA
Ply4a-46	58	GTCAACTCAGAATTCTTATCCAGTG	ACCTTTGCATCTATTTACCTTCCTC
Ply4-1	58	AAGATGCTCAACTTCACTAACAGAAT	TCCTTTGCTGTTTATCAAGAACA
Ply5-12	56	AGGAGCAGTCAGCGTAAATG	CTATCATCAAATCCTGTCTCTGCT

Table 2. The New World Monkey V2 panel used in this study, including the species, common name, sample origin, and sample ID number for each DNA sample.

#	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	*Ateles geoffroyi*	black-handed spider monkey	Coriell	NG 05352
7	*Alouatta sara*	Bolivian red howler monkey	SDFZ	OR749
8	*Callithrix jacchus*	Common marmoset	NERPRC	cj393-99, A02-738
9	*Callithrix pygmea*	Pygmy marmoset	SDFZ	OR690
10	*Callimico goeldii*	Goeldi's marmoset	Alan Harris	955
11	*Saguinus labiatus*	red-chested mustached tamarin	Coriell	NG05308
12	*Saguinus fuscicollis nigrifrons*	Geoffroys saddle-back tamarin	SDFZ	OR621
13	*Cebus*	Capuchin monkey	KP	CA003
14	*Cebus*	Capuchin monkey	KP	30156
15	*Cebus*	Capuchin monkey	KP	30157
16	*Saimiri s. sciureus*	Squirrel monkey	SDFZ	KB4544
17	*Aotus trivirgatus*	Owl monkey	ATCC	CRL1556
18	*Pithecia p. pithecia*	Northern white-faced saki	SDFZ	OR842
19	*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 3. The Squirrel Monkey panel used in this study, including the species, common name, sample origin, and sample ID for each DNA sample.

#	Species	Common name	Origin	ID
1	*Home sapiens*	Human	ATCC	HeLa CCL-2
2	*Callithrix jacchus*	Common marmoset	NERPRC	cj393-99, A02-738
3	*Saimiri s. sciureus*	common Squirrel monkey	SDFZ	KB4544
4	*Saimiri sciureus*	common squirrel monkey	LSU Museum	LSUMZ M-7827
5	*Saimiri sciureus*	common squirrel monkey	Burke	UWBM# 75531
6	*Saimiri sciureus*	common squirrel monkey	Burke	UWBM# 75532
7	*Saimiri sciureus*	common squirrel monkey	Coriell	NG05311
8	*Saimiri sciureus*	common squirrel monkey	Kristof Zyskowski	YPM MAM 015317
9	*Saimiri sciureus*	common squirrel monkey	Kristof Zyskowski	YPM MAM 015340
10	*Saimiri sciureus*	common squirrel monkey	Chris C. Conroy	MVZ Mamm 193661
11	*Saimiri sciureus*	common squirrel monkey	Chris C. Conroy	MVZ Mamm 193685
12	*Saimiri sciureus*	common squirrel monkey	KCCMR	6118
13	*Saimiri sciureus*	common squirrel monkey	KCCMR	658
14	*Saimiri sciureus sciureus*	common squirrel monkey	SDZICR	KB18803
15	*Saimiri boliviensis*	Bolivian squirrel monkey	LSU Museum	LSUMZ M-4970
16	*Saimiri boliviensis*	Bolivian squirrel monkey	Chris C. Conroy	MVZ Mamm 196088
17	*Saimiri boliviensis*	Bolivian squirrel monkey	John A. Vanchiere	H604 Liver DNA
18	*Saimiri boliviensis*	Bolivian squirrel monkey	John A. Vanchiere	H804 Liver DNA
19	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2212
20	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2514
21	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2240
22	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2427
23	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	4248
24	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2573
25	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2269
26	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2560
27	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	2233
28	*Saimiri boliviensis*	Bolivian squirrel monkey	KCCMR	5161
29	*Saimiri boliviensis peruviensis*	Puruvian squirrel monkey	KCCMR	3526
30	*Saimiri boliviensis peruviensis*	Puruvian squirrel monkey	KCCMR	2291
31	*Saimiri boliviensis peruviensis*	Peruvian Black-Capped Squirrel Monkey	SDZICR	KB17911
32	*Saimiri oerstedii oerstedii*	Panamanian Red-Backed Squirell Monkey	SDZICR	KB7456
33	*Saimiri sciureus macrodon*	Ecuadorian Squirrel Monkey	SDZICR	KB17915
34	*Saimiri sp.*	squirrel monkey	Chris C. Conroy	MVZ Mamm 196089

ATCC: From cell lines provided by the American Type Culture Collection
NERPRC: New England Regional Primate Research Center
SDFZ: San Diego Frozen Zoo, Conservation and Research for Endangered Species (CRES)
LSUMZ: LSU Museum of Natural Science Collection of Genetic Resources, 119 Foster Hall, Baton Rouge, LA
Burke: The Burke Museum of Natural History and Culture, University of Washington, Seattle, WA
Coriell: Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ
YPM: Yale Peabody Museum of Natural History, Yale University, 170 Whitney Ave. New Haven, CT
MVZ: Museum of Vertebrate Zoology, University of California, Berkeley
JAV: John A. Vanchiere, M.D., Ph.D. Chief, Pediatric Infectious Diseases, LSUHSC-Shreveport, LA
KCCMR: Michale E. Keeling Center for Comparative Medicine and Research, The University of Texas MD Anderson Cancer Center, Bastrop, TX
SDZICR: San Diego Zoo Global Biomaterials Review Group, San Diego Zoo Institute for Conservation Research