2013

Alu elements and primate genomics

Adam Trent McLain
Louisiana State University and Agricultural and Mechanical College

Recommended Citation

McLain, Adam Trent, "Alu elements and primate genomics" (2013). *LSU Doctoral Dissertations*. 3287.
https://digitalcommons.lsu.edu/gradschool_dissertations/3287

ALU ELEMENTS AND PRIMATE GENOMICS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Adam Trent McLain
B.A., Virginia Polytechnic Institute and State University, 2002
M.A., Hunter College of the City University of New York, 2009
December 2013

ACKNOWLEDGEMENTS

I would like to take a little space here to thank all the people who have made this work not only possible, but worth doing. Dr. Mark Batzer welcomed me into his lab, took me under his wing, and taught me how to be a scientist. He has been the best advisor a person could hope for, and this dissertation wouldn't exist without his help. I also want to thank my Ph.D. committee for all their guidance, helpful advice, commentary, criticism, and questions. Dr. Christopher Austin, Dr. Robb Brumfield, Dr. Joomeyong Kim, and Dr. John Larkin, I couldn't have done it without your help.

All my collaborators deserve a thank-you as well. My research projects would not have come to fruition without you. So, a big shout-out to Josh Meyer, Chris Faulk, Michael Oldenberg, Matt Bourgeois, Camille Abshire, Scott Herke, Christian Roos, Mitchell Fullerton, Will Genstler, Thomas Beckstrom, Glenn Carmen, Erin Conlin, Pieter de Jong, Lucia Carbone, and the late Alan Mootnick.

From the Batzer Lab I would like to thank Jerilyn Walker and Miriam Konkel for their friendship and endless helpful advice about virtually everything, and George Cook for always making science something new and interesting.

I would like to thank the Louisiana Board of Regents and the Department of Biological Sciences for their financial investment in my training and my future, and all of the support this aid provided. Thanks to Prissy Milligan for all of her helpful advice.

I would also like to thank all my friends in Baton Rouge, you have helped it be a pleasure to live here. Thank you to my family for always supporting me regardless of the wild paths I choose to take. Lastly, special thanks and love to my wife Rebecca and to Freyja, Loki, Ella, and Persephone for keeping the home fires burning.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BLAT	Blast-Like Alignment Tool
EN	Endonuclease
HERV	Human Endogenous Retrovirus
L1	Long Interspersed Element-1
LTR	Long Terminal Repeat
ME	Mobile Element
NAHR	Non-Allelic Homologous Recombination
NHEJ	Non-Homologous End Joining
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
RM	RepeatMasker
RT	Reverse Transcriptase
SINE	Short Interspersed Element
SVA	SINE-r; VNTR; HERV-like region
TE	Transposable Element
TPRT	Target-Primed Reverse Transcription
TSD	Target Site Duplication
UTR	Untranslated Region
VNTR	Variable Number of Tandem Repeats

ABSTRACT

Alu elements are a primate-specific non-autonomous retrotransposon of ~300 base pairs (bp). A Short INterspersed Element (SINE), *Alu* elements are found in high copy number in genomes across the primate radiation. *Alu* elements have a history of effective use as a phylogenetic marker due to the presence-absence nature of their retrotransposition, and have been implicated in the cause of a number of genetic diseases. Various subfamilies of *Alu* elements have gone through periods of increased activity during the history of the primate radiation, with the overwhelming majority of elements classified as genomic "fossils" which no longer propagate but can serve as useful markers in elucidating phylogenies.

Lemurs (Infraorder: Lemuriformes) are a monophyletic radiation of strepsirrhine primates indigenous to the Island of Madagascar. The Lemuriformes are taxonomically subdivided into 5 families and >100 species, with numerous questions remaining to be answered about evolutionary relationships at the family, genus, and species levels. We computationally examined the grey mouse lemur (*Microcebus murinus*) genome for the presence of informative *Alu* elements and constructed a phylogenetic tree of 22 lemur species and 2 out-group species using the 138 loci retrieved.

Alu element activity level in primate genomes varies widely across species examined. The genome of the Western lowland gorilla (*Gorilla gorilla gorilla*) was sequenced in 2012 and made available for study. We examined this genome for the presence of *Alu*Y elements. *Alu*Y is the youngest *Alu* subfamily and the one most active in the hominid lineage. Our study elucidated 1,075 western lowland gorilla-specific *Alu*Y

elements classified into 10 subfamilies based on diagnostic mutations in the sequence of the element.

CHAPTER ONE:
BACKGROUND AND INTRODUCTION

Mobile Genetic Elements

Mobile genetic elements are sequences of non-coding, repetitive DNA that are capable of moving inter- and intra-chromosomally within genomes. Mobile elements have a long evolutionary history and have been enormously successful, at propagating themselves across host genomes (Deininger and Batzer 2002). Barbara McClintock first identified mobile elements while doing research on color variations in maize kernels during the 1940's. Mobile element movements in maize are associated with kernel color variation (McClintock 1950; McClintock 1956). This discovery garnered her the 1983 Nobel Prize in Physiology or Medicine, and has had tremendous implications for knowledge about the structure, function, and dynamic nature of genomes. Research into mobile elements now plays a role in disease research, phylogenetics, population genetics, and evolutionary biology. Mobile elements, despite being non-coding inhabitants of genomes, are recognized as having had a large impact on the evolutionary biology of many species (Batzer and Deininger 2002; Kazazian 2004; Cordaux and Batzer 2009; Konkel, Walker et al. 2010).

A minimum of 45% of the human genome, and possibly as much as 2/3, is composed of the various types of mobile DNA elements (Lander, Linton et al. 2001; Cordaux and Batzer 2009; de Koning, Gu et al. 2011). This is in stark contrast to the amount of actual coding sequence identified in the human genome, which is ~1.4% (Lander, Linton et al. 2001). Retrotransposons, mobile elements that reproduce by a copy-and-paste mechanism via an RNA intermediate, have been particularly active in mammals (Deininger and Batzer 2002). In humans, retrotransposons have increased in copy number to the point that they now comprise a minimum of ~42% of the genome, a feat made possible by their copy-and-paste nature (Cordaux and

Batzer 2009). New copies are made, and old copies are left behind, thus doubling each sequence every time a retrotransposition event occurs (Smit, Toth et al. 1995; Batzer and Deininger 2002; Cordaux and Batzer 2009).

Retrotransposons can be divided into two distinct types, autonomous and non-autonomous. Autonomous retrotransposons, as their name implies, encode their own enzymatic machinery, and thus have the capacity to carry out their own retrotransposition. Autonomous retrotransposons include long terminal repeat (LTR) and non-LTR elements. In the human genome, human endogenous retroviruses (HERVs) are an example of LTR elements, while long interspersed elements (LINEs) are the most common non-LTR elements and perhaps the most important LTR elements in the field of primate genomics (Cordaux and Batzer 2009). Non-autonomous retrotransposons do not encode their own retrotransposition machinery and must "hijack" the enzymatic machinery of autonomous retrotransposons to propagate in genomes. Non-autonomous elements in the human genome include short interspersed elements (SINEs), the most prominent of which are *Alu* element as well as the larger SVA retrotransposable element. SVA elements consist of a SINE-like region, a variable number of tandem repeats (VNTR) region, and a HERV-like region (Ostertag, Goodier et al. 2003; Wang, Xing et al. 2005). Both SVA elements and *Alu* elements are primate specific. *Alu*, a primate-specific SINE and the focus of this dissertation, is a particularly important non-autonomous retrotransposon in primate evolutionary history, and appears in high copy number in all surveyed primate genomes. The propagation of *Alu* is dependent on the parasitic relationship of that element with the autonomous retrotransposon LINE-1 (L1), as *Alu* hijacks the machinery of L1 to copy-and-paste itself within primate genomes (Batzer and Deininger 2002; Deininger and Batzer 2002; Konkel, Walker et al. 2010).

An understanding of L1 elements and their method of retrotransposition are crucial to an understanding of *Alu* elements. L1 is ~150 million years old, first arising very early in the mammalian lineage. L1s are common across mammalian genomes. In humans L1s make up ~17% of the genome, and ~512,000 copies are present (Lander, Linton et al. 2001; Brouha, Schustak et al. 2003; Hancks and Kazazian 2012; Rodic and Burns 2013). Unlike *Alu*, L1 is an autonomous retrotransposon and full-length L1 elements can make and insert copies in host genomes without requiring the enzymatic assistance of a second element. A full-length L1 is ~6 kilo bases (kb) in length, and is made up of an internal polymerase II promoter, two open reading frames (ORFs), 3' and 5' untranslated regions (UTRs), and a poly-A tail (Kazazian and Moran 1998; Deininger and Batzer 2002; Konkel, Walker et al. 2010). ORF2, which encodes for the endonuclease responsible for nicking the DNA target site where the new insertion will place itself, is of particular importance to this process (Mathias, Scott et al. 1991; Feng, Moran et al. 1996; Jurka 1997; Kolosha and Martin 1997).

Alu arose from the 7SL RNA sequence, the RNA component of the signal recognition particle ribonucleoprotein complex, ~65 million years ago. The ubiquitous distribution of *Alu* elements across the genomes of widely-diverged primate species and its absence in all other mammals, even close evolutionary relatives to the primate lineage, is indicative of its appearance at the very early stages of primate evolutionary history. A normal, full-length *Alu* element is ~300 base pairs in length and is dimeric in structure, with the left monomer comprised of intact A and B promoter boxes that encode an RNA polymerase III promoter, a central A-rich region between the two monomers, and a right monomer ending in a poly-A tail of varying length. Approximately 1.1 million *Alu* elements are extant in the human genome, comprising ~10-11% of the total size of the genome. Numerous distinct subfamilies of *Alu* elements exist, distinguished from one another by the presence of signature variations in their respective base pair composition. The three major *Alu* subfamilies are *Alu*J, *Alu*S, and *Alu*Y, with many derivatives of those subfamilies also identified. The putative age of these various subfamilies and their distributions across primate genomes play an important role in the use of *Alu* as a phylogenetic marker (Ullu and Tschudi 1984; Okada 1991; Batzer and Deininger 2002; Kriegs, Churakov et al. 2007; Konkel, Walker et al. 2010).

Alu elements have been used with great success in phylogenetic, population genetic, and disease research applications (Cordaux and Batzer 2009; Konkel, Walker et al. 2010). Because of the presence-absence nature of *Alu* insertions they are useful as phylogenetic markers; a shared marker indicates a common ancestry, and the ancestral state is almost always the absence of the element at the locus in question. Additionally, *Alu* elements are quick and inexpensive to assay using the polymerase chain reaction (PCR). These factors combine to make *Alu* SINEs a nearly perfect marker for phylogenetic and population genetic research in primates (Ray, Xing et al. 2006).

Target Primed Reverse Transcription

Using the L1 enzymatic machinery, both *Alu* and L1 elements propagate in primate genomes via a process termed target primed reverse transcription (TPRT). TPRT allows these elements to insert new copies of themselves in the host genome while the ancestral copies remain at the original loci (Luan, Korman et al. 1993; Cost, Feng et al. 2002). L1 elements encode for a protein called ORF2p via the ORF2 component of their sequence that is instrumental in both endonuclease and reverse transcriptase activities. The endonuclease cuts the bottom strand of DNA at a cleavage site with the sequence motif (3'-AA/TTTT-5') (Luan, Korman et al. 1993; Luan and Eickbush 1995; Cost, Feng et al. 2002). The poly-A tail of the retrotransposon mRNA then anneals to the -TTTT sequence. L1 reverse transcriptase then primes at the annealing site and reverse transcription occurs at the free 3' hydroxyl group present at the cleavage site. The top strand is then cleaved via mechanisms not entirely understood, the cDNA copy of the mRNA joins the bottom strand, and top strand synthesis occurs using the cDNA copy as a template (Fanning and Singer 1987; Luan and Eickbush 1995; Feng, Moran et al. 1996). The top strand cleavage occurs in a manner that the new insertion is flanked by target site duplications (TSDs), flanking DNA sequence of ~7-20 bp that matches that of the original insertion and created when the staggered nicks between top and bottom strands fill in (Szak, Pickeral et al. 2002).

The ultimate impact of TPRT is dependent upon where the new element inserts in the genome. Most new insertions occur in non-coding or non-regulatory sequence, due to the large percentage of the genome that falls into this category, and do not have a negative or positive impact on the host genome (Cordaux, Lee et al. 2006). When new insertions do occur in a regulatory region the results can be catastrophic. An element inserted into the promoter or regulatory region of a gene can have a negative impact on gene expression, and even if supressed by methylation can still have ancillary effects on the surrounding region of the genome because the methylation activity may silence transcriptional activity nearby (Slotkin and Martienssen 2007; Lees-Murdock and Walsh 2008; Hollister and Gaut 2009; Kano, Godoy et al. 2009; van der Heijden and Bortvin 2009). The effect of insertions in coding or regulatory regions in host genomes may be manifested in a variety of genetic diseases (Deininger and Batzer 1999; Boissinot, Entezam et al. 2001; Schmid 2003; Callinan and Batzer 2006; Hancks and Kazazian 2012). As mentioned previously, though the end result of TPRT is most often neutral, its occurrence may provide a wealth of information about the evolutionary history of a lineage for phylogenetic researchers (Deininger and Batzer 2002; Cordaux, Lee et al. 2006).

Alu Elements and Phylogenetics

The effective use of SINEs as phylogenetic markers was first demonstrated in 1993 in a study seeking to resolve relationships between Pacific salmonid species (Murata, Takasaki et al. 1993). Subsequent to this original study, SINE phylogenetics have been applied across a wide range of species to determine evolutionary relationships (Shedlock and Okada 2000; Shedlock, Takahashi et al. 2004). In particular, *Alu* elements have proven to be extremely useful tools for elucidating evolutionary relationships between primate species. The essentially homoplasy free presence of an *Alu* element of the same subfamily at a given locus between two or more primate species is almost always definitive evidence of shared ancestry between the species (Ray, Xing et al. 2006). The possibility of confounding events is very small, and easily resolved by the sequencing and examining of the element in question (Ray, Xing et al. 2006; Konkel, Walker et al. 2010). In the past 15 years *Alu*-based phylogenetic methods have been used with great success to resolve evolutionary relationships among the Tarsiers (Zietkiewicz, Richer et al. 1999; Schmitz, Ohme et al. 2001), New World (Ray, Xing et al. 2005) and Old World monkeys (Xing, Wang et al. 2005; Xing, Wang et al. 2007; Li, Han et al. 2009), gibbons (Meyer, McLain et al. 2012), and great apes (Salem, Ray et al. 2003). Applying this same methodology to the Malagasy primate infraorder Lemuriformes could thus be expected to yield similarly informative results.

The lemurs (Infraorder: Lemuriformes) are a monophyletic radiation of Strepsirrhine primates endemic to the island of Madagascar. Morphological and genetic evidence points to the descent of all extant lemur species from a single common ancestor that arrived on Madagascar ~55-60 million years ago, most likely via a rafting event across the Mozambique Channel from the eastern African mainland. In the ensuing millions of years lemurs have diversified to fill a wide variety of ecological and geographical niches (Yoder, Cartmill et al. 1996; Yoder and Yang

2004). Madagascar's own geographical diversity and isolation from the mainland propelled speciation, and today more than 100 species of lemur are recognized, often separated from their close evolutionary relatives by mountains or riverine boundaries. Madagascar, at only 4% of the world's landmass, contains ~15% of extant primate diversity (Mittermeier, Louis et al. 2010).

Lemurs are taxonomically divided into 5 families, 15 genera, and more than 100 species. The precise number of lemur species differs depending upon the source consulted (Mittermeier, Louis et al. 2010). The family Daubentoniidae is universally classified as the basal lineage of the Infraorder, but there is still disagreement regarding the relative placement of the other four families (Indriidae, Cheirogaleidae, Lepilemuridae, and Lemuridae) in relation to one another (Horvath, Weisrock et al. 2008; Orlando, Calvignac et al. 2008). Many lemur species are endangered. Some of these endangered lemur species are further classified as critically endangered. Consequently, establishing the phylogentic relationships between these animals may prove crucial to the justification of future efforts for their conservation.

In Chapter Two of this dissertation we discuss the use of *Alu* elements to construct a SINE based phylogeny of the Lemuriformes that seeks to resolve questions about species- and genera-level relationships. The sequenced, unassembled genome of the grey mouse lemur (*Microcebus murinus*) was downloaded from the website of the Broad Institute's '29 Mammals Project' and computationally analyzed for the presence of lemur-specific *Alu* elements (Smit, Hubley et al. 1996-2010; Kent 2002; Lindblad-Toh, Garber et al. 2011). Additionally, available sequenced DNA from other available lemur species was obtained from the NCBI database and subjected to the same computational analysis. Primers were constructed in the flanking regions of identified putative lemur-specific *Alu* elements, and these elements were then assayed via PCR to determine what, if any, species, genus, and/or family specificity they displayed

(Untergasser, Nijveen et al. 2007). Sufficient elements were recovered at nearly every node of the tree to meet the requirements of the Waddell test of statistical significance commonly used in *Alu*-based phylogenetic studies, and a robust phylogenetic tree was elucidated (Waddell, Kishino et al. 2001). A total of 138 lemur-specific *Alu* elements were assayed via PCR on a primate phylogenetic panel comprised of 22 lemur species from all five families, as well as two out-group species.

The resulting phylogenetic tree, generated using the Dollo parsimony method, reaffirmed Daubentoniidae as the basal lineage of the five lemur families. Additional loci were recovered that indicated a divergence between the remaining four families, with a Lepilemuridae-Cheirogaleidae grouping and an Indriidae-Lemuridae grouping. At the genus level we demonstrated strong support for a *Microcebus-Mirza* grouping as sister genera, and resolved the genus-level relationships in family Lemuridae, placing *Lemur* and *Hapalemur* together to the exclusion of *Eulemur* and establishing *Varecia* as the basal lineage within the family. Additionally, we discuss in detail how this phylogeny compares to earlier studies and the implications of this for lemur taxonomy and conservation (McLain, Meyer et al. 2012).

***Alu*Y Activity in the Western Lowland Gorilla Genome**

In Chapter Three of this dissertation we examine *Alu* subfamily activity in the western lowland gorilla genome. The genome of the Western lowland gorilla (*Gorilla gorilla gorilla*) was sequenced in 2012 at the Sanger Institute in an effort to gain a better understanding of one of mankind's closest relatives. The genome of "Kamilah", a female western lowland gorilla living at the San Diego Zoo, was initially assembled from 5.4 gigabase-pairs (Gbp) of capillary sequence and 166.8 Gbp of Illumina read pairs, and further refined using bacterial artificial chromosome (BAC) and fosmid end pair capillary technology (Scally, Dutheil et al. 2012). This data set was downloaded and analyzed for the presence of *Alu* elements using an in-house installation of the RepeatMasker program (Smit, Hubley et al. 1996-2010). The resulting data was parsed into separate files based on the *Alu* subfamily designations assigned by RepeatMasker (Smit, Hubley et al. 1996-2010). The focus of this study was potentially active *Alu* subfamilies in the gorilla genome, and *Alu*Y, as the youngest identified family of *Alu* elements, was targeted as the most likely to yield active subfamilies.

The file containing elements designated as members of the *Alu*Y subfamily was then further parsed to remove any elements under the length of 250bp using the estimation that shorter elements were likely to be older elements present in multiple species and therefore not useful for our analysis. The COSEG program, designed to identify repeat subfamilies using significant co-segregating mutations, was then run on the dataset to identify and group specific subfamilies together. COSEG ignores non-diagnostic mutations during analysis, providing an accurate representation of relationships between subfamilies of elements by ignoring potentially misleading mutational events (Price, Eskin et al. 2004). Subfamilies were defined as groups of >10 elements. A statistically significant subset of each putative subfamily was then subjected to wet-bench verification.

Loci singled out for verification were examined for further evidence of gorilla-specificity using the BLAST-Like Alignment Tool (BLAT) available at the UCSC Genome Browser website (http://genome.ucsc.edu) (Kent 2002). Putative gorilla-specific loci were compared to the available genomes of 5 other primate species, human (hg19), chimpanzee (panTro2), orangutan (ponAbe2), gibbon (GGSC Nleu 1.0/nomLeu1), and rhesus macaque (rheMac2). Elements found to be absent in these species and with sufficient orthologous flanking were marked for primer design.

The *Alu*Y subfamily and its derivatives were found to be active in the gorilla genome. A total of 1,075 western lowland gorilla specific *Alu*Y elements in 10 subfamilies were identified and computationally verified. A subset of additional, probable gorilla-specific *Alu*Y elements was also identified, but a lack of orthologous flanking material in other available primate genomes made assay and verification by PCR impossible and so these probable gorilla-specific elements were left out of the final analysis.

References

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.

Boissinot, S., A. Entezam, et al. (2001). "Selection against deleterious LINE-1-containing loci in the human lineage." Mol Biol Evol **18**(6): 926-935.

Brouha, B., J. Schustak, et al. (2003). "Hot L1s account for the bulk of retrotransposition in the human population." Proc Natl Acad Sci U S A **100**(9): 5280-5285.

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

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

Cordaux, R., J. Lee, et al. (2006). "Recently integrated Alu retrotransposons are essentially neutral residents of the human genome." Gene **373**: 138-144.

Cost, G. J., Q. Feng, et al. (2002). "Human L1 element target-primed reverse transcription in vitro." EMBO J **21**(21): 5899-5910.

de Koning, A. P., W. Gu, et al. (2011). "Repetitive elements may comprise over two-thirds of the human genome." PLoS Genet **7**(12): e1002384.

Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-193.

Deininger, P. L. and M. A. Batzer (2002). "Mammalian retroelements." Genome Res **12**(10): 1455-1465.

Fanning, T. G. and M. F. Singer (1987). "LINE-1: a mammalian transposable element." Biochim Biophys Acta **910**(3): 203-212.

Feng, Q., J. V. Moran, et al. (1996). "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell **87**(5): 905-916.

Hancks, D. C. and H. H. Kazazian, Jr. (2012). "Active human retrotransposons: variation and disease." Curr Opin Genet Dev **22**(3): 191-203.

Hollister, J. D. and B. S. Gaut (2009). "Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression." Genome Res **19**(8): 1419-1428.

Horvath, J. E., D. W. Weisrock, et al. (2008). "Development and application of a phylogenomic toolkit: resolving the evolutionary history of Madagascar's lemurs." Genome Res **18**(3): 489-499.

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

Kano, H., I. Godoy, et al. (2009). "L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism." Genes Dev **23**(11): 1303-1312.

Kazazian, H. H., Jr. (2004). "Mobile elements: drivers of genome evolution." Science **303**(5664): 1626-1632.

Kazazian, H. H., Jr. and J. V. Moran (1998). "The impact of L1 retrotransposons on the human genome." Nat Genet **19**(1): 19-24.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.

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 U S A **94**(19): 10155-10160.

Konkel, M. K., J. A. Walker, et al. (2010). "LINEs and SINEs of primate evolution." Evolutionary Anthropology **19**: 236-249.

Kriegs, J. O., G. Churakov, et al. (2007). "Evolutionary history of 7SL RNA-derived SINEs in Supraprimates." Trends Genet **23**(4): 158-161.

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

Lees-Murdock, D. J. and C. P. Walsh (2008). "DNA methylation reprogramming in the germ line." Epigenetics **3**(1): 5-13.

Li, J., K. Han, et al. (2009). "Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements." Gene **448**(2): 242-249.

Lindblad-Toh, K., M. Garber, et al. (2011). "A high-resolution map of human evolutionary constraint using 29 mammals." Nature **478**(7370): 476-482.

Luan, D. D. and T. H. Eickbush (1995). "RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element." Mol Cell Biol **15**(7): 3882-3891.

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.

Mathias, S. L., A. F. Scott, et al. (1991). "Reverse transcriptase encoded by a human transposable element." Science **254**(5039): 1808-1810.

McClintock, B. (1950). "The origin and behavior of mutable loci in maize." Proc Natl Acad Sci U S A **36**(6): 344-355.

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

McLain, A. T., T. J. Meyer, et al. (2012). "An alu-based phylogeny of lemurs (infraorder: Lemuriformes)." PLoS One **7**(8): e44035.

Meyer, T. J., A. T. McLain, et al. (2012). "An Alu-based phylogeny of gibbons (hylobatidae)." Mol Biol Evol **29**(11): 3441-3450.

Mittermeier, R. A., E. E. Louis, et al. (2010). Lemurs of Madagascar. Arlington, Virginia, Conservation International.

Murata, S., N. Takasaki, et al. (1993). "Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution." Proc Natl Acad Sci U S A **90**(15): 6995-6999.

Okada, N. (1991). "SINEs." Curr Opin Genet Dev **1**(4): 498-504.

Orlando, L., S. Calvignac, et al. (2008). "DNA from extinct giant lemurs links archaeolemurids to extant indriids." BMC Evol Biol **8**: 121.

Ostertag, E. M., J. L. Goodier, et al. (2003). "SVA elements are nonautonomous retrotransposons that cause disease in humans." Am J Hum Genet **73**(6): 1444-1451.

Price, A. L., E. Eskin, et al. (2004). "Whole-genome analysis of Alu repeat elements reveals complex evolutionary history." Genome Res **14**(11): 2245-2252.

Ray, D. A., J. Xing, et al. (2005). "Alu insertion loci and platyrrhine primate phylogeny." Mol Phylogenet Evol **35**(1): 117-126.

Ray, D. A., J. Xing, et al. (2006). "SINEs of a nearly perfect character." Syst Biol **55**(6): 928-935.

Rodic, N. and K. H. Burns (2013). "Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms?" PLoS Genet **9**(3): e1003402.

Salem, A. H., D. A. Ray, et al. (2003). "Alu elements and hominid phylogenetics." Proc Natl Acad Sci U S A **100**(22): 12787-12791.

Scally, A., J. Y. Dutheil, et al. (2012). "Insights into hominid evolution from the gorilla genome sequence." Nature **483**(7388): 169-175.

Schmid, C. W. (2003). "Alu: a parasite's parasite?" Nat Genet **35**(1): 15-16.

Schmitz, J., M. Ohme, et al. (2001). "SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates." Genetics **157**(2): 777-784.

Shedlock, A. M. and N. Okada (2000). "SINE insertions: powerful tools for molecular systematics." Bioessays **22**(2): 148-160.

Shedlock, A. M., K. Takahashi, et al. (2004). "SINEs of speciation: tracking lineages with retroposons." Trends Ecol Evol **19**(10): 545-553.

Slotkin, R. K. and R. Martienssen (2007). "Transposable elements and the epigenetic regulation of the genome." Nat Rev Genet **8**(4): 272-285.

Smit, A., R. Hubley, et al. (1996-2010). "RepeatMasker Open-3.0." from http://www.repeatmasker.org.

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

Szak, S. T., O. K. Pickeral, et al. (2002). "Molecular archeology of L1 insertions in the human genome." Genome Biol **3**(10): research0052.

Ullu, E. and C. Tschudi (1984). "Alu sequences are processed 7SL RNA genes." Nature **312**(5990): 171-172.

Untergasser, A., H. Nijveen, et al. (2007). "Primer3Plus, an enhanced web interface to Primer3." Nucleic Acids Res **35**(Web Server issue): W71-74.

van der Heijden, G. W. and A. Bortvin (2009). "Transient relaxation of transposon silencing at the onset of mammalian meiosis." Epigenetics **4**(2): 76-79.

Waddell, P. J., H. Kishino, et al. (2001). "A phylogenetic foundation for comparative mammalian genomics." Genome Inform **12**: 141-154.

Wang, H., J. Xing, et al. (2005). "SVA elements: a hominid-specific retroposon family." J Mol Biol **354**(4): 994-1007.

Xing, J., H. Wang, et al. (2005). "A mobile element based phylogeny of Old World monkeys." Mol Phylogenet Evol **37**(3): 872-880.

Xing, J., H. Wang, et al. (2007). "A mobile element-based evolutionary history of guenons (tribe Cercopithecini)." BMC Biol **5**: 5.

Yoder, A. D., M. Cartmill, et al. (1996). "Ancient single origin for Malagasy primates." Proc Natl Acad Sci U S A **93**(10): 5122-5126.

Yoder, A. D. and Z. Yang (2004). "Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context." Mol Ecol **13**(4): 757-773.

Zietkiewicz, E., C. Richer, et al. (1999). "Phylogenetic affinities of tarsier in the context of primate Alu repeats." Mol Phylogenet Evol **11**(1): 77-83.

CHAPTER TWO:
AN *ALU*-BASED PHYLOGENY OF LEMURS (INFRAORDER: LEMURIFORMES)*

Introduction

Lemurs (infraorder: Lemuriformes) are an ecologically and phenotypically diverse radiation of strepsirrhine primates endemic to the island of Madagascar. Varying in size from the tiny mouse lemurs (*Microcebus*), the smallest living primates at 23-29 cm, to the indris (*Indri*) at 64-72 cm, lemurs display stunning diversity in length, weight, diet, behavior, and pelage (Mittermeier, Louis et al. 2010). The prevailing genetic and morphological evidence supports the monophyletic descent of all extant and extinct lemur species from a common ancestor that arrived on Madagascar between 55-60 million years ago. The ancestor most likely dispersed via a rafting event across the Mozambique Channel from the African mainland (Yoder, Cartmill et al. 1996; Roos, Schmitz et al. 2004; Yoder and Yang 2004; Masters, de Wit et al. 2006; Tattersall 2007; Mittermeier, Louis et al. 2010). In the ensuing period lemurs have diversified to occupy a wide array of ecological niches across Madagascar. At present five families and approximately 100 species are recognized (the number of species differs depending upon the author(s) consulted), with dozens of new species discovered or elevated from subspecies status since the 1980s. In the first decade of the 21st century alone researchers described 41 new lemur species (Mittermeier, Louis et al. 2010). Concerns have been raised regarding overzealousness in the description of new species, but it cannot be disputed that lemurs are a much more speciose radiation than they were thought to be a few decades ago (Tattersall 2007).

A broad review of extant species in 2010 (Mittermeier, Louis et al. 2010) recognized 101 species of lemur in 15 genera grouped into five families: Cheirogaleidae, the mouse, dwarf, and fork-marked lemurs, with 31 species in five genera (*Allocebus*, *Cheirogaleus*, *Microcebus*,

*Reprinted by permission of *PLoS ONE*

Mirza, and *Phaner*); Daubentoniidae, the aye-aye, with one species in one genus (*Daubentonia*); Indriidae, the indris, sifakas, and wooly lemurs, with 19 species in three genera (*Avahi*, *Indri*, and *Propithecus*); Lemuridae, the ring-tailed, brown (or true), ruffled, and bamboo (or gentle) lemurs, with 24 species in five genera (*Eulemur*, *Hapalemur*, *Lemur*, *Prolemur*, and *Varecia*); and Lepilemuridae, the sportive lemurs, with 26 species in one genus (*Lepilemur*). There are, in addition, 17 known species of extinct lemurs classified into eight genera and three families. Many of these extinctions are recent and related to the arrival of humans on Madagascar ~2kya (Shapiro, Seiffert et al. 2005; Mittermeier, Louis et al. 2010).

Lemur phylogeny remains controversial at all taxonomic levels, and a number of studies have been conducted over the last four decades using various approaches, including morphology (Eaglen 1983; Groves and Eaglen 1988; Shapiro, Seiffert et al. 2005; Lebrun, de Leon et al. 2010; Mittermeier, Louis et al. 2010), karyotyping (Eckhardt 1970; Ishak, Warter et al. 1988), mitochondrial DNA analysis (Delpero, Masters et al. 2001), nuclear DNA (Pastorini, Thalmann et al. 2003; Roos, Schmitz et al. 2004; Horvath, Weisrock et al. 2008; Perelman, Johnson et al. 2011) combinations of mitochondrial and nuclear markers (Yoder and Yang 2004), combinations of molecular and morphological characters (Wyner, Desalle et al. 1999), retrotransposon analysis (Roos, Schmitz et al. 2004; Herke, Xing et al. 2007), and the genetic and morphological relationships of parasites across species (Bochkob, Klimov et al. 2011). The large body of literature on lemur phylogeny is in agreement on several key points, including the monophyly of the infraorder and the grouping of lemurs alongside the infraorder Lorisiformes within the strepsirrhine clade, the tooth-combed primates, as sister taxa to all other living primates (Horvath, Weisrock et al. 2008; Perelman, Johnson et al. 2011). Additionally, it is generally agreed that the family Daubentoniidae is the basal lineage within the lemuriforms, having been first to diverge from a shared common ancestor with the other four extant families (Karanth, Delefosse et al. 2005). Because of its extreme temporal divergence and phenotypic variation from other lemurs the family Daubentoniidae is sometimes separated in its own infraorder, Chiromyiformes (Groves 2001; Roos, Schmitz et al. 2004; Perelman, Johnson et al. 2011).

Outside of these points of agreement are the contentious relationships among the four remaining families. Two competing family-level phylogenies have emerged during the past decade. Both agree on the basal position of Daubentoniidae to the other four families. Horvath, Weisrock et al. (2008) place the Indriidae, Cheirogaleidae*,* and Lepilemuridae together, and all three as basal to Lemuridae, a position supported by Yoder (1997), Pastorini, Thalmann et al. (2003), and Yoder and Yang (2004). Alternatively, Orlando, Calvignac et al. (2008), in an analysis that also included subfossil lemurs, placed Indriidae and Lemuridae together to the exclusion of Cheirogaleidae and Lepilemuridae, a position supported by Delpero, Masters at al. (2001), Roos, Schmitz et al. (2004), and Bochkob, Klimov et al. (2011) (Orlando, Calvignac et al. 2008). These conflicting phylogenies rest primarily on the position of Indriidae, and whether this family represents a basal lineage to Lemuridae or is of more recent divergence, with Cheirogaleidae and Lepilemuridae basal to both. We use *Alu* elements, a family of primate-specific mobile elements, to resolve these conflicting phylogenetic analyses.

SINEs (Short INterspersed Elements) are a class of non-autonomous retrotransposons of <500 base pairs (bp) length that use RNA intermediaries to copy and insert themselves elsewhere within host genomes (Luan, Korman et al. 1993; Luan and Eickbush 1995; Cost, Feng et al. 2002; Konkel, Walker et al. 2010). SINEs are particularly useful genetic markers in the establishment of evolutionary relationships for several reasons. First, they are nearly-homoplasy-free markers (Batzer and Deininger 1991; Ray, Xing et al. 2006). The ancestral state is known to be the absence of the element, and each new element to arise is a distinct evolutionary event within a lineage. Thus, individuals sharing the same SINE at an orthologous locus are thought to be of common ancestry (Batzer and Deininger 1991; Murata, Takasaki et al. 1993; Batzer, Stoneking et al. 1994; Shedlock and Okada 2000; Okada, Shedlock et al. 2004; Ray, Xing et al. 2006; Ray 2007; Konkel, Walker et al. 2010). Second, once a SINE has inserted into a genome it is very rarely precisely excised. Thirdly, SINEs are relatively easy to evaluate using a locus specific PCR assay, making them potentially useful markers for conservationists (Batzer and Deininger 1991; Ray, Xing et al. 2006).

The use of SINEs as evolutionary and phylogenetic markers was first applied nearly two decades ago to resolve phylogenetic relationships among fish species (Murata, Takasaki et al. 1993). Since this early work the reliability of SINEs as phylogenetic markers has been well documented across many species, and the *Alu* family of primate-specific SINEs has been demonstrated to be particularly useful at elucidating phylogenetic relationships between primate species (Zietkiewicz, Richer et al. 1999; Schmitz, Ohme et al. 2001; Salem, Ray et al. 2003; Roos, Schmitz et al. 2004; Shedlock, Takahashi et al. 2004; Ray, Xing et al. 2005; Schmitz, Roos et al. 2005; Xing, Wang et al. 2005; Herke, Xing et al. 2007; Ray 2007; Xing, Wang et al. 2007; Li, Han et al. 2009; Osterholz, Walter et al. 2009; Roos, Zinner et al. 2011; Meyer, McLain et al. 2012).

Alu elements are a SINE of ~300bp length found only in primate genomes. Originally derived from 7SL RNA in a common ancestor of all living primates, *Alu* elements have propagated to the point where they comprise a significant component of primate genomes (Batzer and Deininger 2002; Kriegs, Churakov et al. 2007; Cordaux and Batzer 2009). *Alu* elements are classified into subfamilies, with *Alu*J being the oldest and therefore present in the genomes of all living primates (Jurka and Smith 1988; Churakov, Grundmann et al. 2010; Konkel, Walker et al. 2010). Younger lineage-specific subfamilies exist across the primate radiation, with some subfamilies presently active and others no longer producing new copies or subfamilies (Konkel, Walker et al. 2010). Liu, Alkan et al. (2009) assigned the subfamily designation *Alu*L to elements found in Lemuriformes (Liu, Alkan et al. 2009). Earlier studies have examined aspects of lemur phylogeny using SINEs. Roos, Schmitz et al. (2004) used a combination of SINE and mitochondrial markers to reconstruct a phylogeny of the strepsirrhine radiation, while Herke, Xing et al. (2007) examined relationships among some lemur species as part of a larger study involving the building of an *Alu*-based key for primate species identification. However, no exclusively *Alu*-based phylogeny focused on this infraorder has ever been reported. Here, using a combination of computational methods, PCR display methodology, and DNA sequencing, we use 138 *Alu* insertions specific to the Malagasy strepsirrhine lineage, including 22 loci previously reported by Herke, Xing et al. (2007) and 17 loci previously reported by Roos, Schmitz et al. (2004) to construct a phylogeny of the Lemuriformes.

Computational Methodology

Genomic sequence of the grey mouse lemur (*Microcebus murinus*; GenBank accession number ABDC01000000) generated at 1.9X coverage by the Broad Institute for the 29 Mammals Project was obtained in the form of 669,735 genomic contigs at 1.6 Gb total length (Lindblad-Toh, Garber et al. 2011). The majority of our PCR primers were designed using sequence from the *Microcebus murinus* genome. Additionally, genomic sequence was obtained from GenBank for the following species: *Lemur catta* (ring-tailed lemur), *Eulemur macaco* (black lemur), *Eulemur coronatus* (crowned lemur), *Propithecus coquereli* (Coquerel’s sifaka), *Daubentonia madagascariensis* (aye-aye), and *Cheirogaleus medius* (fat-tailed dwarf lemur). Sequences for these species were searched for putative lemur-specific *Alu* insertions based upon seven previously identified *Alu*L consensus sequences (Liu, Alkan et al., 2009) using an in-house installation of the RepeatMasker program with a custom library (Smit, Hubley et al. 1996-2004).

In-house Perl scripts were used to parse the RepeatMasker output for easier examination. Elements identified by RepeatMasker in *Microcebus murinus* genomic contigs as members of an *Alu*L subfamily and >280bp in length were compared to four non-lemuriform primate genomes, human (hg19), chimpanzee (panTro2), orangutan (ponAbe2), and rhesus macaque (rheMac2), using the BLAST-Like Alignment Tool (BLAT) available at http://genome.ucsc.edu (Kent 2002). Elements found to be absent in these outgroups, and that had orthologous flanking sequence that would allow primer design were marked for further examination. In the case of putative Lemuriformes-specific *Alu* elements obtained from other lemur species via GenBank, the Ensembl BLAT tool was used to compare the sequence to the genomic data of *Microcebus murinus* (Flicek, Amode et al. 2012). This was done to differentiate between *Alu* subfamilies distinct to particular lemuriform lineages and more effectively locate subfamilies specific to particular genera. The CLC Main Workbench v.5 software suite was used to align sequences and identify regions suitable for primer building (http://www.clcbio.com/index.php?id=92). Oligonucleotide primers for PCR assay were designed in the regions flanking the element using the Primer3Plus program (Untergasser, Nijveen et al. 2007). These primers were tested computationally against available primate genomes using the *in-silico* PCR tool on the UCSC Genome Bioinformatics website. Additionally, 22 primer pairs from Herke et al. (2007) and 17 primer pairs from Roos, Schmitz et al. (2004) were added to our analysis to provide additional resolution of phylogenetic relationships (Figure 2.1) among the Lemuriformes (Table S.1).

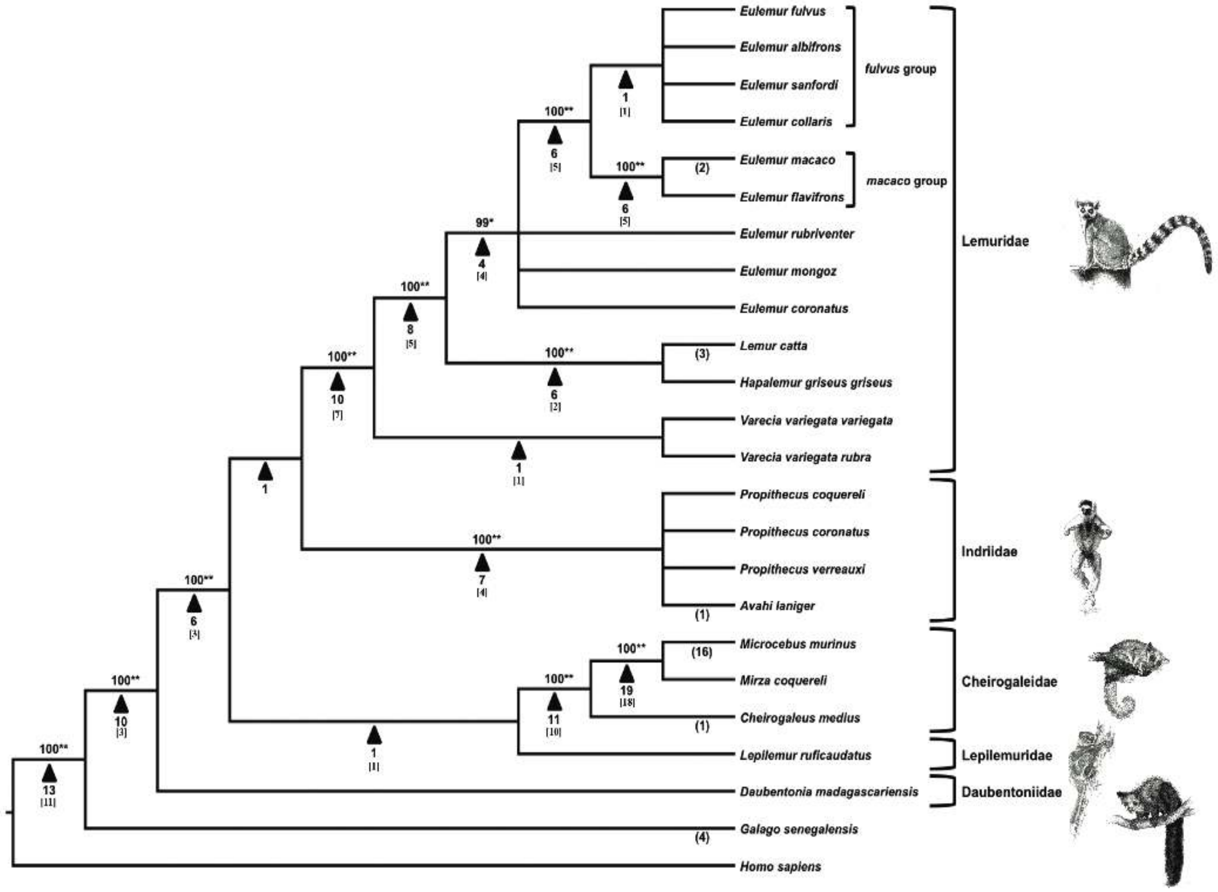

Figure 2.1: The most parsimonious tree generated from analysis of 138 *Alu* insertions in Lemuriformes. The amplification patterns of the *Alu* insertions were used to construct a Dollo parsimony tree of phylogenetic relationships with *G. senegalensis* and *H. sapiens* as outgroups using the MESQUITE and PAUP* programs. Numbers above branches are bootstrap values. The significance level of each node supported by insertions as determined by likelihood testing is indicated by either * ($p < 0.05$) or ** ($p < 0.01$). Numbers below arrows indicate the number of unambiguous loci supporting that node. Numbers in brackets below arrows indicate the number of loci at a given informative node identified by McLain et al. (2012). Numbers in parentheses represent insertions that are only present in one species or group. These insertions are not parsimony-informative. Consistency index (CI): 1.000; Homoplasy index (HI): 0.000; Retention index (RI): 1.000.

PCR and DNA sequencing

Primers were tested for amplification with lemur DNA templates corresponding to the species in which the primers were designed using a temperature gradient PCR (48°-65°C) to determine the proper annealing temperature for analysis of non-human samples. All loci were screened on a primate panel composed of *Homo sapiens* (HeLa) genomic DNA and samples from 23 strepsirrhine primates, including 22 lemur species and one out-group (non-Lemuriformes) strepsirrhine primate, *Galago senegalensis* (Senegal bushbaby) (Table S.1). For species with limited amounts of genomic DNA available, the samples were subjected to whole genome pre-amplification using the GenomiPhi genome amplification kit (Amersham, Sunnyvale, CA).

PCR amplification of each locus was performed in 25µl reactions using 15ng of template DNA, 200nM of each primer, 200µM dNTPs in 50mM KCl, 1.5mM $MgCl_2$, 10mM Tris-HCl (pH 8.4), and 2 units of *Taq* DNA polymerase. PCR reaction conditions were as follows: an initial denaturation at 95°C for 1 minute, followed by 32 cycles of denaturation at 95°C, annealing at the previously determined optimal annealing temperature, and extension at 72°C for 30 seconds each, followed by a final extension of 72°C for 1 minute. PCR products were analyzed on 2% agarose gels stained with 0.25ug ethidium bromide and visualized with UV fluorescence (Figure 2.2). A list of all loci, corresponding primer pairs, and optimal annealing temperatures for each are available as Table S.1 in the Supplemental Information for this study.

In the case of loci exhibiting relationships inconsistent with the most parsimonious tree, DNA cloning and sequencing was performed to verify the subfamily identity of the element in each species, in order to be certain that the relationships ascertained visually from PCR and gel electrophoresis did in fact share sequence identity. Gel slices were cut and purified using the Wizard SV Gel and PCR Clean-Up System (www.promega.com). Products from the individual PCRs were then cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA) and sequenced using chain termination sequencing on an ABI 3130 Genetic Analyzer (Sanger, Nicklen et al. 1992).

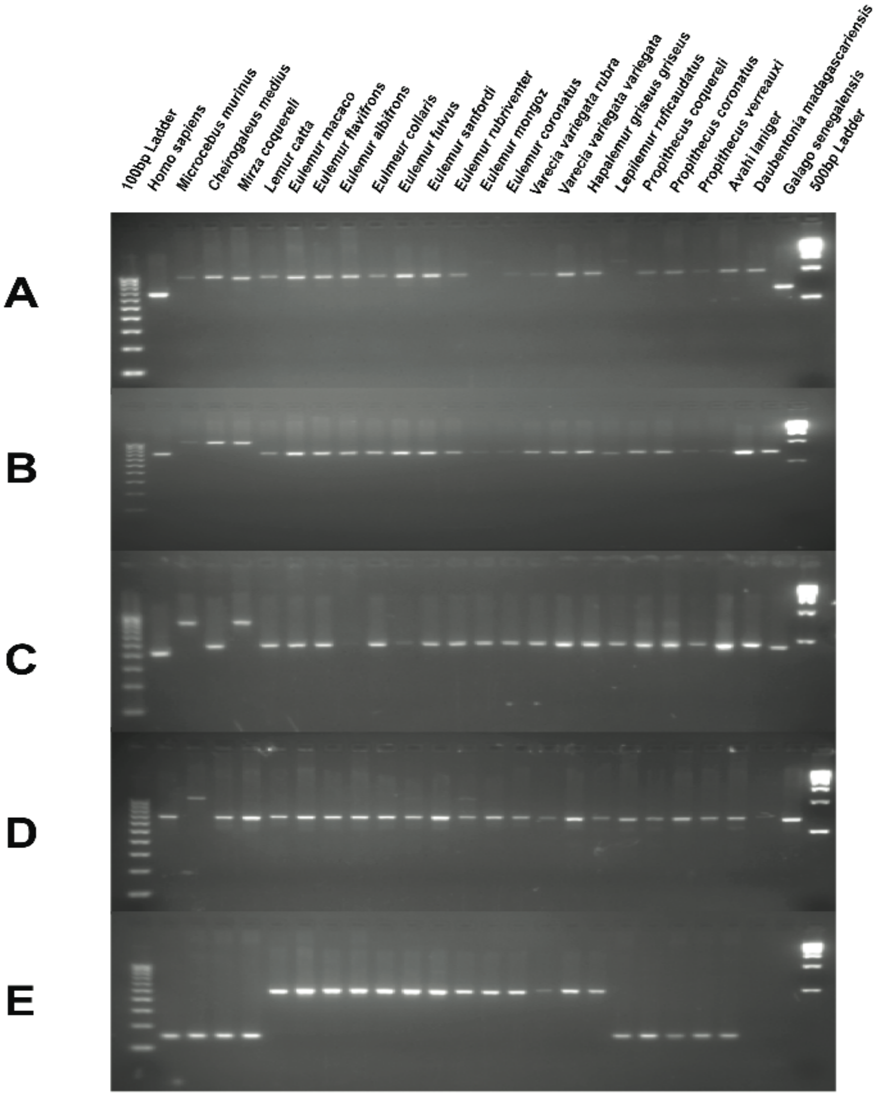

Figure 2.2: PCR amplification of polymorphic *Alu* insertions in Lemuriformes. Gel photographs displaying the methodology for establishing evolutionary relationships using *Alu* elements. The presence and absence of elements, supplemented by sequencing to eliminate the possibility of confounding events, is used to determine which species are more closely related. A total of 5 gel electrophoresis results on a 24-species primate panel are shown with *H. sapiens* and *G. senegalensis* as outgroups. **A:** Amplification of locus Str71B, an *Alu* insertion shared by the infraorder Lemuriformes. **B:** Amplification of locus MmA39, an *Alu* insertion shared by the family Cheirogaleidae. **C:** Amplification of locus MmA27, an *Alu* insertion shared by the sister genera *Microcebus* and *Mirza*. **D:** Amplification of locus Str59, an *Alu* insertion specific to the genus *Microcebus*. **E:** Amplification of locus Em6, an *Alu* insertion affirming the monophyly of the family Lemuridae to the exclusion of other lemur species and outgroups.

The resulting sequence data was then aligned and examined to determine if the locus in question was, in fact, the same element across multiple species, or a confounding event such as a parallel independent insertion or independent lineage sorting (Murata, Takasaki et al. 1993; Shedlock and Okada 2000; Ray, Xing et al. 2006).

Phylogenetic Analysis

A Dollo parsimony matrix was created using MESQUITE 2.75, a multi-faceted analysis program for applications in evolutionary biology (Maddison and Maddison 2011) (Table S.2). Dollo parsimony is particularly suited to *Alu*-based phylogenies because *Alu* elements are synapomorphic, presence-absence characters and nearly homoplasy-free genetic markers (Ray, Xing et al. 2005; Xing, Wang et al. 2005; Ray, Xing et al. 2006; Li, Han et al. 2009). All loci were set to Dollo.up for parsimony analysis. If an *Alu* insertion was found to be present via PCR assay it was coded as “1” for the given locus in the matrix. If the insertion was absent it was coded as “0”. Loci that could not be resolved for a given species at a certain locus were coded with a “?”. The PAUP* version 4.0b10 software was then used to perform a heuristic search on the data (Swofford 2003). A total of 10,000 bootstrap replicates were performed and a statistical test for evaluating SINE insertions based on a likelihood model was used to assess the statistical significance of each branch on the resulting tree (Waddell, Kishino et al. 2001). The tree was then visualized using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Computational Data Mining in the *Microcebus murinus* Genome

RepeatMasker identified a total of 294,218 putative *Alu* elements in the genome of *Microcebus murinus*, of which 229,774 were identified as belonging to the *Alu*L subfamily based on our consensus sequence. Another 16,224 *Alu*s were identified as members of the ancient *Alu*J subfamilies, with 3,169 identified as *Alu*Jb and 13,055 identified as *Alu*Jo, respectively.

Churakov et al. (2010) examined the *Alu*Jb subfamily in Strepsirrhines and found that it was likely inactive. Therefore the putative *Alu*Jbs identified in our RepeatMasker run were probably wrongly annotated by the program. Caution must be used when examining the total number of *Alu* elements identified, as the *Microcebus murinus* genome is unassembled and there are certainly contigs in our analysis that overlap. We guarded against examining the same locus twice by reviewing all loci in alignments to determine that they were, in fact, unique. Other elements were identified as belonging to various *Alu* subfamilies to which it is evolutionarily impossible for them to belong, such as *Alu*Y, a catarrhine-specific subfamily. These loci require additional examination for confirmation, but are likely members of Lemuriformes-specific subfamilies that are as-yet undocumented and were therefore not included in our custom RepeatMasker library. We focused on putative *Alu*L loci with sufficient orthologous flanking in available primate genomes in order to build the most optimal primers for elucidating phylogenetic relationships among lemur species. The large divergence time between the most recent common ancestor of the Lemuriformes and any other assembled primate genomes available for analysis necessitated the examination of a large number of potentially informative *Alu*L sites in order to construct functional primers. Sequence data generated from this project has been deposited in GenBank under the accession numbers (JX22863-JX228922; JX195193-JX195195; JX195187-JX195189). Sequence data for primers generated by Roos, Schmitz et al. (2004) is available from GenBank under the accession numbers (AY441478–AY441759). Sequence data for primers generated by Herke, Xing et al. (2007) is available from GenBank under the accession numbers (DQ822046–DQ822070 and DQ843660–DQ843663).

Phylogeny of the Lemuriformes

We identified 138 *Alu* insertion loci in multiple species of the five Lemuriformes families. 111 of these were phylogenetically informative, and 27 were autapomorphic insertions. The majority of primer design was completed in the available nuclear DNA sequence from the *Microcebus murinus* genome. In the case of PCR amplifications suggesting a branching pattern different from that of the majority of loci, sequencing universally revealed either a near-parallel independent insertion or another type of insertion/deletion event not affecting the topology of the tree. Loci that did not challenge the majority of loci at a node or did not display an incongruous pattern of relationships within the Lemuriformes were not subjected to sequencing. It is possible, but unlikely, that this had an effect of the topology of the tree. Most of the loci included in this study were identified computationally in lemur nuclear DNA sequence available via GenBank. The analysis of these loci resulted in a single most parsimonious tree (Figure 2.1; CI=1.000; HI=0.0000; RI=1.000). With the exception of two branches of our tree, discussed in more detail below, every clade was robustly supported (p-value < 0.05) under the maximum likelihood test developed in Waddell, Kishino, et al. (2001). The branches also had high levels of support based upon bootstrap analysis.

Debate over the phylogenetic relationships within Lemuriformes during the past decade has centered on the relationships of the four families Lemuridae, Lepilemuridae, Indriidae, and Cheirogaleidae to one another, with Daubentoniidae generally recognized as the basal lineage. The topology of our tree strongly supports the monophyly of the infraorder Lemuriformes, with 10 shared insertions recovered in support of this node. We recovered six loci placing Daubentoniidae as the most basal lineage among the five families. The position of Daubentoniidae is unsurprising in light of earlier studies, which established the aye-aye as basal to the other Lemuriformes (Poux and Douzery 2004; Roos, Schmitz et al. 2004; Yoder and Yang 2004; Horvath, Weisrock et al. 2008). The other four families segregate into a Lepilemuridae – Cheirogaleidae and an Indriidae – Lemuridae clade. Each of these clades is supported by a single insertion locus, MmM97 (GenBank accession numbers JX195194 and JX195195) and LI1 (GenBank accession numbers AY441631-AY441638), respectively. Cloning and sequencing of MmM97 confirmed that these two insertions are in fact the same element. LI1 was previously sequenced and confirmed by Roos et al. (2004) with identical results. Other studies have supported a grouping of Lepilemuridae and Cheirogaleidae as sister taxa (Horvath, Weisrock et al. 2008; Perelman, Johnson et al. 2011). The grouping of Lemuridae and Indriidae as sister taxa is also supported by previous studies (Yoder 1997; Delpero, Masters et al. 2001; Groves 2001; Roos, Schmitz et al. 2004).

Support for the monophyly of the Lemuridae was recovered from 10 insertions. *Lemur* and *Hapalemur* were determined to be sister taxa, a position supported by six loci. A further eight loci support a *Lemur*-*Hapalemur* clade as sister to *Eulemur*, with *Varecia* sister taxa to the other three genera. These findings support the established phylogeny of lemurid taxonomy (Roos, Schmitz et al. 2004; Horvath, Weisrock et al. 2008; Orlando, Calvignac et al. 2008; Mittermeier, Louis et al. 2010; Bochkob, Klimov et al. 2011; Perelman, Johnson et al. 2011), in particular, the taxonomic separation of *Eulemur* from *Lemur* (Simons and Rumpler 1988). We also recovered three loci unique to *Lemur catta*, supporting the established convention that *Lemur catta* is the sole species in the genus *Lemur* (Mittermeier, Louis et al. 2010; Perelman, Johnson et al. 2011).

In *Eulemur* we recovered strong support for the unity of the *macaco* group (*Eulemur macaco* and *Eulemur flavifrons*, formerly identified as subspecies of *Eulemur macaco* and now elevated to full species status) to the exclusion of other members of the genus. We were unable to further elucidate relationships among the other species definitively, particularly within the closely related and monophyletic *fulvus* group. Groves (2001) elevated the former subspecies in the *fulvus* group to full species status, a position supported by morphological and genetic evidence (Tattersall and Schwartz 1991; Djlelati, Brun et al. 1997; Pastorini, Forstner et al. 2000). The genus *Eulemur*, created by Simons and Rumpler (1988) to house the "true" lemurs when they were removed from *Lemur* after the species *Lemur catta* was designated the sole occupant of that genus, is believed to have diversified from a common ancestor over a relatively rapid span of time beginning ~8mya. This rapid speciation was possibly driven by a wetter climate and changing plant life in Madagascar (Cerling, Harris et al. 1997; Wells 2003; Horvath, Weisrock et al. 2008). The recent divergence times within this genus likely contributed to the difficulty of identifying species-specific *Alu* elements.

The relationships between the 12 *Eulemur* species currently recognized whether all of them should be accorded full species status or remain subspecies, has been debated at length (Yoder and Irwin 1999; Tattersall 2007; Mittermeier, Louis et al. 2010). Further confounding relationships and pointing to possibly overzealous species description in *Eulemur* are instances of observed hybridization between described species (Rumpler, Warter et al. 1985; Rabarivola, Meyers et al. 1991; Djlelati, Brun et al. 1997). We chose to follow Mittermeier, Louis et al. (2010) in recognizing each of the nine taxa available for our study (Table S.2) as full species, though we were unable to obtain lineage-specific *Alu* elements to support this hypothesis at the individual species level. We have opted to use the binomial identification for six of the species that were previously delineated as subspecies, for instance, *Eulemur albifrons* instead of *Eulemur fulvus albifrons*. A total of six shared insertion loci affirm the relationship between the two species of the *macaco* group and the four species of the *fulvus* group to the exclusion of the other three *Eulemur* species on our panel. An additional two loci were identified as being specific to *Eulemur macaco*, and may prove useful in the future as markers for species identification.

Within the family Cheirogalidae we recovered a strongly supported sister-group relationship between *Microcebus* and *Mirza,* with *Cheirogaleus* recovered as the basal lineage. A total of 19 loci supported the *Microcebus*-*Mirza* grouping to the exclusion of *Cheirogaleus*, which supports the findings of earlier phylogenetic studies (Yoder 1994; Roos, Schmitz et al. 2004; Rumpler 2004; Horvath, Weisrock et al. 2008; Perelman, Johnson et al. 2011). While we were only able to obtain samples from *Microcebus murinus* for our study, 18 species are currently recognized in *Microcebus* (Mittermeier, Louis et al. 2010). The 16 *Microcebus*-specific loci identified in this study might be useful in future analyses to clarify relationships within this speciose genus. One or more of the *Alu* elements we identified could certainly be polymorphic between species in this genus, something we were unable to clarify with only a single species on our panel.

In the Indriidae clade we recovered eight loci. Two of these loci, PcC1 and PcC2, were taken from nuclear DNA sequence available via GenBank and are present in all four Indriidae species represented in our dataset. Among the other six loci, three (MmA2c, MmA20A, and Str67A) were obtained from the sequencing of ambiguous loci and the remaining three were taken from Roos, Schmitz et al. (2004). Of these six loci, four are present in all four Indriidae species examined in our study. One locus, MmA2c, (GenBank accession number JX195193) is specific to *Avahi laniger*, the eastern wooly lemur.

While *Alu*-based phylogenies are generally reliable, confounding events can occur that result in incongruent tree topologies (Xing, Wang et al. 2005; Ray, Xing et al. 2006; Li, Han et al. 2009). In this case it is necessary to resolve relationships between species by DNA sequencing and comparative analysis of the element in question to establish the precise nature of a given locus. An example of a confounding event in the form of a parallel independent insertion is locus MmA20 (GenBank accession numbers JX195187-JX195189), which appeared to group the Cheirogaleidae with the Indriidae to the exclusion of the other species on our panel. This did not agree with the topology of our tree. Sequencing of this locus in both families demonstrated the presence of a near-parallel independent insertion event, with two *Alu* elements from independent subfamilies present at nearly the same location in the genome in the two different genera, that is, within the amplicon produced by the primers designed for this locus. MmA20 was then scored as Cheirogaleidae-specific, and MmA20A was scored as Indriidae-specific. Other loci found to contain parallel independent insertions include MmA2, M11, Str67A, and LcC2. Additionally, Ray, Xing et al. (2005) present an excellent illustrating of potentially confounding *Alu* insertion events in their study of platyrrhine primate phylogeny.

Conclusions

The robust phylogenetic relationships presented in this study support existing morphological and genetic research about relationships at the species and genus levels within the infraorder Lemuriformes. We offer support for a resolution of the previously unresolved relationships between the four families Lemuridae, Indriidae, Cheirogaleidae, and Lepilemuridae with a statistically robust tree (HI=0.000) demonstrating that Daubentoniidae is the basal lineage among Lemuriformes, with the common ancestor of the remaining families later separating into Lemuridae – Indriidae and Cheirogalidae – Lepilemuridae clades. Additionally, we largely resolve the branching patterns within the Cheirogaleidae and Lemuridae families. The methods used to examine these relationships further affirm the strengths of SINE-based phylogenetic studies. Given the known proliferation of *Alu* elements found in primate genomes during previous studies we expect that these primate synapomorphies will continue to be useful phylogenetic tools in the future.

References

Batzer, M. A. and P. L. Deininger (1991). "A human-specific subfamily of Alu sequences." Genomics **9**(3): 481-487.

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.

Batzer, M. A., M. Stoneking, et al. (1994). "African origin of human-specific polymorphic Alu insertions." Proc Natl Acad Sci U S A **91**(25): 12288-12292.

Bochkob, A. V., P. B. Klimov, et al. (2011). "Phylogeny and coevolutionary association of makialgine mites (Acari, Psoroptidae, Makialginae) provide insight into the evolutionary history of their hosts, strepsirrhine primates." Zool J Linn Soc **162**: 1-14.

Cerling, T. E., J. M. Harris, et al. (1997). "Global vegitation change throughout the Miocene/Pliocene boundry." Nature **389**: 153-158.

Churakov, G., N. Grundmann, et al. (2010). "A novel web-based TinT application and the chronology of the Primate Alu retroposon activity." BMC Evol Biol **10**: 376.

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

Cost, G. J., Q. Feng, et al. (2002). "Human L1 element target-primed reverse transcription in vitro." EMBO J **21**(21): 5899-5910.

Delpero, M., J. C. Masters, et al. (2001). "Phylogenetic relationships among the Malagasy lemuriformes (Primates: Strepsirrhini) as indicated by mitochondrial sequence data from the 12S rRNA gene." Zool J Linn Soc **133**: 83-103.

Djlelati, R., B. Brun, et al. (1997). "Meiotic study of hybrids in the genus Eulemur and taxonomic considerations." Am J Primatol **42**: 235-245.

Eaglen, R. (1983). "Lemur phylogeny and parsimony." Int J Primatology **4**(3): 249-273.

Eckhardt, R. (1970). "A Chromosome Arm Number Index and Its Application to the Phylogeny and Classification of Lemurs." Am J Phys Anthropol **31**: 85-88.

Flicek, P., M. R. Amode, et al. (2012). "Ensembl 2012." Nucleic Acids Res **40**(Database issue): D84-90.

Groves, C. P. (2001). Primate Taxonomy. Washington, D.C., Smithsonian Press.

Groves, C. P. and R. H. Eaglen (1988). "Systematics of the Lemuridae (Primates, Strepsirhini)." J Hum Evol **17**: 513-538.

Herke, S. W., J. Xing, et al. (2007). "A SINE-based dichotomous key for primate identification." Gene **390**(1-2): 39-51.

Horvath, J. E., D. W. Weisrock, et al. (2008). "Development and application of a phylogenomic toolkit: resolving the evolutionary history of Madagascar's lemurs." Genome Res **18**(3): 489-499.

Ishak, B., S. Warter, et al. (1988). "Phylogenetic Relation Between Lepilemuridae and Other Lemuriform Families." Am J Primatol **15**: 275-280.

Jurka, J. and T. Smith (1988). "A fundamental division in the Alu family of repeated sequences." Proc Natl Acad Sci U S A **85**(13): 4775-4778.

Karanth, K. P., T. Delefosse, et al. (2005). "Ancient DNA from giant extinct lemurs confirms single origin of Malagasy primates." Proc Natl Acad Sci U S A **102**(14): 5090-5095.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.

Konkel, M. K., J. A. Walker, et al. (2010). "LINEs and SINEs of primate evolution." Evolutionary Anthropology **19**: 236-249.

Kriegs, J. O., G. Churakov, et al. (2007). "Evolutionary history of 7SL RNA-derived SINEs in Supraprimates." Trends Genet **23**(4): 158-161.

Lebrun, R., M. P. de Leon, et al. (2010). "Deep evolutionary roots of strepsirrhine primate labyrinthine morphology." J Anat **216**(3): 368-380.

Li, J., K. Han, et al. (2009). "Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements." Gene **448**(2): 242-249.

Lindblad-Toh, K., M. Garber, et al. (2011). "A high-resolution map of human evolutionary constraint using 29 mammals." Nature **478**(7370): 476-482.

Liu, G. E., C. Alkan, et al. (2009). "Comparative analysis of Alu repeats in primate genomes." Genome Res **19**(5): 876-885.

Luan, D. D. and T. H. Eickbush (1995). "RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element." Mol Cell Biol **15**(7): 3882-3891.

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.

Maddison, W. P. and D. R. Maddison. (2011). "Mesquite: a modular system for evolutionary analysis Version 2.75." from http://mesquiteproject.org.

Masters, J. C., M. J. de Wit, et al. (2006). "Reconciling the origins of Africa, India and Madagascar with vertebrate dispersal scenarios." Folia Primatol (Basel) **77**(6): 399-418.

Meyer, T. J., A. T. McLain, et al. (2012). "An Alu-based phylogeny of gibbons (hylobatidae)." Mol Biol Evol **29**(11): 3441-3450.

Mittermeier, R. A., E. E. Louis, et al. (2010). Lemurs of Madagascar. Arlington, Virginia, Conservation International.

Murata, S., N. Takasaki, et al. (1993). "Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution." Proc Natl Acad Sci U S A **90**(15): 6995-6999.

Okada, N., A. M. Shedlock, et al. (2004). "Retroposon mapping in molecular systematics." Methods Mol Biol **260**: 189-226.

Orlando, L., S. Calvignac, et al. (2008). "DNA from extinct giant lemurs links archaeolemurids to extant indriids." BMC Evol Biol **8**: 121.

Osterholz, M., L. Walter, et al. (2009). "Retropositional events consolidate the branching order among New World monkey genera." Mol Phylogenet Evol **50**(3): 507-513.

Pastorini, J., M. R. Forstner, et al. (2000). "Relationships among brown lemurs (Eulemur fulvus) based on mitochondrial DNA sequences." Mol Phylogenet Evol **16**(3): 418-429.

Pastorini, J., U. Thalmann, et al. (2003). "A molecular approach to comparative phylogeography of extant Malagasy lemurs." Proc Natl Acad Sci U S A **100**(10): 5879-5884.

Perelman, P., W. E. Johnson, et al. (2011). "A molecular phylogeny of living primates." PLoS Genet **7**(3): e1001342.

Poux, C. and E. J. Douzery (2004). "Primate phylogeny, evolutionary rate variations, and divergence times: a contribution from the nuclear gene IRBP." Am J Phys Anthropol **124**(1): 1-16.

Rabarivola, C., D. Meyers, et al. (1991). "Distribution and Morphological Characters of Intermediate Forms Between the Black Lemur (Eulemur macaco macaco) and the Sclater's Lemur (E.m. flavifrons)." Primates **32**(2): 269-273.

Ray, D. A. (2007). "SINEs of progress: Mobile element applications to molecular ecology." Mol Ecol **16**(1): 19-33.

Ray, D. A., J. Xing, et al. (2005). "Alu insertion loci and platyrrhine primate phylogeny." Mol Phylogenet Evol **35**(1): 117-126.

Ray, D. A., J. Xing, et al. (2006). "SINEs of a nearly perfect character." Syst Biol **55**(6): 928-935.

Roos, C., J. Schmitz, et al. (2004). "Primate jumping genes elucidate strepsirrhine phylogeny." Proc Natl Acad Sci U S A **101**(29): 10650-10654.

Roos, C., D. Zinner, et al. (2011). "Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys." BMC Evol Biol **11**: 77.

Rumpler, Y. (2004). "Complementary approaches of cytogenetics and molecular biology to the taxonomy and study of speciation processes in lemurs." Evol Anthropol **13**(67-78).

Rumpler, Y., S. Warter, et al. (1985). "Cytogenetic study of complex hybrids in the genus Lemur (Primates, Prosimians)." Folia Primatol (Basel) **44**: 108-116.

Salem, A. H., D. A. Ray, et al. (2003). "Alu elements and hominid phylogenetics." Proc Natl Acad Sci U S A **100**(22): 12787-12791.

Sanger, F., S. Nicklen, et al. (1992). "DNA sequencing with chain-terminating inhibitors. 1977." Biotechnology **24**: 104-108.

Schmitz, J., M. Ohme, et al. (2001). "SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates." Genetics **157**(2): 777-784.

Schmitz, J., C. Roos, et al. (2005). "Primate phylogeny: molecular evidence from retroposons." Cytogenet Genome Res **108**(1-3): 26-37.

Shapiro, L. J., C. V. Seiffert, et al. (2005). "Morphometric analysis of lumbar vertebrae in extinct Malagasy strepsirrhines." Am J Phys Anthropol **128**(4): 823-839.

Shedlock, A. M. and N. Okada (2000). "SINE insertions: powerful tools for molecular systematics." Bioessays **22**(2): 148-160.

Shedlock, A. M., K. Takahashi, et al. (2004). "SINEs of speciation: tracking lineages with retroposons." Trends Ecol Evol **19**(10): 545-553.

Simons, E. L. and Y. Rumpler (1988). "Eulemur: new generic name for species of Lemur other than Lemur catta." CR Acad Sci Paris **3**(307): 547-551.

Smit, A., R. Hubley, et al. (1996-2004). "RepeatMasker Open-3.0." from http://www.repeatmasker.org.

Swofford, D. L. (2003). PAUP: Phylogenetic analysis using parsimony. Sunderland, MA, Sinauer Associates, Inc.

Tattersall, I. (2007). "Madagascar's Lemurs: Cryptic diversity or taxanomic inflation?" Evol. Anthropol. **16**(1): 12-23.

Tattersall, I. (2007). Origins of the Malagasy Strepsirrhine Primates. Lemurs: Ecology and Adaptation. New York, NY, Springer Publishing**:** 3-17.

Tattersall, I. and J. H. Schwartz (1991). "Phylogeny and nomenclature in the "lemur-group" of Malagasy strepsirhine primates." Anthrop Pap Am Mus Nat Hist **69**: 1-18.

Untergasser, A., H. Nijveen, et al. (2007). "Primer3Plus, an enhanced web interface to Primer3." Nucleic Acids Res **35**(Web Server issue): W71-74.

Waddell, P. J., H. Kishino, et al. (2001). "A phylogenetic foundation for comparative mammalian genomics." Genome Inform **12**: 141-154.

Wells, N. A. (2003). Some hypotheses on the Mesozoic and Cenozoic paleoenvironmental history of Madagascar. The Natural History of Madagascar. S. M. Goodman and J. P. Benstead. Chicago, IL, University of Chicago Press**:** 16-74.

Wyner, Y. M., R. Desalle, et al. (1999). "Phylogeny and Character Behavior in Family Lemuridae." Mol Phylogenet Evol **15**(1): 124-134.

Xing, J., H. Wang, et al. (2005). "A mobile element based phylogeny of Old World monkeys." Mol Phylogenet Evol **37**(3): 872-880.

Xing, J., H. Wang, et al. (2007). "A mobile element-based evolutionary history of guenons (tribe Cercopithecini)." BMC Biol **5**: 5.

Yoder, A. D. (1994). "Relative Position of the Cheirogaleidae in Strepsirhine Phylogeny: A Comparison of Morphological and Molecular Methods and Results." Am J Phys Anthropol **94**: 25-46.

Yoder, A. D. (1997). "Back to the Future: A synthesis of Strepsirrhine systematics." Evol Anthropol **6**: 11-22.

Yoder, A. D., M. Cartmill, et al. (1996). "Ancient single origin for Malagasy primates." Proc Natl Acad Sci U S A **93**(10): 5122-5126.

Yoder, A. D. and I. Irwin (1999). "Eulemur is a phylogenetically defensible taxon (but species relationships within the genus are problematic)." Lemur News **4**: 31-32.

Yoder, A. D. and Z. Yang (2004). "Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context." Mol Ecol **13**(4): 757-773.

Zietkiewicz, E., C. Richer, et al. (1999). "Phylogenetic affinities of tarsier in the context of primate Alu repeats." Mol Phylogenet Evol **11**(1): 77-83.

CHAPTER THREE:
*ALU*Y SUBFAMILY ACTIVITY IN THE WESTERN LOWLAND GORILLA GENOME*

Introduction

Alu elements are a family of primate-specific SINEs (**S**hort **IN**terspersed **E**lements) of ~300 base pairs (bp) long and present in the genomes of all living primates (Batzer and Deininger 2002; Cordaux and Batzer 2009; Konkel, Walker et al. 2010). *Alu* elements were derived from 7SL RNA, the RNA component of the signal recognition particle, in the common ancestor of all living primates (Ullu and Tschudi 1984). In the past ~65 million years *Alu* elements have become widely distributed in primate genomes (Roy-Engel 2008; Konkel, Walker et al. 2010). *Alu* elements are now present at copy numbers of >1,000,000 in all surveyed great ape genomes (Konkel, Walker et al. 2010) (Table S.1). Despite their high copy number the majority of *Alu* elements are genomic fossils, non-propagating relics passed down over millions of years after earlier periods of replicative activity (Cordaux, Lee et al. 2006; Konkel, Walker et al. 2010). A relatively small number of "master" elements are responsible for the continued spread of all active subfamilies (Deininger, Batzer et al. 1992; Han, Xing et al. 2005).

As non-autonomous retrotransposons, *Alu* elements do not encode the enzymatic machinery necessary for self-propagation (Batzer and Deininger 2002; Konkel, Walker et al. 2010). Instead, "master" *Alu* elements (Deininger, Batzer et al. 1992) insert new copies of themselves into host genomes. This is accomplished by appropriating the replication machinery (Batzer and Deininger 2002; Dewannieux, Esnault et al. 2003) of a much larger, autonomous retrotransposon called LINE1 (L1) via a process termed target-primed reverse transcription (TPRT) (Luan, Korman et al. 1993; Luan and Eickbush 1995; Feng, Moran et al. 1996; Cost, Feng et al. 2002).

*Submitted as a research article to *Mobile DNA*

The effective use of SINEs as phylogenetic markers was first demonstrated in 1993 in a study seeking to resolve relationships among Pacific salmonid species (Murata, Takasaki et al. 1993). Subsequent to this study, SINE-based phylogenetic methods have been applied across a wide range of species to determine evolutionary relationships (Shedlock and Okada 2000; Shedlock, Takahashi et al. 2004). In particular, *Alu* elements have proven to be extremely useful tools for elucidating evolutionary relationships between primate species (Minghetti and Dugaiczyk 1993; Konkel, Walker et al. 2010). The essentially homoplasy free presence of an *Alu* element of the same subfamily at a given locus between two or more primate species is almost always definitive evidence of shared ancestry (Ray, Xing et al. 2006). The possibility of confounding events is very small, and easily resolved by the sequencing and examining of the element in question (Ray, Xing et al. 2006; Konkel, Walker et al. 2010). In the past 15 years *Alu*-based phylogenetic methods have been used with great success to resolve evolutionary relationships among the Tarsiers (Zietkiewicz, Richer et al. 1999; Schmitz, Ohme et al. 2001), New World (Ray, Xing et al. 2005) and Old World monkeys (Xing, Wang et al. 2005; Xing, Wang et al. 2007; Li, Han et al. 2009), gibbons (Meyer, McLain et al. 2012), lemurs (Roos, Schmitz et al. 2004; McLain, Meyer et al. 2012), and great apes (Salem, Ray et al. 2003).

In addition to phylogenetic applications *Alu* elements also function as effective markers for the study of population genetics via examination of polymorphic elements between members of the same species (Perna, Batzer et al. 1992; Batzer, Stoneking et al. 1994; Batzer and Deininger 2002). *Alu* elements are also linked to numerous genetic diseases, and the insertion of an element at an importune genomic location can have grave consequences for the individual involved (Deininger and Batzer 1999; Cordaux and Batzer 2009; Hancks and Kazazian 2012).

Additionally, *Alu* elements are thought to be a causal factor in genomic instability (Hedges and Deininger 2007; Konkel and Batzer 2010; Cook, Konkel et al. 2011; Cook, Konkel et al. 2013).

Alu elements are classified in multiple major subfamilies and numerous smaller, derivative subfamilies based on specific sequence mutations (Slagel, Flemington et al. 1987; Willard, Nguyen et al. 1987; Britten, Baron et al. 1988; Jurka and Smith 1988). All extant primates share older elements, while all primate lineages examined also have younger, lineage-specific subfamilies (Batzer, Deininger et al. 1996). *Alu* subfamily evolution is parallel, not linear, and various subfamilies have been found to be actively retrotransposing at the same time in all primate genomes surveyed; each primate lineage thus possesses its own *Alu* subfamilies (Cordaux, Hedges et al. 2004; Price, Eskin et al. 2004; Konkel, Walker et al. 2010).

The *Alu*J subfamily is the most ancient *Alu* lineage, and was largely active from ~65 million years ago (mya) to ~55 mya, at which point *Alu*S evolved and supplanted *Alu*J as the predominant active subfamily (Jurka and Smith 1988; Batzer, Deininger et al. 1996). Due to the antiquity of the lineage, *Alu*J subfamilies are present in all extant primates, including Strepsirrhines (Roos, Schmitz et al. 2004; Liu, Alkan et al. 2009). *Alu*S, on the other hand, evolved from *Alu*J after the Strepsirrhine-Haplorrhine divergence, and so is only found in New World and Old World primates (Jurka and Smith 1988; Kapitonov and Jurka 1996; Batzer and Deininger 2002). The *Alu*Y subfamily subsequently evolved from *Alu*S in the Old World primate lineage, and remains the predominant active subfamily in catarrhines (Batzer, Deininger et al. 1996; Kapitonov and Jurka 1996; Konkel, Walker et al. 2010).

A number of *Alu*Y-derived subfamilies continue to be active in great apes (Konkel, Walker et al. 2010), and polymorphic lineage-specific *Alu* elements have been well documented between existing human populations (Batzer and Deininger 2002), indicating a continued activity level for these mobile elements. A rate of 1 new element in every ~20 live births has been proposed as the current rate of *Alu* element activity in the extant human population, but the large size of this population coupled with human generation time would make it very difficult for new elements to come to fixation outside of small population groups (Cordaux, Hedges et al. 2006; Xing, Zhang et al. 2009). Research into *Alu* element activity in Sumatran and Bornean orangutans has indicated a comparatively low-level of continued retrotransposition activity in these apes (Locke, Hillier et al. 2011), suggesting some alteration of the propagation of *Alu* within this lineage (Walker, Konkel et al. 2012).

The western lowland gorilla (*Gorilla gorilla gorilla*), a subspecies of the western gorilla (*Gorilla gorilla*), is a critically endangered great ape endemic to the forests and lowland swamps of central Africa (Fleagle 1999; Strier 2007). Western lowland gorillas are gregarious, living in family groups comprised of a dominant male, multiple females, subadult males, and juvenile offspring (Fleagle, Janson et al. 1999). Western lowland gorillas are in danger of extinction due to human activity. Their wild population size is shrinking in the face of anthropogenic pressure and diseases such as Ebola (Strier 2007). Gorillas are a close evolutionary relative of humans and the *Pan* lineage of chimpanzees and bonobos, with the most widely accepted date for a common ancestor 6-9 mya (Chen and Li 2001; Glazko and Nei 2003; Salem, Ray et al. 2003; Steiper and Young 2006), though a date as early as 10 mya has been recently proposed (Langergraber, Prufer et al. 2012).

The genome of "Kamilah", a female western lowland gorilla living at the San Diego Zoo, was initially assembled from 5.4 Gbp of capillary sequence and 166.8 Gbp of Illumina read pairs, and further refined using bacterial artificial chromosome (BAC) and fosmid end pair capillary technology (Scally, Dutheil et al. 2012). This sequence is available from the Wellcome Trust-Sanger Institute.

Previous analyses of *Alu* elements in gorillas have been limited to analysis in the context of wider research projects (Salem, Ray et al. 2003; Sen, Han et al. 2006; Lee, Han et al. 2008; Ventura, Catacchio et al. 2011; Hormozdiari, Konkel et al. 2013) and have not focused specifically on subfamily analysis. Here we examine the western lowland gorilla genome (build gorGor3.1) (Scally, Dutheil et al. 2012) to identify gorilla-specific *Alu*Y subfamilies and assess the activity levels, copy number, and age of these subfamilies.

Results and discussion

Computational examination of the western lowland gorilla genome

A total of 1,085,174 *Alu* elements were identified in the genome of the western lowland gorilla (Appendix A; Table S.3). Of these, 286,801 were identified as belonging to the ancient *Alu*J subfamily, and 599,237 were identified as members of the *Alu*S subfamily. 57,427 elements were too degraded or truncated to be assigned a subfamily designation by RepeatMasker, and were simply identified as "*Alu*". We identified 141,709 members of the *Alu*Y subfamily. This subfamily is of particular interest due to its relatively young age and known continued mobility in other great ape genomes (Smit, Hubley et al. 1996-2010; Konkel, Walker et al. 2010). Approximately one-third (57,458) of these putative *Alu*Y elements were >250bp in length. Gorilla-specific elements were subsequently identified by comparison of orthologous loci in the genomes of human, common chimpanzee, and orangutan (Giardine, Riemer et al. 2005). Putative unique, gorilla-specific *Alu*Y insertions were estimated at 4,127 copies.

This number is similar (96.5%) to the 4,274 gorilla-specific *Alu* elements identified using other approaches (Ventura, Catacchio et al. 2011). Individual examination demonstrated that the majority of our 4,127 loci were in fact shared insertions. These loci were manually examined for gorilla specificity using BLAT (Kent 2002). This manual examination excluded 2,858 loci from further analysis due to the presence of shared insertions missed by Lift Over (2,626 insertions) or the lack of orthologous flanking regions in the genomes of other species that preclude PCR verification (232 insertions). This resulted in a total of 1,269 likely gorilla-specific *Alu* insertion loci for inclusion in subfamily structure analysis.

These 1,269 loci were analyzed for subfamily structure using the COSEG program. COSEG removed 194 probable gorilla-specific *Alu* insertions from the dataset due to the presence of truncations or deletions in diagnostic regions of the element, leaving 1,075 probable gorilla-specific *Alu* insertion loci for further analysis (Appendix A; Table S.5). COSEG then divided the loci into 10 subfamilies based on diagnostic mutations in the sequence of the individual *Alu* elements and provided subfamily consensus sequences (Price, Eskin et al. 2004). The consensus sequences were then aligned with known human *Alu*Y subfamilies from the RepBase database of repetitive elements (Jurka, Kapitonov et al. 2005). A gorilla-specific nomenclature system was created to designate subfamilies using the suffix "*Gorilla*" preceded by the subfamily affiliation based on a comparison to identified human subfamilies (e.g. "*Alu*Yc5a1_*Gorilla*"). Subfamilies were named in accordance with established practice for *Alu* subfamily nomenclature (Batzer, Deininger et al. 1996). The first identified *Alu*Yc5-derived subfamily was, for example, designated *Alu*Yc5a3_*Gorilla*. The "a" denotes the fact that this is the first Yc5-derived subfamily identified.

The "3" denotes the number of diagnostic mutations by which this gorilla-specific subfamily differs from the human *Alu*Yc5 consensus sequence (Batzer, Deininger et al. 1996). Subfamily age estimates were calculated using the BEAST program (Drummond and Rambaut 2007).

*Alu*Y subfamily activity in the western lowland gorilla genome

Computational and PCR analysis of the western lowland gorilla genome has identified 1,075 independent, gorilla-specific *Alu*Y insertion loci. Computational analysis of this dataset indicates the presence of 10 distinct subfamilies identifiable by the presence of diagnostic mutations specific to each lineage (Figure 3.1). The 1,075 elements identified in our study almost certainly do not represent the total number of *Alu*Y specific to western lowland gorilla genome. Any loci under our arbitrary length of >250 were excluded from our dataset. It is also likely that a number of *Alu*Y loci are located in portions of the genome where sequence data is incomplete; within repeat regions, for example. Additionally, some *Alu*Y loci were excluded when no orthologous genomic region was present in the species being used for comparison.

The largest newly identified gorilla-specific *Alu* subfamily was designated as *Alu*Y_*Gorilla*. This designation was established via computational evaluation and manual alignment of the 759 elements assigned to this subfamily. The consensus sequence for these elements was found to be 100% identical to the canonical *Alu*Y human consensus sequence (Figure 3.1). This subset of classic *Alu*Y elements continued to propagate in the *Gorilla* lineage after the divergence from the shared common ancestor with the *Homo*-*Pan* lineage. We assayed and verified a total of 135 loci from this subfamily via PCR (18%). The 43 elements belonging to the *Alu*Ya1_*Gorilla* subfamily differ from the *Alu*Y consensus sequence by one diagnostic mutation at nucleotide position 133. We assayed and verified via PCR 21 elements in this subfamily (49%). This sequence should not be confused with the *Homo*-*Pan Alu*Ya subfamily.

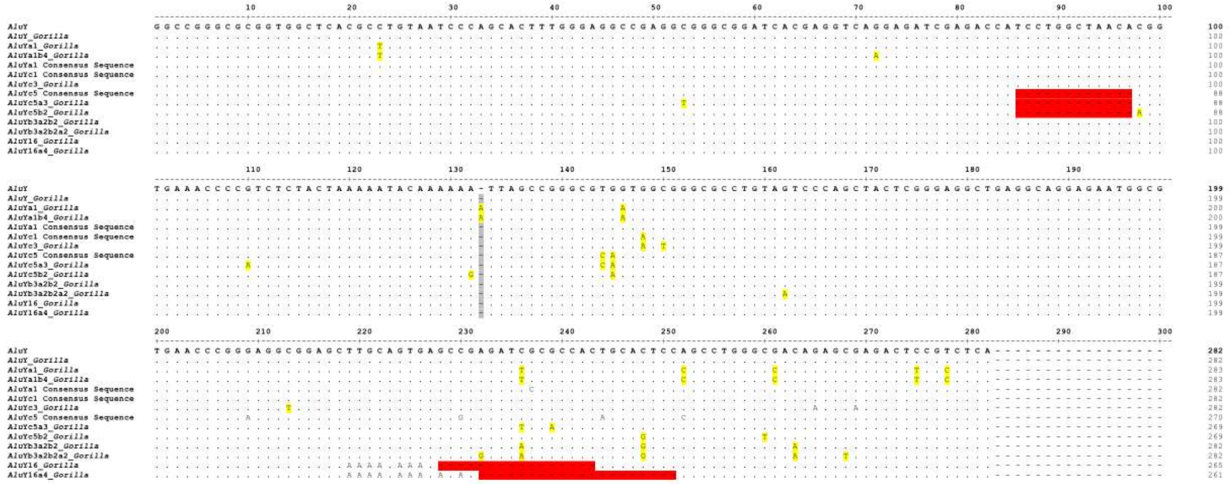

Figure 3.1: *Alu* sequence alignment. The consensus sequence for the *Alu*Y subfamily is shown at the top, with western lowland gorilla-specific *Alu* subfamilies listed below. The dots below the consensus denote the same base with insertions and deletions noted by dashes and mutations with the appropriate bases. The consensus sequences for the *Alu*Ya1, *Alu*Yc1, and *Alu*Yc5 subfamilies included for comparative purposes. Subfamily-specific diagnostic mutations are highlighted in yellow. Lineage-specific deletions are highlighted in red. *Alu*Y_*Gorilla* is 100% identical to the *Alu*Y consensus sequence. The shared 12-bp deletion identifying the *Alu*Yc5-derived *Gorilla* subfamilies is located at position 86. The 16-bp and 20-bp deletions identifying the *Alu*Y16_*Gorilla* and *Alu*Y16a4_*Gorilla* subfamilies are visible at position 228 and 232.

The *Alu*Ya1b4 subfamily is derived from *Alu*Ya1_*Gorilla* and is a small and very likely young subfamily of 13 elements that shared the diagnostic mutation at position 133 of Ya1 but has also accrued four additional diagnostic mutations. We assayed and verified via PCR seven elements in this subfamily (54%). A second identified *Alu*Y lineage in gorilla is the *Alu*Yc3_*Gorilla* subfamily. We assayed and verified via PCR 20 of the 69 elements in this subfamily (29%). The consensus sequence for the 69 members identified in this subfamily is a 100% match to the human *Alu*Yc3 subfamily consensus sequence (Figure 3.1).

Two additional gorilla-specific *Alu*Yc-derived subfamilies share the characteristic 12bp deletion at position 87-98 that is a hallmark of human *Alu*Yc5. These two subfamilies possess independent diagnostic mutations that make them distinct from the *Alu*Yc5 consensus sequence.

These two subfamilies are designated as *Alu*Yc5a3_*Gorilla* (55 elements identified) and *Alu*Yc5b2_*Gorilla* (46 elements identified*). Alu*Yc5a3_*Gorilla* has three additional diagnostic mutations differentiating it from the *Alu*Yc5 consensus as a mark of identification. In keeping with *Alu* subfamily naming convention this subfamily has thus been deemed "Yc5a3", "a" as the first Yc5-like subfamily identified in the gorilla genome and "3" for the three diagnostic mutations differentiating it from the canonical Yc5 consensus. We assayed and verified 27 members of this subfamily via PCR (49%). *Alu*Yc5b2 also shares the characteristic 12bp deletion of the human *Alu*Yc5, but has two independent diagnostic mutations (Figure 3). We assayed and verified via PCR 19 members of this subfamily (41%). It is probable that *Alu*Yc5a3_*Gorilla* and AluYc5b2_*Gorilla* derived from *Alu*Yc5 around the time of the *Homo*/*Pan*-*Gorilla* speciation event.

A third lineage nearly identical to human *Alu*Yb3a2 was identified as *Alu*Yb3a2b2_*Gorilla* (25 elements identified). This *Alu* subfamily contains two additional diagnostic mutations. Termed *Alu*Yb3a2b2_*Gorilla*, this lineage is an independent evolution in the *Gorilla gorilla gorilla* genome and not a derivative of the human-specific *Alu*Yb3a2. The *Alu*Yb lineage is human specific, meaning any identical or apparently derived *Alu* lineages in other primate genomes must be examples of independent evolution (Carter, Salem et al. 2004). This is confirmed by the lack of orthologs at the same location in the human genome. We assayed and verified 14 members of this subfamily via PCR (56%). An additional subfamily present at only 17 copies and derived from *Alu*Yb3a2b2_*Gorilla* was identified and termed *Alu*Yb3a2b2a2_*Gorilla*, due to two diagnostic mutations separating these otherwise identical subfamilies. We assayed and verified via PCR nine elements in this subfamily (53%). The low copy-number of these subfamilies coupled with their lack of impairing point mutations, even with the caveat that some subfamily members may have been overlooked, leads us to posit that they are among the youngest and potentially still active subfamilies in the western lowland gorilla genome.

Two additional subfamilies were identified that, while clearly *Alu*Y derived, do not follow the consensus sequences of established subfamilies available via RepBase. The first, termed *Alu*Y16_*Gorilla* is identified clearly by the presence of an A-rich insert at position 219 followed by a 16bp deletion, and is present in 30 copies. We assayed and verified via PCR 10 members of this subfamily (33%). The second subfamily, apparently derived from the first and designated *Alu*Y16a4_*Gorilla*, is present in 18 copies and can be distinguished from *Alu*Y16_*Gorilla* by a 20bp deletion occurring after the A-rich region at position 219. 17 elements from this subfamily were assayed via PCR (94%), with 100% of these 17 being verified as gorilla-specific. One locus (gorGor3.1 chrX:74544052-74544324) lacked sufficient orthologous 5' sequence in non-gorilla outgroups to successfully design a working primer, but was included in the total count based on computational verification. The accumulation of non-diagnostic mutations in these two subfamilies may indicate that they are more ancient.

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

One means of identifying potential master elements (Deininger, Batzer et al. 1992) is to look for subfamily members with mutation-free polyA-tails (Roy-Engel, Salem et al. 2002). A possible source element for the *Alu*Yc5a3 subfamily, for instance, was identified on chr2B:83106386-83106708, with a mutation-free 16bp polyA-tail. A possible source element for the *Alu*Y_*Gorilla* subfamily was identified on chrX:5135584-5135921, with a mutation-free 30bp polyA-tail and intact promoter region. A posited source element for the *Alu*Yc5b2 subfamily was identified on chr9:17925753-17926051, also with a mutation-free 30bp polyA-tail and intact promoter region.

*Alu*Y retrotransposition rates appear to be lower in the western lowland gorilla genome than in the human or chimpanzee genomes (Hedges, Callinan et al. 2004), while higher than that seen in the orangutan genome (Locke, Hillier et al. 2011; Walker, Konkel et al. 2012). Factors influencing rates of retrotransposition are myriad (Cordaux, Hedges et al. 2006; Konkel, Walker et al. 2010). Active retrotransposons are frequently polymorphic within a population, and are easily lost during events like speciation or population bottlenecks (Hedges and Batzer 2005; Belancio, Hedges et al. 2008). The number of active elements, and the amplification rate of elements surviving such an event, can vary greatly and impact overall retrotransposition activity in the host genome.

A possible explanation for this lower activity level include inhibition of retrotransposition in the *Gorilla* lineage by the interaction of host factors such as members of the APOBEC family of proteins with the enzymatic machinery of L1 (Schumann 2007; Konkel, Walker et al. 2010). Interference with L1 and *Alu* retrotransposition by APOBEC has been documented (Bogerd, Wiegand et al. 2006; Hulme, Bogerd et al. 2007; Schumann 2007). Additionally, environmental stress factors may impact retrotransposition rates (Farkash, Kao et al. 2006). It is possible that one or a combination of these retrotransposition-inhibiting factors could be responsible for the lower level of *Alu*Y activity in the western lowland gorilla genome.

A median joining tree of relationships between gorilla-specific *Alu*Y subfamilies was generated from a stepwise alignment (Edgar 2004) using the Network program (Figure 3.2) (Bandelt, Forster et al. 1999; Cordaux, Hedges et al. 2004). The tree generated supports the divergence of all gorilla-specific subfamilies from the *Alu*Y_*Gorilla* subfamily, and analysis of subfamily ages using BEAST places the date for this subfamily divergence at the stem of the *Gorilla* lineage. Initial divergence of gorilla-specific subfamilies occurred shortly after the speciation event separating the *Gorilla* lineage from the *Homo*-*Pan* lineage 6-9 mya (Chen and Li 2001; Glazko and Nei 2003; Salem, Ray et al. 2003; Steiper and Young 2006), and master elements have continued to produce copies of each subfamily at varying rates since (Deininger, Batzer et al. 1992).

Divergence dates of gorilla-specific *Alu*Y subfamilies

BEAST analysis of individual subfamily ages suggests no delay or change in transposon activity in western lowland gorilla following the divergence of the *Gorilla* and *Homo-Pan* lineages. The age of the gorilla-specific lineages ranges from 6.5-6.71 mya based on a baseline divergence of 7 mya for the most recent common ancestor of *Gorilla* and *Homo-Pan*. This indicates that all of the identified subfamilies originated around the time of the speciation event that separated these two lineages. This result is consistent with the ongoing propagation of these subfamilies before, during, and after the speciation event at a relatively constant rate. This indicates that all of the identified subfamilies originated around the time of the speciation event that separated these two lineages.

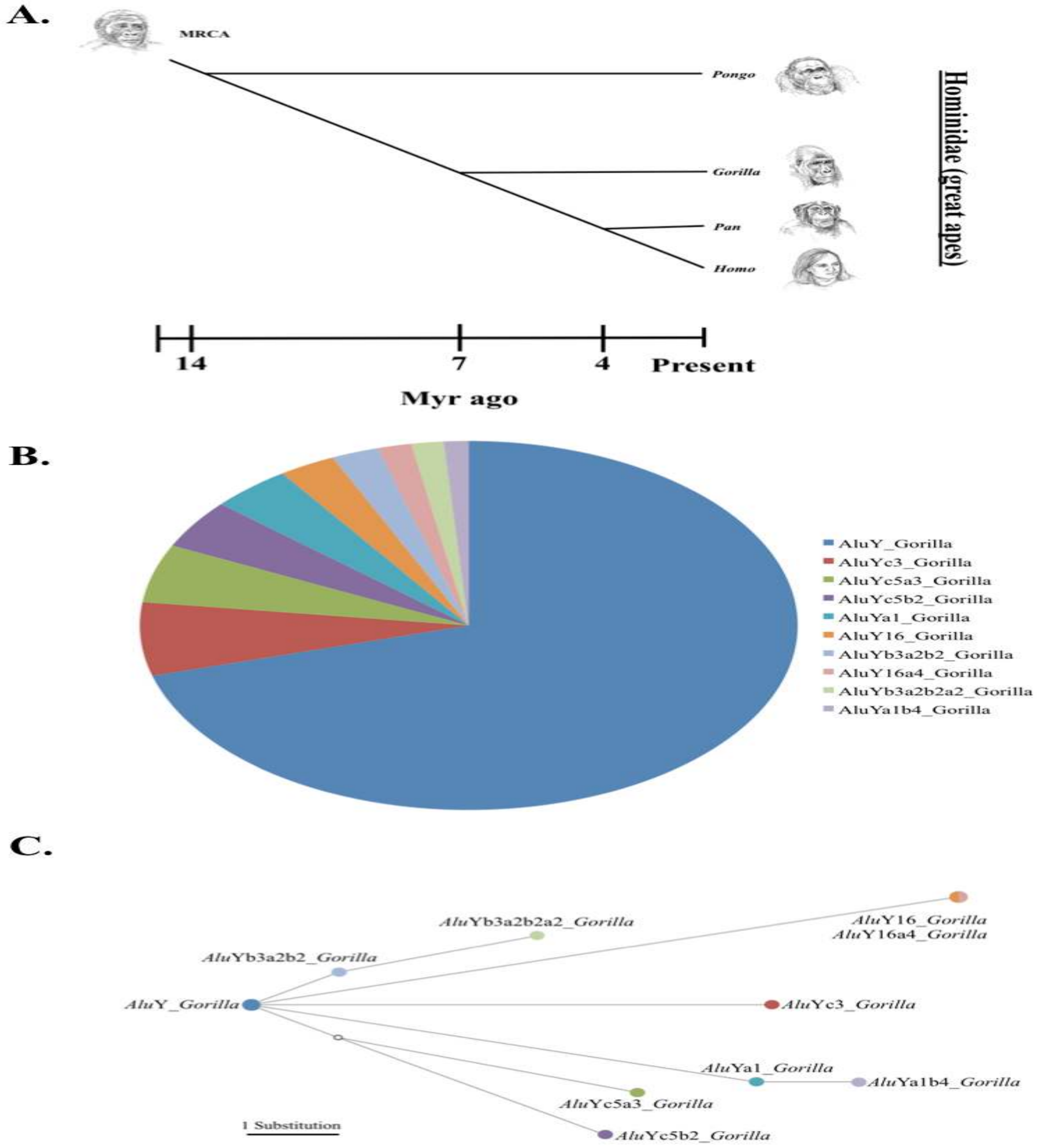

Figure 3.2: Analysis of gorilla-specific *Alu* subfamilies.
A) A schematic diagram of a tree of evolutionary relationships of the four genera in Family Hominidae (great apes) based on divergence dates of 6-9mya for the *Gorilla-Homo/Pan* speciation event [28, 53-55] **B)** A pie chart showing a color-coded distribution of Gorilla-specific *Alu*Y subfamilies. *Alu*Y_*Gorilla* is the largest subfamily, representing slightly less than 3/4 of the total copy number identified **C)** A stepwise analysis of the relationships between Gorilla-specific *Alu*Y subfamilies generated from a Network analysis of the consensus sequences for each subfamily. The color of the dots representing each subfamily are correlated with the colors in the pie chart in Figure 3.2.B.

This result is consistent with the ongoing propagation of these subfamilies before, during, and after the speciation event at a relatively constant rate. This indicates that the "master genes" (Deininger, Batzer et al. 1992) from which these subfamilies are derived already existed and were retrotranspositionally active prior to the aforementioned speciation event, and have remained active subsequently. Examination of *Alu* elements indicates retrotranspositionally active elements are relatively rare, and that most *Alu* activity is the result of a small number of "master" copies engaging in retrotranspositional activity over time (Deininger, Batzer et al. 1992). Our results suggest that the ten gorilla-specific *Alu*Y subfamilies identified in this study diverged and are still diverging from master elements already present in the genome of the common ancestor of the *Gorilla* and *Homo*-*Pan* lineages. A table listing each subfamily, the "master gene" or ancestral *Alu* subfamily from which it was likely derived, the % divergence from the consensus sequence of the master element, copy number, and suggested age of the most recent common ancestral element are available in Appendix A as Table S.6.

Conclusions

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

Materials and methods

Computational methodology

The genome of the Western lowland gorilla (*Gorilla gorilla gorilla*) was downloaded and analyzed for the presence of *Alu* elements using an in-house installation of the RepeatMasker program (Smit, Hubley et al. 1996-2010). The *Gorilla gorilla gorilla* genome is available for download and analysis via the website of the Wellcome Trust-Sanger Institute. The resulting dataset was parsed into separate files based on the *Alu* subfamily designations assigned by RepeatMasker. The file containing elements designated as members of the *Alu*Y subfamily was then further parsed to remove 84,251 elements under the length of 250bp using the estimation that shorter elements were likely to be older elements present in multiple species and therefore not useful for our analysis. The "Fetch Sequences" function from the online version of the Galaxy suite of programs (Giardine, Riemer et al. 2005; Blankenberg, Von Kuster et al. 2010; Goecks, Nekrutenko et al. 2010) was then used to retrieve the individual DNA sequence present at each of these loci using the gorilla genome build gorGor3.1, and the Lift Over function was used to examine these loci for gorilla lineage specificity by comparison to the closely related genomes of human (*Homo sapiens*; hg19), chimpanzee (*Pan troglodytes*; panTro2), and Sumatran orangutan (*Pongo pygmaeus abelii*; ponAbe2). An additional 200 bp of flanking sequence on each side of the loci assayed was included in this analysis for validation of orthologous loci between the nine primate species considered in this study.

Loci selected for verification were examined for further evidence of gorilla-specificity using the BLAST-Like Alignment Tool (BLAT) available at the UCSC Genome Browser website. Putative gorilla-specific loci were compared to the available genomes of 3 other primate species, human (hg19), chimpanzee (panTro2), and orangutan (ponAbe2) (Kent 2002; Kent, Sugnet et al. 2002). Elements found to be absent in these species and with sufficient orthologous flanking across species were marked for PCR primer design and experimental validation. Loci determined to be shared insertions, as well as those lacking sufficient orthologous flanking for effective primer design, were removed from further consideration (Kent 2002).

The COSEG program, designed to identify repeat subfamilies using significant co-segregating mutations, was then run on the remaining putative gorilla-specific insertions to identify and group specific subfamilies together. COSEG ignores non-diagnostic mutations during analysis, providing an accurate representation of relationships between subfamilies of elements by ignoring potentially misleading mutational events (Price, Eskin et al. 2004). COSEG uses a minimum subfamily size of 50 elements as the default setting. We arbitrarily defined subfamilies as groups of >10 elements to increase the detail of subfamily structure resolved. A subset of a minimum of 10% of each identified subfamily was then chosen for verification using locus-specific PCR, with a total of 279 loci assayed and verified.

A multi-species alignment comprised of the species listed above was created for each locus using BioEdit (Hall 1999). Oligonucleotide primers for the PCR assays were designed in shared regions flanking each putative gorilla-specific element chosen for experimental verification using the Primer3Plus program (Untergasser, Nijveen et al. 2007). These primers were then tested computationally against available primate genomes using the *in-silico* PCR tool on the UCSC Genome Bioinformatics website (Kent, Sugnet et al. 2002).

Subfamily age estimates were calculated using the BEAST program (**B**ayesian **E**volutionary **A**nalysis by **S**ampling **T**rees) (Drummond and Rambaut 2007). BEAST has previously been used to estimate dates of divergence using transposon data (Hellen and Brookfield 2013). For each subclade, the consensus sequence for each subfamily was determined from the COSEG output (Price, Eskin et al. 2004). The progenitor element was determined by RepeatMasker analysis of each consensus sequence. Elements were aligned using the SeaView software program and MUSCLE algorithm (Edgar 2004; Gouy, Guindon et al. 2010). The progenitor element was then used as an out-group to root the tree of each subclade. BEAST was calibrated with a baseline divergence date of 7 mya for the split between the *Gorilla* and *Homo-Pan* lineages. A divergence date of 7 mya is within the generally accepted 6-9 mya range for this divergence (Chen and Li 2001; Glazko and Nei 2003; Salem, Ray et al. 2003; Steiper and Young 2006). BEAST was run with the following parameters: Site Heterogeneity= "gamma"; Clock= "strict clock"; Species Tree Prior= "birth death process"; Prior for Time of Most Recent Common Ancestor (tmrca)= "Normal distribution" with mean of 7.0 million years and 1.0 standard deviation"; ucld.mean= "uniform model" with initial rate set at 0.033; Length of Chain= "10,000,000"; All other parameters were left at default settings.

The Network program was run on gorilla-specific *Alu*Y subfamily consensus sequences to generate a stepwise tree of relationships between identified subfamilies (Bandelt, Forster et al. 1999; Cordaux, Hedges et al. 2004). The consensus sequences for the gorilla-specific *Alu*Y subfamilies were aligned using the MUSCLE algorithm (Edgar 2004) and converted to the .rdf file format using the DNAsp program (Librado and Rozas 2009). The .rdf file was then imported to Network, and a median-joining analysis was run. The resulting output file demonstrating evolutionary relationships between subfamilies is presented in Figure 3.2.C.

PCR and DNA sequencing

To verify gorilla-specificity, locus specific PCR was performed with a 9-species primate panel comprised of DNA samples from the following species: Western lowland gorilla (*Gorilla gorilla gorilla*); Human HeLa (*Homo sapiens*); Common chimpanzee (*Pan troglodytes*); Bonobo

(*Pan paniscus*); Bornean orangutan (*Pongo pygmaeus*); Sumatran orangutan (*Pongo abelii*); Northern white-cheeked gibbon (*Nomascus leucogenys*); Rhesus macaque (*Macaca mulatta*); African green monkey (*Chlorocebus aethiops*).

PCR amplification of each locus was performed in 25µl reactions using 15ng of template DNA, 200nM of each primer, 200µM dNTPs in 50mM KCl, 1.5mM $MgCl_2$, 10mM Tris-HCl (pH 8.4), and 2 units of *Taq* DNA polymerase. PCR reaction conditions were as follows: an initial denaturation at 95°C for 1 minute, followed by 32 cycles of denaturation at 95°C, annealing at the previously determined optimal annealing temperature (60°C with some exceptions), and extension at 72°C for 30 seconds each, followed by a final extension of 72°C for 1 minute. PCR products were analyzed to confirm gorilla specificity of all loci on 2% agarose gels stained with 0.25ug ethidium bromide and visualized with UV fluorescence (Figure 3.3). A list of all 279 assayed loci, corresponding primer pairs, and optimal annealing temperatures for each are available as in Appendix A as Table S.4. Additionally, all PCR-tested loci containing unidentified bases in the original sequence data were subjected to chain-termination sequencing to verify base pair composition (Sanger, Air et al. 1977). Sequence data generated from this project for gorilla-specific *Alu*Y subfamilies has been deposited in GenBank under the accession numbers (KF668269-KF668278).

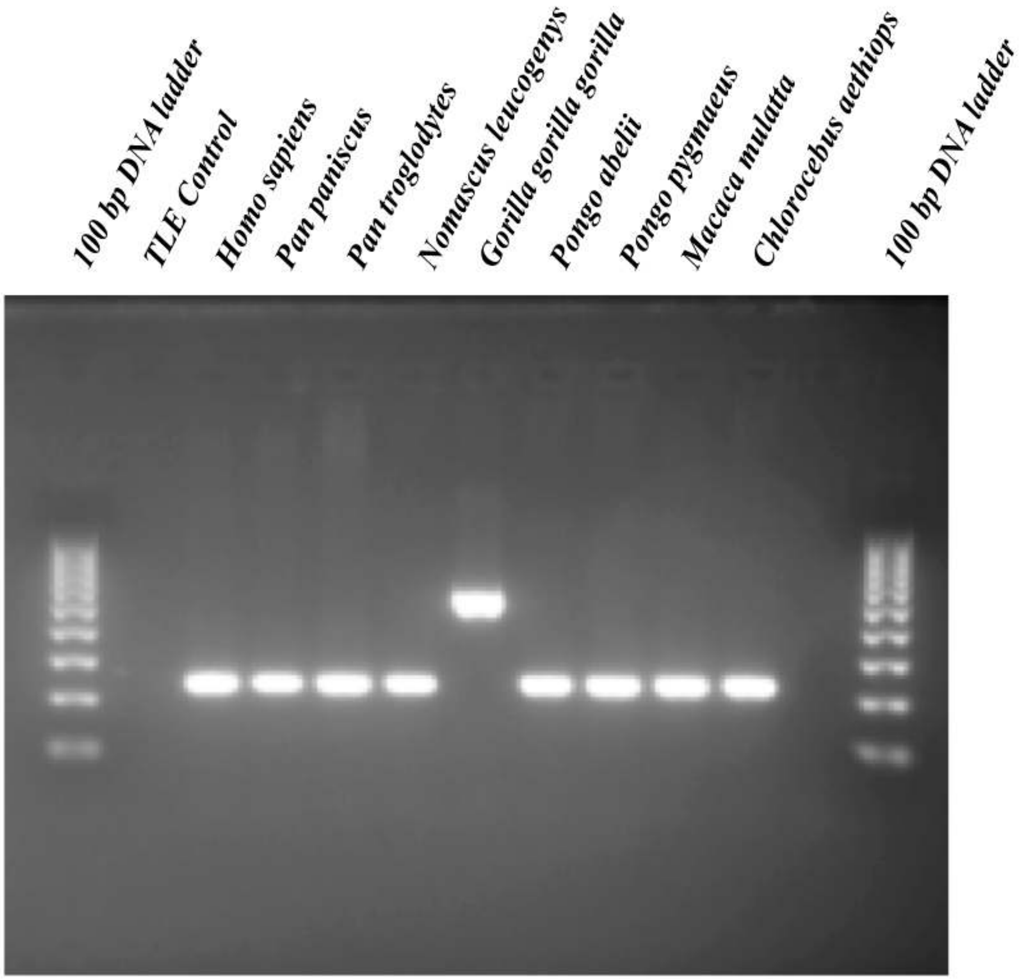

Figure 3.3: Phylogenetic assay of a western lowland gorilla-specific *Alu* insertion (Primer Pair Gor112). An agarose gel chromatograph of the gorilla specific *Alu* insertion Gor112. The filled site is approximately 550 bp (lane 7) and the empty site is 250 bp (lanes 3-6 & 8-11). Lanes (1) 100 bp DNA ladder; (2) negative control; (3) human; (4) bonobo; (5) common chimpanzee; (6) northern white-cheeked gibbon;(7) western lowland gorilla; (8) Sumatran orangutan; (9) Bornean orangutan; (10) Rhesus macaque; (11) green monkey; (12) empty; (13) 100 bp DNA ladder.

References

Bandelt, H. J., P. Forster, et al. (1999). "Median-joining networks for inferring intraspecific phylogenies." Mol Biol Evol **16**(1): 37-48.

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.

Batzer, M. A., 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." Proc Natl Acad Sci U S A **91**(25): 12288-12292.

Belancio, V. P., D. J. Hedges, et al. (2008). "Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health." Genome Res **18**(3): 343-358.

Blankenberg, D., G. Von Kuster, et al. (2010). "Galaxy: a web-based genome analysis tool for experimentalists." Curr Protoc Mol Biol **Chapter 19**: Unit 19 10 11-21.

Bogerd, H. P., H. L. Wiegand, et al. (2006). "Cellular inhibitors of long interspersed element 1 and Alu retrotransposition." Proc Natl Acad Sci U S A **103**(23): 8780-8785.

Britten, R. J., W. F. Baron, et al. (1988). "Sources and evolution of human Alu repeated sequences." Proc Natl Acad Sci U S A **85**(13): 4770-4774.

Carter, A. B., A. H. Salem, et al. (2004). "Genome-wide analysis of the human Alu Yb-lineage." Hum Genomics **1**(3): 167-178.

Chen, F. C. and W. H. Li (2001). "Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees." Am J Hum Genet **68**(2): 444-456.

Cook, G. W., M. K. Konkel, et al. (2011). "Alu pair exclusions in the human genome." Mob DNA **2**: 10.

Cook, G. W., M. K. Konkel, et al. (2013). "A comparison of 100 human genes using an alu element-based instability model." PLoS One **8**(6): e65188.

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. (2004). "Retrotransposition of Alu elements: how many sources?" Trends Genet **20**(10): 464-467.

Cordaux, R., D. J. Hedges, et al. (2006). "Estimating the retrotransposition rate of human Alu elements." Gene **373**: 134-137.

Cordaux, R., J. Lee, et al. (2006). "Recently integrated Alu retrotransposons are essentially neutral residents of the human genome." Gene **373**: 138-144.

Cost, G. J., Q. Feng, et al. (2002). "Human L1 element target-primed reverse transcription in vitro." EMBO J **21**(21): 5899-5910.

Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-193.

Deininger, P. L., M. A. Batzer, et al. (1992). "Master genes in mammalian repetitive DNA amplification." Trends Genet **8**(9): 307-311.

Dewannieux, M., C. Esnault, et al. (2003). "LINE-mediated retrotransposition of marked Alu sequences." Nat Genet **35**(1): 41-48.

Drummond, A. J. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evol Biol **7**: 214.

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Res **32**(5): 1792-1797.

Farkash, E. A., G. D. Kao, et al. (2006). "Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay." Nucleic Acids Res **34**(4): 1196-1204.

Feng, Q., J. V. Moran, et al. (1996). "Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition." Cell **87**(5): 905-916.

Fleagle, J. G. (1999). Primate adaptation and evolution. San Diego ; London, Academic Press.

Fleagle, J. G., C. H. Janson, et al. (1999). Primate communities. Cambridge, Cambridge University Press.

Giardine, B., C. Riemer, et al. (2005). "Galaxy: a platform for interactive large-scale genome analysis." Genome Res **15**(10): 1451-1455.

Glazko, G. V. and M. Nei (2003). "Estimation of divergence times for major lineages of primate species." Mol Biol Evol **20**(3): 424-434.

Goecks, J., A. Nekrutenko, et al. (2010). "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." Genome Biol **11**(8): R86.

Gouy, M., S. Guindon, et al. (2010). "SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." Mol Biol Evol **27**(2): 221-224.

Hall, T. A. (1999). "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT." Nucleic Acids Symp. Ser. **41**: 95-98.

Han, K., J. Xing, et al. (2005). "Under the genomic radar: the stealth model of Alu amplification." Genome Res **15**(5): 655-664.

Hancks, D. C. and H. H. Kazazian, Jr. (2012). "Active human retrotransposons: variation and disease." Curr Opin Genet Dev **22**(3): 191-203.

Hedges, D. J. and M. A. Batzer (2005). "From the margins of the genome: mobile elements shape primate evolution." Bioessays **27**(8): 785-794.

Hedges, D. J., P. A. Callinan, et al. (2004). "Differential alu mobilization and polymorphism among the human and chimpanzee lineages." Genome Res **14**(6): 1068-1075.

Hedges, D. J. and P. L. Deininger (2007). "Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity." Mutat Res **616**(1-2): 46-59.

Hellen, E. H. and J. F. Brookfield (2013). "The diversity of class II transposable elements in mammalian genomes has arisen from ancestral phylogenetic splits during ancient waves of proliferation through the genome." Mol Biol Evol **30**(1): 100-108.

Hormozdiari, F., M. K. Konkel, et al. (2013). "Rates and patterns of great ape retrotransposition." Proc Natl Acad Sci U S A **110**(33): 13457-13462.

Hulme, A. E., H. P. Bogerd, et al. (2007). "Selective inhibition of Alu retrotransposition by APOBEC3G." Gene **390**(1-2): 199-205.

Jurka, J., V. V. Kapitonov, et al. (2005). "Repbase Update, a database of eukaryotic repetitive elements." Cytogenet Genome Res **110**(1-4): 462-467.

Jurka, J. and T. Smith (1988). "A fundamental division in the Alu family of repeated sequences." Proc Natl Acad Sci U S A **85**(13): 4775-4778.

Kapitonov, V. and J. Jurka (1996). "The age of Alu subfamilies." J Mol Evol **42**(1): 59-65.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.

Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.

Konkel, M. K. and M. A. Batzer (2010). "A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome." Semin Cancer Biol **20**(4): 211-221.

Konkel, M. K., J. A. Walker, et al. (2010). "LINEs and SINEs of primate evolution." Evolutionary Anthropology **19**: 236-249.

Langergraber, K. E., K. Prufer, et al. (2012). "Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution." Proc Natl Acad Sci U S A **109**(39): 15716-15721.

Lee, J., K. Han, et al. (2008). "Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons." PLoS One **3**(12): e4047.

Li, J., K. Han, et al. (2009). "Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements." Gene **448**(2): 242-249.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics **25**(11): 1451-1452.

Liu, G. E., C. Alkan, et al. (2009). "Comparative analysis of Alu repeats in primate genomes." Genome Res **19**(5): 876-885.

Locke, D. P., L. W. Hillier, et al. (2011). "Comparative and demographic analysis of orang-utan genomes." Nature **469**(7331): 529-533.

Luan, D. D. and T. H. Eickbush (1995). "RNA template requirements for target DNA-primed reverse transcription by the R2 retrotransposable element." Mol Cell Biol **15**(7): 3882-3891.

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.

McLain, A. T., T. J. Meyer, et al. (2012). "An alu-based phylogeny of lemurs (infraorder: Lemuriformes)." PLoS One **7**(8): e44035.

Meyer, T. J., A. T. McLain, et al. (2012). "An Alu-based phylogeny of gibbons (hylobatidae)." Mol Biol Evol **29**(11): 3441-3450.

Minghetti, P. P. and A. Dugaiczyk (1993). "The emergence of new DNA repeats and the divergence of primates." Proc Natl Acad Sci U S A **90**(5): 1872-1876.

Murata, S., N. Takasaki, et al. (1993). "Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution." Proc Natl Acad Sci U S A **90**(15): 6995-6999.

Perna, N. T., M. A. Batzer, et al. (1992). "Alu insertion polymorphism: a new type of marker for human population studies." Hum Biol **64**(5): 641-648.

Price, A. L., E. Eskin, et al. (2004). "Whole-genome analysis of Alu repeat elements reveals complex evolutionary history." Genome Res **14**(11): 2245-2252.

Ray, D. A., J. Xing, et al. (2005). "Alu insertion loci and platyrrhine primate phylogeny." Mol Phylogenet Evol **35**(1): 117-126.

Ray, D. A., J. Xing, et al. (2006). "SINEs of a nearly perfect character." Syst Biol **55**(6): 928-935.

Roos, C., J. Schmitz, et al. (2004). "Primate jumping genes elucidate strepsirrhine phylogeny." Proc Natl Acad Sci U S A **101**(29): 10650-10654.

Roy-Engel, A. M., M.A. Batzer, P.L. Deininger (2008). Evolution of human retrosequences: Alu. Encyclopedia of Life Sciences. Chichester, United Kingdom, John Wiley & Sons, Ltd.

Roy-Engel, A. M., A. H. Salem, et al. (2002). "Active Alu element "A-tails": size does matter." Genome Res **12**(9): 1333-1344.

Salem, A. H., D. A. Ray, et al. (2003). "Alu elements and hominid phylogenetics." Proc Natl Acad Sci U S A **100**(22): 12787-12791.

Sanger, F., G. M. Air, et al. (1977). "Nucleotide sequence of bacteriophage phi X174 DNA." Nature **265**(5596): 687-695.

Scally, A., J. Y. Dutheil, et al. (2012). "Insights into hominid evolution from the gorilla genome sequence." Nature **483(**7388): 169-175.

Schmitz, J., M. Ohme, et al. (2001). "SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates." Genetics **157**(2): 777-784.

Schumann, G. G. (2007). "APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition." Biochem Soc Trans **35**(Pt 3): 637-642.

Sen, S. K., K. Han, et al. (2006). "Human genomic deletions mediated by recombination between Alu elements." Am J Hum Genet **79**(1): 41-53.

Shedlock, A. M. and N. Okada (2000). "SINE insertions: powerful tools for molecular systematics." Bioessays **22**(2): 148-160.

Shedlock, A. M., K. Takahashi, et al. (2004). "SINEs of speciation: tracking lineages with retroposons." Trends Ecol Evol **19**(10): 545-553.

Slagel, V., E. Flemington, et al. (1987). "Clustering and subfamily relationships of the Alu family in the human genome." Mol Biol Evol **4**(1): 19-29.

Smit, A., R. Hubley, et al. (1996-2010). "RepeatMasker Open-3.0." from http://www.repeatmasker.org.

Steiper, M. E. and N. M. Young (2006). "Primate molecular divergence dates." Mol Phylogenet Evol **41**(2): 384-394.

Strier, K. B. (2007). Primate behavioral ecology. Boston, Allyn and Bacon.

Ullu, E. and C. Tschudi (1984). "Alu sequences are processed 7SL RNA genes." Nature **312**(5990): 171-172.

Untergasser, A., H. Nijveen, et al. (2007). "Primer3Plus, an enhanced web interface to Primer3." Nucleic Acids Res **35**(Web Server issue): W71-74.

Ventura, M., C. R. Catacchio, et al. (2011). "Gorilla genome structural variation reveals evolutionary parallelisms with chimpanzee." Genome Res **21**(10): 1640-1649.

Walker, J. A., M. K. Konkel, et al. (2012). "Orangutan Alu quiescence reveals possible source element: support for ancient backseat drivers." Mob DNA **3**: 8.

Willard, C., H. T. Nguyen, et al. (1987). "Existence of at least three distinct Alu subfamilies." J Mol Evol **26**(3): 180-186.

Xing, J., H. Wang, et al. (2005). "A mobile element based phylogeny of Old World monkeys." Mol Phylogenet Evol **37**(3): 872-880.

Xing, J., H. Wang, et al. (2007). "A mobile element-based evolutionary history of guenons (tribe Cercopithecini)." BMC Biol **5**: 5.

Xing, J., Y. Zhang, et al. (2009). "Mobile elements create structural variation: analysis of a complete human genome." Genome Res **19**(9): 1516-1526.

Zietkiewicz, E., C. Richer, et al. (1999). "Phylogenetic affinities of tarsier in the context of primate Alu repeats." Mol Phylogenet Evol **11**(1): 77-83.

CHAPTER FOUR:
CONCLUSION

Alu elements are primate-specific SINEs of ~300bp length that have, over the past three decades, been recognized and studied as important genetic markers in phylogenetics, comparative genomics, population genetics, and disease research (Deininger and Batzer 1999; Batzer and Deininger 2002; Ray, Xing et al. 2006; Konkel, Walker et al. 2010). The prolific spread of these mobile elements during the primate radiation has been great enough that they now comprise >10% of the human genome, and significant percentages of the genomes of other primates surveyed (Deininger and Batzer 2002). Phylogenetic studies of all parts of the primate order have now been conducted using *Alu* as the primary marker and trees with strong statistical support have been produced elucidating relationships between species and genera (Zietkiewicz, Richer et al. 1999; Schmitz, Ohme et al. 2001; Salem, Ray et al. 2003; Ray, Xing et al. 2005; Xing, Wang et al. 2005; Herke, Xing et al. 2007; Xing, Wang et al. 2007; Li, Han et al. 2009; McLain, Meyer et al. 2012; Meyer, McLain et al. 2012). Additionally, *Alu* elements have been used effectively in the study of variation and divergence between extant human populations (Batzer, Stoneking et al. 1994; Watkins, Rogers et al. 2003; Witherspoon, Marchani et al. 2006). With the continued expansion of sequenced primate genomes available to researchers it is likely that *Alu*-based studies will be able to delve even deeper into the evolutionary history of this lineage, and provide ever more accurate phylogenetic reconstructions and analyses of variance between populations.

Chapter Two of this dissertation examines evolutionary relationships within the Strepsirrhine infraorder Lemuriformes, a branch of primates endemic to Madagascar and a few surrounding islands. Lemurs are monophyletic, with all extant species (~103) derived from a single ancestral population that likely colonized Madagascar during a rafting even from mainland

Africa ~55-60 mya. The magnificent diversity seen among Malagasy primates is a result of the unique geography and climatic history of the island, allowing for relatively rapid divergence and speciation events (Yoder, Cartmill et al. 1996; Yoder and Yang 2004; Mittermeier, Louis et al. 2010). Our computational analysis of genomic DNA from the grey mouse lemur (*Microcebus murinus*) as well as other nuclear DNA sequences from various species available from GenBank, revealed the presence of numerous Lemuriformes-specific *Alu* elements (Smit, Hubley et al. 1996-2004; Kent 2002; Lindblad-Toh, Garber et al. 2011). Using the 138 of these loci deemed phylogenetically informative, a tree was constructed via the Dollo parsimony method comprised of 22 species of lemur from all five extant genera as well as two out-group species, the galago or "Senegalese bushbaby" (*Galago senegalensis*) and human (*Homo sapiens sapiens*). In particular, our tree presented two major genera-level groupings: Indriidae-Lemuridae, and Lepilemuridae-Cheirogaleidae, with Daubentoniidae confirmed as the most diverged genus from the other four (McLain, Meyer et al. 2012). This tree demonstrated strong statistically significance at all nodes, and additional research into the genera- and species-level relationships may provide sufficient loci to establish these groupings beyond question (Waddell, Kishino et al. 2001). As additional whole genome sequences become available for more lemur species this tree can be expanded upon and further strengthened. Lemurs are a fascinating and grievously endangered clade of primates, and stringent scientific research and conservation work is necessary to preserve remaining populations. It is hoped that this study provides some contribution to the future survival of these magnificent animals.

In Chapter Three we examine active *Alu* subfamilies in the western lowland gorilla (*Gorilla gorilla gorilla*) genome. Western lowland gorillas are large, endangered great apes endemic to the montane and lowland rainforests and swamps of central Africa. They are the most populous of extant gorilla species, and like all gorillas a close evolutionary relative of humans and the *Pan* lineage of chimpanzees and bonobos. Together the living and extinct gorilla, human, and chimpanzee/bonobo lineages comprise the subfamily Homininae. Large-bodied, herbivorous (with the occasional insect), and gregarious, western lowland gorillas live in family groups centered on a dominant male (a "silverback") and multiple females, lesser males, and dependent offspring (Fleagle 1999; Scally, Dutheil et al. 2012). The genome of the western lowland gorilla was searched for *Alu* elements using the RepeatMasker program, and the *Alu*Y subfamily was targeted as being most likely to yield active elements (Smit, Hubley et al. 1996-2004).

A total of 1,075 western lowland gorilla-specific *Alu*Y insertions were identified using computational and laboratory research. These insertions were then computationally classified into 10 subfamilies based on diagnostic mutations in the sequence structure of the element. This represents the clearest picture thus far of *Alu*Y subfamily activity in the western lowland gorilla, and provides exciting opportunities for future comparative work between this great ape and other species. Several thousand additional elements were classified as "likely" specific to the western lowland gorilla genome, but a lack of sufficient orthologous flanking for primer design in related species made PCR verification of the status of these elements impossible. At a future date they could be re-examined to provide an even clearer picture of *Alu*Y subfamily activity in this species.

Finally, our work put forth in this dissertation further reinforces that *Alu* elements are a reliable phylogenetic marker and an informative tool in the study of primate evolution. The nearly homoplasy-free nature and known polarity of these elements makes them ideal for constructing phylogenetic trees. The exploration of active *Alu*Y subfamilies in a closely related primate like the western lowland gorilla makes it possible to compare and contrast activity levels with humans and chimpanzees/bonobos, and thereby gain insight into the evolutionary history of all of these lineages. Non-human primates are in continual danger from human activity, and all possible information that can be obtained about them is valuable to the effort to conserve existing populations. As a primate-specific SINE, *Alu* elements have a unique capacity to elucidate phylogenies, answer questions about population genetics, and provide information that is invaluable to researchers studying numerous genetic diseases. As the genomic era continues to progress apace there is little remaining doubt that these sequences, formerly classified as "junk", will continue to be used as invaluable sources of genomic insights.

References

Batzer, M. A. and P. L. Deininger (2002). "Alu repeats and human genomic diversity." Nat Rev Genet **3**(5): 370-379.

Batzer, M. A., M. Stoneking, et al. (1994). "African origin of human-specific polymorphic Alu insertions." Proc Natl Acad Sci U S A **91**(25): 12288-12292.

Deininger, P. L. and M. A. Batzer (1999). "Alu repeats and human disease." Mol Genet Metab **67**(3): 183-193.

Deininger, P. L. and M. A. Batzer (2002). "Mammalian retroelements." Genome Res **12**(10): 1455-1465.

Fleagle, J. G. (1999). Primate adaptation and evolution. San Diego ; London, Academic Press.

Herke, S. W., J. Xing, et al. (2007). "A SINE-based dichotomous key for primate identification." Gene **390**(1-2): 39-51.

Kent, W. J. (2002). "BLAT--the BLAST-like alignment tool." Genome Res **12**(4): 656-664.

Konkel, M. K., J. A. Walker, et al. (2010). "LINEs and SINEs of primate evolution." Evolutionary Anthropology **19**: 236-249.

Li, J., K. Han, et al. (2009). "Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements." Gene **448**(2): 242-249.

Lindblad-Toh, K., M. Garber, et al. (2011). "A high-resolution map of human evolutionary constraint using 29 mammals." Nature **478**(7370): 476-482.

McLain, A. T., T. J. Meyer, et al. (2012). "An alu-based phylogeny of lemurs (infraorder: Lemuriformes)." PLoS One **7**(8): e44035.

Meyer, T. J., A. T. McLain, et al. (2012). "An Alu-based phylogeny of gibbons (hylobatidae)." Mol Biol Evol **29**(11): 3441-3450.

Mittermeier, R. A., E. E. Louis, et al. (2010). Lemurs of Madagascar. Arlington, Virginia, Conservation International.

Ray, D. A., J. Xing, et al. (2005). "Alu insertion loci and platyrrhine primate phylogeny." Mol Phylogenet Evol **35**(1): 117-126.

Ray, D. A., J. Xing, et al. (2006). "SINEs of a nearly perfect character." Syst Biol **55**(6): 928-935.

Salem, A. H., D. A. Ray, et al. (2003). "Alu elements and hominid phylogenetics." Proc Natl Acad Sci U S A **100**(22): 12787-12791.

Scally, A., J. Y. Dutheil, et al. (2012). "Insights into hominid evolution from the gorilla genome sequence." Nature **483**(7388): 169-175.

Schmitz, J., M. Ohme, et al. (2001). "SINE insertions in cladistic analyses and the phylogenetic affiliations of Tarsius bancanus to other primates." Genetics **157**(2): 777-784.

Smit, A., R. Hubley, et al. (1996-2004). "RepeatMasker Open-3.0."

Waddell, P. J., H. Kishino, et al. (2001). "A phylogenetic foundation for comparative mammalian genomics." Genome Inform **12**: 141-154.

Watkins, W. S., A. R. Rogers, et al. (2003). "Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms." Genome Res **13**(7): 1607-1618.

Witherspoon, D. J., E. E. Marchani, et al. (2006). "Human population genetic structure and diversity inferred from polymorphic L1(LINE-1) and Alu insertions." Hum Hered **62**(1): 30-46.

Xing, J., H. Wang, et al. (2005). "A mobile element based phylogeny of Old World monkeys." Mol Phylogenet Evol **37**(3): 872-880.

Xing, J., H. Wang, et al. (2007). "A mobile element-based evolutionary history of guenons (tribe Cercopithecini)." BMC Biol **5**: 5.

Yoder, A. D., M. Cartmill, et al. (1996). "Ancient single origin for Malagasy primates." Proc Natl Acad Sci U S A **93**(10): 5122-5126.

Yoder, A. D. and Z. Yang (2004). "Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context." Mol Ecol **13**(4): 757-773.

Zietkiewicz, E., C. Richer, et al. (1999). "Phylogenetic affinities of tarsier in the context of primate Alu repeats." Mol Phylogenet Evol **11**(1): 77-83.

APPENDIX A:
SUPPLEMENTAL DATA

Table S.1 - A table listing all primers used in this study with optimal annealing temperatures and source. Primers highlighted in green were designed specifically for this study. Primers highlighted in yellow were designed by Herke et al. (2007). Primers highlighted in red were designed by Roos et al. (2004).

Locus	Tm	Forward Primer	Reverse Primer
M1	60.4	GATTAGCTCAATGTGTTTTCATT	CTAGAAGCCCATTTCAGCAGC
M2	63	TGGTCTTTTTCCATTGCCTG	GGATCTTAAAAAACAATTAGAGGAGG
M3	63	TTGCTGCAAATCTGGGTGTC	TTGGATTTAAAGGAACATAATGGC
M4	60.4	AGCACCCAGACTGTGGTCTC	GCTTATCTTGAATAGATGCTTTTTAG
M6	55.5	TGAAGAGGTTTAGAAACAGAATG	TATTTAGTTGAGAAGCAAATGAATC
M7	60	CAGAAATAGACATTTCCCTTCC	CCAATTCAGCAGTGGTAAAG
M11	56	TTGCCCAGGCTGAATTTGG	GGGCACACAGCTGCAGTG
M12	56	CCCGTTTTTGTGACTGCAC	CAGACCTCCTGGACTCACC
M15	55.5	ATTAGGATGCCCAAAGACCC	GCCACCCACAAAAGACAAGT
M17	55.5	GAACACCTTGACTTACTGGAAG	CCCAGCTGTAATCCACAG
M18	56	AGTCAGTGGGCAAATTGGC	CAACTTTCCATTTGCATTTGG
M34	55.5	ACCTGTAAACACACCTCTATCAGA	GATTCACTGAAGAAATGACAGTCTC
MmA2	59	CCGATGTTCATGAATTGGAA	CCTGGAACATTTGTCAAAAATCA
MmA6	57	CTCAGGGCTTTTCAAACCAG	GATATTTATTAGCTGCCATTT
MmA7	57	GCATAAATTGAAGCACAGAGAA	TGGGGAGAGTGTAGAATACTTAGAA
MmA11	58	AAGGACCATAGACTTCAACCAGA	GTATGACGTGTATCCAGGAATTGA
MmA14	56	TAAATTCAAAGCATTAATGA	GGAGTTAATGCCAAAGATATGTT
MmA18	58	TTGTCTAGCTTAATGATAAAACTATAGAAA	ACCTGTCTTTTCCCCTCTGATAT
MmA19	58	ATATGCTGATTGGAGAGTATGAAGA	ATGGTTAGTCTTTCTAATTTTAGCCATT
MmA20	58	ATGGATTGGAGTTGGAATTTTCTAA	CAACCATGGTTCTACTGCAGAA
MmA22	58	GAAGCAGTGGTCTCTGCTCTCTAC	TCTAGGTTAGTGTTGGAGAATTTGT
MmA25	58	AGGTTTGCACATTAGAAGCCTT	GTTTCATGTGTATATCCCAGCAC
MmA27	61	TATGCCCAGGAAGCGCAGGCT	AGCCCATCCGGAGCTATGA
MmA28	58	GGTCAAAGTGCTTCCTTAAGG	TCCTCTCATCTTTTTAACTCAGCT
MmA29	58.5	GGATATATGCCTGAGAAAACAA	CCTAGTTTCATGAAAATTTATGC
MmA33	63	GACTTTTGGCAAAGGCCAA	CTGGCTTCTAGGCACCTCTGC
MmA36	58	ACAGCAGGATTCCAAATGGA	GAGGTAGTTGCCTCGAACTC
MmA39	62	AATGCCAGCCATGCCACT	GGGTCGTACAAAGGAGAAGCC
MmA42	58	GAGGCAGAGGAAAAGGTAATTAT	TCTCTTACATACTACTGCCACCT
MmA43	56	GGACAGCATATTAGCAAACTAAA	CCTTCCAATTATGAATGCTTTA

MmA45	56	GATTCCTGGAACACCTAAGAA	TCCTCTAAGTTTACTTTGAGCCT
MmA46	60	CATGAGCTTCCAGAGCCAGT	GCTCAGGAACCAATTTGGAA
MmA49	60	CCTGACAATGTTTCCAGTAGCA	CCCTCTGGAGGTCAAGCTGAT
MmA51	58	TTTCCTCACTTGGGAGAAA	TCTCTTAACTCTAGAAGACAGTTAACC
MmA52	60	GCCTGAGTTTTTCCAAAAGTT	TCCTTCTCAAGATAGTATTTGCC
MmA53	58	GGAGACATGTAGGTCATTGC	GCTGTAATAAACCACAGCCA
MmA54	57	GAAGAAAGACCACAGAGAAGAA	CATGAATTGGAAAGTACCTGC
MmA55	56	TTAGAAGAAGATGGAGAAGTGC	ATTCCTCTTACATGTTTACTAGCAT
MmA56	60	AGATGTGCCAGTTCCCAGTT	CTCTGTTTTAGGAATTTTGTCATCTA
MmA58	57	GGAGTGAGGTTCTGTGAACA	TCCTAACAGTGAAACTAAGCTTC
MmM27	56	ATAGTCAATGCCTTCACAGACAG	ATGAGTTGCTTGATTTGTCCAG
MmM35	58	TCCAAAGATCTTTAGACACTAGAAG	CACCAGTGCTTAATAGGAAGAC
MmM36	59	GAACACCTCTCCTTCCAGAATATAC	GACAATGGCCATACGCATC
MmM38	57	ACTCACCACTTTTCTCAGGC	ACTAAGACATGGGCCTATGC
MmM39	55	ATCTCAATGTTCTCAGGCAC	AATCTTGGAAGAGACCAGGAG
MmM41	55.5	GGTTCAAGTCTAGCATCTTGC	AAAATGAGAGTATGCATAATAACATG
MmM47	55	AGAGATATAGCAATCAACTGCA	GAATCAGGATATAAATTGTTAAAGG
MmM48	55.5	CCAATTTGATAGGGTAAAAGG	GCTCAGTGTCCTGCATCTAC
MmM49	56	TTTAAGGAAGTCAGGAAGGTAAGG	CATCAATATCCAGAGGGTTCC
MmM51	56	AATGGATCCTGAACATCTTTC	GCCATAACAATATTAAAGAGAAGTC
MmM57	55	GGCATTTCCTATGAGTGAAG	TAGTACCTAGCATAAAGTCAGTGC
MmM58	57.5	CATCACTTTGTCCAGCAACATTAG	CTTGCACTCCATGCTCTGACCTCC
MmM61	56	ACAGTTATTTGCATGCTAGTCAC	AAAGGAATTTGATGACAGTTAAGA
MmM73	56.5	GTTGGAGGTTGTTAGTTCAGG	ATCAAATTAAGTGGGAAATAGGTACTA
MmM97	55	AGAAACTGATTGTTACATTCCTTC	ACTGAAAGGACAAGCCAAG
MmM101	58	CTGGCAATGAACAAGACTTAGG	AATAGGAGCATCACATGAATGC
MmB2	57	GCAGGGTGGCATAACACATA	GCTTAAGAAACTTGCCTGACATA
MmB7	59	GTAAATTCCACTTTTATATCTGATAAGC	CACATAGCCTGACTGATAAACTATTAA
CmC1	51	AATGCATGCATTAAAATGTA	ATTGAAAAGGCCATTCA
EcM2	55	GCACAGGCTGGAGAAGCTAG	CATAATTTTGCTCAAGGCCAT
Em3	63	GTTACAAAGGGGCAAGGATGAG	GCTTGTTGCTTTTAACTGATGC
Em6	60	GCTCTCAAACACCAATCCCTAC	TGTCACCTTCAGATTATCCCAGT
Em7	64	CATTTAACCTTGCTTCCTCCAT	CAAGGTACCGTCTGTCACTTCA
EmA2	59.5	CCAACAAACCTTTGCACTACA	CATGCAATTTTGTCTTCCTTCA
EmA6	59.2	CCTTTGCACTACACTTGATGGT	CATGCAATTTTGTCTTCCTTCA
EmA8	58	GCTCTCAAACACCAATCCCTA	TGTCACCTTCAGATTATCCCAGT
EmM3	57	AATGCCAGCTATTATCTGCG	CCTCACCAGGGTGTGGTC
LcC2	62	CCCAGAACTTTAGCTTCCTTC	CATTCTCAACTGCTTGTCTCAG
LcM2	55	AATTTGTGCTCAGAAAAGAAGTC	CACAAGGCTGTCAAATGC
LcM3	55	TCCCTTCCTGAAACTCATG	TCAAGAGATGATGAGAGTTAGTAT

			CT
LcM7	55.5	TTTAGGATTATCAACCACTGCT	GGGTCAGTCTATTTTGAATCTG
LcM8	55.5	TGAAGCCTCTATGTTATTGACAA	CCATCAGTCTGGACCTGTG
PcC1	52	ATGGTGGAGCTGTTTCC	CTAAGACTATGAAGGACAGACC
PcC2	54	ATTAACTCAGCAAACATTGA	ACCTGGACCAGACACTAG
Str57A	55	tgtttcaaagatcaaatgtcatgg	cccaaggggagacagaagtt
Str57B	58	agratcctggaatactgrtcct	gctgagatgyatccctccyg
Str57C	55	caccacaactaccctggaagr	agaactctccttbacttaacctgtttc
Str58	52	atagccthtgcaggtcraag	gagcttaatatataavtwttctcaac
Str59	60	tacccagtgctcatttggatc	gcaatcagcaaatgcttgttttag
Str61	60	gcccagtttgtaattcttgatttc	gcaacccagctwacttcagagg
Str62	60	ctcactttcraaaatagcctactgc	aaggtcagtgttccytctctkc
Str63A	50	tcctctrtgactctyytgagttag	ggcacawttaagctatattrthcaa
Str64	54	tgaggcaartaaatatcttgakagc	tggagcccttgcwgtyg
Str66A	60	ctgcctcttaccatatttgacttc	gcctattcacatccatcactc
Str66B	60	gatgctaatgtggcttaatctcag	agccagctcttgaaatgtgc
Str67	51	gctttctgcatrggaggtagaataat	ccaatatcdtggcattctgtwaactc
Str68	60	aggctacagattttcaagatttacc	ctttagagccctctgctgga
Str69A	60	gctgtgccaaagaaatcagg	ggcacatttcctttgtttggag
Str69B	60	tgggraaaatgtttggcagat	tgtgattcattcttttggagca
Str70A	56	gtgyatgtrtgttggggcaar	cacttcytagcytgrtggcttt
Str70B	56	matacagtgaatacttgaatgkgc	taatcactaatgaaaatgtatctgarag
Str71B	48	gtcttattggatargaagaanctagg	gcctaaggtdcaaahttgtctc
Str71C	56	tgtgctatcccttttttaggrmtg	caacacaaaayvttgtcttaatrgg
Str72	50	ttcartttcactggraatcctttag	agagatcwtcagaactatgcccttr
Str73A	55	ccatttaggaagatgtgycatg	tgcttcactgagatamaagavga
Str73B	61	tggagaggacctaaattcaaagyttg	gaaagtgagytgggyatgaac
EcM5	55	GTCTGGCTGTTTCCAAATG	GGTGAAAGTTGTTGGGGAA
EmA102	55	CTGTTCCTTTCCGTCGATCT	GGAGGACTGCCACTACCTGA
EmA106	59	GATGGTAGTGATGGTTGCACA	TCACTTTCTCTTTTGCCATCTG
EmA108	58	TGTGGTGTAGTCAGCCAGAGA	AGTAAATCGAGGCAGAAATGC
EmA114	60	CACCTTGGGGAAGATTCAGT	TGAGTGGAGTGTTGCTCTGG
EmA115	59	GGCTCCATTTCTCCTCCTTT	GCCACCCACATCAAACAATA
EmA120	59	GACCTGGAGGGAAATTTTTG	TGAACCACCAACCCTAAAGA
EmA123	59	GGAAGGTTTGGTGTGCATTT	TGCTGTGATCTAAACGAGCAA
EmA126	57	CTTCTATCACCTGACAAAGTTCC	TCCATCTGCCTCAGTTTATCA
EmA129	58	TCAAAGAATCATACAAGTGTCTAATCA	CTCCCTTTTATAGAGCCTATCTCA
EmA130	60	AATGGGGAAACAGGAAAACC	CCAGATCTTGGTTTGGTAATTT
EmA132	60	GGTCAGCCAACTTCCAAAGA	TCTACTTCCGAGGCACCGTA
EmA173	55	CAGAAATGCCACGTGTACTGA	TAGTGCACCTGTCACCCAGA
EmA179	59	CATGTATGCCCACACACACA	CTGCCTCCCAAATGAAACAT
EmA199	59	AGCCAGGAACCATGGAGATA	CCTTGTTCCAAGCCAGTCTC
EmA1102	61	ATTGCGTGCTTGAGTTTTGG	AACTTCTCTCACGTCCTTCTCC
EmA1110	59	TAAGTCATGGCCATTTGGTG	TTGTTTTCATAATGGGTTGTGAG
EmA1111	60	CCCAGCCTCTTGGAAGTAAG	GGATGTGGCCATTCAATGTA
EmA111	59	GSAASTGTTATAATGACAATGATGC	GCTGAGACAGACTGCTGTGG

7			
HL2	55	CCGAAGTAAGTGCTGAAAACGAGG	CCGAAGGGATACAGAGATTACCTGAAG
HLEL1	55	GCCCAGACCCAAATATTCGTCTTCTAA	CCCATGAAGTCATTTTCAATGGCC
IND1	55	GGTGGAGCTGTTTCCATGC	GAGAACATACAGTGTCAGGAA
IND2	55	AGGTCAAATAAGGCATAGTCC	ATGGAATAGCTCTGAAGAAGG
IND3	55	TCTGGTGGAGGTTGGCTT	CCTCCATACTCAGAAGTCCT
MAD2	53	GTTTCACCATGAATTCCTTACT	AAAGGGATTAAAGGAATTTACAC
MAD3	53	AAAATATTTCCCATGAGATTAAATT	ATTGAGTTAATGGCAATAGACTT
LI1	53	CATGGGAGGTAGAATAATGAAT	TGGCATTCTGTAAACTCCACT
BK1	56	GTTATTGAACCCATGAAGTCAT	TGAGTTAGTGCTTAGAACCAC
BK4	56	CCAAAACTTGCAAAAGCTATTC	TCTTCATTTACAAACCCATGCT
BK17	58	GTTTATGATCCTGGTTTATGAG	CAGAGATCATCAGAACTTATGC
BK20	58	GAACACAAAAGAAGAATGGATC	AACAGCTTCTGTGCAATAACC
BK30	55	AGCCTAAGAAGCAAGAGACC	CTTCTGTTCTCCCTTTCCTAAC
BK31	58	GGTTTATCTCCTTTATCAACAAC	GAAGGCACATAAGCTACACTC
BK32	58	TGGTTATTGAACCCATGAAGTC	CACTCCAGTGTCCCTTCTAG
BK33	58	ACACACTACTCCATTACTACC	ATCAGACAAGGTTACTGAATGG
BK34	58	ATGTTGGGTATGTCATTTAACC	GTCTATAAAGTCACATATTAGTG

Table S.2 - A supplementary table showing the character matrix of amplification patterns of all loci in all species. A "1" indicates the locus is a filled site (*Alu* element present) and a "0" indicates the locus is an empty site (*Alu* element absent) in the corresponding species. A "?" indicates no amplification occurred of that primer pair in that species.

Locus ID	*Homo sapiens*	*Microcebus murinus*	*Cheirogaleus medius*	*Mirza coquereli*	*Lemur catta*	*Eulemur macaco*	*Eulemur macaco flavifrons*	*Eulemur fulvus albifrons*	*Eulemur fulvus collaris*	*Eulemur fulvus fulvus*	*Eulemur fulvus sanfordi*	*Eulemur rubriventer*	*Eulemur mongoz*	*Eulemur coronatus*	*Varecia variegata rubra*	*Varecia variegata variegata*	*Hapalemur griseus griseus*	*Lepilemur ruficaudatus*	*Propithecus coquereli*	*Propithecus coronatus*	*Propithecus verreauxi*	*Avahi laniger*	*Daubentonia madagascariensis*	*Galago senegalensis*
M1	0	1	1	1	?	?	1	1	1	?	?	?	?	?	1	?	?	?	?	?	?	?	?	?

M2	0	1	1	1	0	0	0	0	0	0	?	?	?	0	0	0	?	0	0	0	0	0	0	?
M3	0	1	1	1	1	1	1	1	?	1	?	?	?	1	1	1	?	1	1	1	1	?	1	?
M4	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
M6	0	1	1	1	1	1	1	1	1	1	1	1	?	1	1	1	1	1	?	1	1	1	0	?
M7	0	1	1	1	1	1	?	?	?	?	?	?	?	?	?	?	?	1	1	1	1	1	1	?
M11	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M11B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
M12	0	1	1	1	?	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	?	?
M15	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M17	0	1	?	1	0	0	?	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0	?
M18	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M34	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmA2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmA2B	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
MmA2C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
MmA6	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?
MmA7	0	1	1	1	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0	?
MmA11	?	1	0	0	0	0	0	0	?	?	?	?	?	?	?	?	?	0	0	0	0	0	0	0
MmA14	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	?
MmA18	0	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	?
MmA19	?	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	?	0
MmA20	0	1	1	1	0	0	0	0	?	0	0	0	0	?	?	0	?	0	0	0	0	0	0	?
MmA22	0	1	1	1	0	0	0	0	0	0	0	0	0	?	?	?	?	0	0	0	0	0	?	?
MmA25	0	1	0	?	0	0	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
MmA27	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmA28	0	1	1	1	?	1	1	1	1	1	1	1	1	1	1	1	1	?	1	1	1	?	1	?
MmA29	0	1	0	0	0	0	0	?	?	?	?	?	?	0	0	0	0	0	0	0	0	0	0	?
MmA33	0	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	?
MmA36	0	1	?	?	?	0	0	?	?	?	?	0	?	?	?	?	?	?	?	?	?	0	?	?
MmA39	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MmA42	?	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0
MmA43	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmA45	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
MmA46	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0
MmA49	0	1	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
MmA51	0	1	0	1	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	?	?
MmA52	0	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	0
MmA53	0	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?	1	1	1	1	1	?	1	1
MmA54	?	1	?	0	?	0	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0
MmA55	0	1	0	0	0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	0	0	0	0	?
MmA56	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

MmA58	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
MmM27	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	?	0	?	?	?
MmM35	0	1	0	1	0	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	0	0	0	?
MmM36	0	1	0	1	0	?	0	?	0	?	0	?	?	0	0	0	0	0	0	0	0	0	0	0
MmM38	0	1	0	1	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	0	?
MmM39	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmM41	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
MmM47	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	0	0	?
MmM48	0	1	0	0	?	0	0	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	?
MmM49	0	1	0	1	0	?	?	0	0	?	?	?	0	0	0	0	0	0	0	0	0	0	0	?
MmM51	0	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?	1	1	1	1	1	1	0	?
MmM57	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmM58	0	1	0	1	0	?	?	?	?	?	?	?	?	?	?	?	?	?	0	0	0	0	0	0
MmM61	0	1	0	1	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	0	?
MmM73	0	1	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0	0	0	0	?
MmM97	0	1	1	1	?	?	?	?	?	?	?	?	?	?	?	?	?	1	0	0	0	0	0	?
MmM101	0	1	0	1	0	0	0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0	0	0
MmB2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MmB7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CmC1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?
EcM2	0	?	?	?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Em3	0	0	0	0	0	1	1	1	1	1	1	0	0	0	?	0	0	0	?	?	?	?	0	?
Em6	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	?
Em7	?	?	0	?	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?	?	?	?	?	?
EmA2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
EmA6	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?
EmA8	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	?	?
EmM3	0	0	0	0	0	1	1	1	1	1	1	1	1	1	?	0	0	0	0	?	0	?	0	0
LcC2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	?
LcC2B	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	?
LcM2	0	0	0	?	1	1	1	1	1	1	1	1	1	?	1	1	1	0	0	0	0	0	0	?
LcM3	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	?	?
LcM7	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	?
LcM8	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	?
PcC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	1	1	1	0	0
PcC2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	?
Str57A	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	?	0	0
Str57B	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Str57C	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
Str58	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Str59	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Str61	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Str62	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Str63A	0	0	0	0	0	1	1	1	1	1	1	?	?	?	0	0	0	0	0	0	0	0	0	0
Str64	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	?	?
Str66A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	0	0	0	0	0	0	0	1
Str66B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Str67	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Str67A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
Str68	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Str69A	0	0	0	?	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Str69B	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	1
Str70A	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	?	0	?	?	0	0
Str70B	0	?	0	?	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
Str71B	0	1	1	1	1	1	1	1	1	1	1	1	?	1	1	1	1	?	1	1	1	1	1	0
Str71C	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Str72	0	?	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
Str73A	0	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?	1	1	1	1	1	1	?	?
Str73B	0	1	1	1	?	1	1	1	1	1	1	1	1	1	1	1	1	1	?	?	?	1	1	?
EcM5	0	0	?	?	?	1	1	1	1	1	1	?	1	1	1	?	?	0	0	0	0	0	0	?
EmA102	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	?
EmA106	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	?	?	?	?	?	?	?
EmA108	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	?	?	?	?	?	?	0
EmA114	?	?	?	?	0	1	1	0	0	0	0	0	0	0	?	?	0	?	?	?	?	?	?	?
EmA115	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	?	0	0	0	0	?	?	?
EmA120	?	?	0	?	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	?
EmA123	?	?	?	?	0	1	1	1	1	1	1	1	1	1	0	0	0	?	?	?	?	?	?	?
EmA126	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0
EmA129	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	?	?	0	?
EmA130	0	0	0	0	1	1	1	1	1	?	1	1	1	1	0	0	1	0	0	0	0	0	?	0
EmA132	0	0	?	?	1	1	1	1	1	1	1	1	1	1	0	0	1	?	?	?	?	?	?	?
EmA173	?	?	?	?	0	1	1	0	0	0	0	0	0	0	?	?	0	?	?	?	?	?	0	?
EmA179	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EmA199	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
EmA1102	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	?	?	?	?	0	?
EmA1110	?	?	?	?	0	1	1	1	1	1	1	0	0	0	0	0	0	?	?	?	?	?	?	?
EmA1111	?	?	?	?	0	1	1	1	1	1	1	0	0	0	0	0	0	?	?	?	?	?	?	?
EmA1117	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	?	?	?	?	?	?	?
MmA20A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	?
HL2	0	0	?	0	1	0	0	0	0	0	0	0	0	0	0	0	1	?	?	?	?	?	0	?
HLEL1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0
IND1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0

IND2	0	?	?	?	0	0	0	0	0	0	0	0	0	0	0	0	0	?	1	1	1	1	0	0
IND3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0
MAD2	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
MAD3	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
BK1	?	0	?	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	?	?	?	?	0
BK4	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
BK17	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
BK20	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
BK30	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
BK31	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
BK32	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
BK33	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	?	0
BK34	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
LI1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0

Table S.3- Enumeration of *Alu* Elements in Ape Genomes

Species	*Alu* Total	*Alu*Y	*Alu*S	*Alu*J
Western Lowland Gorilla (*Gorilla gorilla gorilla*)	1,085,174	141,709	599,237	286,801
Human (*Homo sapiens*)	1,194,734	143,178	686,962	312,138
Common Chimpanzee (*Pan troglodytes*)	1,181,265	137,537	680,914	310,816
Sumatran Orangutan (*Pongo pygmaeus abelii*)	1,242,224	142,704	699,099	296,058
Northern White-Cheeked Gibbon (*Nomascus leucogenys)*	1,141,075	187,588	620,667	284,494

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

Name	Oligo	Tm	Locus
Gor1-F	GAAGAACTTGCCCAAACCAC	59	chr11:20025155-20026490
Gor1-R	GCCCAGAAACAACACAAACT	57	
Gor2-F	ACCTTTGAGAGCAGGAAGCA	60	chr10:88745976-88747284
Gor2-R	TCTTGAAGGCAGGGACAAAT	59.7	
Gor3-F	TTAGCAGAGACAGGGGTTCA	58	chr3:5496399-

			5497699
Gor3-R	GCTGAGATTACCAAAGTGAGGAA	59	
Gor4-F	GGAGAAAGAAGAAAAAGGAAAACC	60	chr7:118932715-118934524
Gor4-R	CATTATTTCCCAGCACAGACA	58.6	
Gor5-F	GGAACAAAGAAAAATGTCTTATGGA	59	chr12:89649593-89650910
Gor5-R	AATCTCTTTGTGCCTCAGTTTC	57	
Gor6-F	CATTGCTTTAAAACTCACTTTGATT	58	chr7_110662045_110662357
Gor6-R	GCAGAAACCAATGAAAGAACC	58	
Gor7-F	CCCAGACTGGTGTCCAACTT	60	chr11_98496017_9849632
Gor7-R	CATTACTGTTCAAAATGTTAATTTGC	58	
Gor8-F	ATGGCCCTCAAGTGATTGTC	59	chr8_38237333_38237643
Gor8-R	TGACACCATGAGACCCTGTC	59	
Gor9-F	TGGAGTCACCTGTTTGCTTG	59.9	chr1_19983001_19983330
Gor9-R	GGATCACTTGAACCCAGGAA	59	
Gor10-F	GCAAGGCCAGTCAGAAACTG	60	chr10_118640620_118640942
Gor10-R	AAAAGAATGGCAAGTTTTCTTCC	61	
Gor11-F	TTGGCATATTTTCATGCCTAAT	58	chr7_73620519_73620821
Gor11-R	TCAAAGGAAATGGTCCCTGA	60	
Gor12-F	TGGCACTTACTCCTATGTCTAAACT	57	chr15_28440037_28440341
Gor12-R	CCAAGTGTAAGTGGCTGAGCTAT	59	
Gor13-F	CAGCCTTTTAACTCCTGCTG	57	chr12_106523736_106524045
Gor13-R	TGACTCTGGTTACTTTTGCCATT	60	
Gor14-F	TTTTACAGAGATCAACTGGCTCA	59	chr13_11699647_11699955
Gor14-R	GCATATGTAGTGGATGTCTGCTTT	59.6	
Gor15_F	TGAGAACATGCAGTATTTGA	59	chr19_34866989_348

	CTTTC		67294
Gor15_R	TTTTGACATTGCACTTTTTCAT	57	
Gor16-F	TGATTGGGAAAAATTTGACTACA	58	chr6_121604525_121604839_
Gor16-R	TTTGGGGGTTTTATTGCCTA	59	
Gor17-F	AGGAGAGGGAGCTGGTATGG	60	chr14:82763727-82764022
Gor17-R	ACTGCAGGAAAGGCGTTAGA	60	
Gor18-F	GGGACAAGAGCGAGACTTTG	60	chr1:204853308-204853609
Gor18-R	AGCACATGACTCCGGATAGG	60	
Gor19-F	GCCAGACTAATAAAGGAGAAAAGAG	58	chr6:60911397-60911701
Gor19-R	TTCTGTGGGATCAGTGGTGA	60	
Gor20-F	TTGGAGAATGTTTCATCTGCAC	60	chr16:18374400-18374707
Gor20-R	GGACAGGGCAATTACACAA	58	
Gor21-F	TTTTCTAGTGTCCCATTTTAGTTCA	58	chr15_61251539_61251846
Gor21-R	CCCAGGGCAACTACCAAAA	60	
Gor22-F	GGTGTCCAGAAAGTTCATCATCTA	59	chr1:132758646-132758947
Gor22-R	CAGAGGGGGACTGTAGTT	57	
Gor23-F	CCAATTGGTTCCATTTCCAA	60	chr7:83827494-83827766
Gor23-R	AAGCAAGATGGTGAATATGAGGA	60.5	
Gor24-F	TGTGACATTATTGTTGTTCTGATTTT	58	chr7: 77324451-77324761
Gor24-R	TGCGTTTGAGAAACATGTCAG	59	
Gor25-F	GGTGAGGCCTTTAAGAGGTG	58	chr6:147497831-147498149
Gor25-R	GGGTGGAAGACAGAAGCAAG	58	
Gor26-F	CACCGCTATACCCCTATCCA	59	chr13:69081094-69081401
Gor26-R	TTGCTTCACATCCTTGCTAA	59.9	
Gor27-F	TGAGCATGGGAAGTTATTCCA	60	chr4:67790814-67791122
Gor27-R	TTCAAGCTTAGAACCAAATCAAT	57	

Gor28-F	GATCTCGGCTCACTGCAAG	59	chr5:43934028-43934338
Gor28-R	ATGCAAAAATTAGCCAGGTG	57	
Gor29-F	CAAAGGGGACAGCACAATTT	60	chr11:88231370-88231680
Gor29-R	GACTCCTGAAATGCTTTTCCTC	59	
Gor30-F	GTGAGAGGAACCAGGCAAAA	60	chr17:880013-880323
Gor30-R	GGTGGGGAAAGAAAATGCT	59	
Gor31-F	CAGAAAAGGGGTGGAAACAT	58	chr3:95264819-95265125
Gor31-R	TTGTTTGTTTGCCCTGTTAGC	60	
Gor32-F	TGCTCTCCAATATGCCTGTG	59	chr12:90894351-90894651
Gor32-R	TGATAGAGAGGACCAACAAAACAA	60	
Gor33-F	GGGAATGTAGCAAGCAAGGA	60	chr6:77494827-77495135
Gor33-R	CAGGGCTAACAGTCACATGG	59	
Gor34-F	GGGAAACCAATGCAGCTAGA	60	chr3:20589548-20589852
Gor34-R	CCCAGTGTTGTGGAGTTTCC	60.2	
Gor35-F	TCCTTTATGGGTTACATAATTTGAGA	59	chr15:77545627-77545929
Gor35-R	GGCTTCCCCAAAATTCAAAG	60	
Gor36-F	TGTCAAAATCCTAAGCTGTCCA	59.7	chr18:14644863-14645173
Gor36-R	TCAGAACCAGCTAAGGGAAAA	79	
Gor37-F	GCTTTGTCTTTTCAAATGTGGAG	60	chr14:44885847-44886174
Gor37-R	CTCCTGAACCCAAAGTGGAG	59.7	
Gor38-F	GAAAGATCATGAATTGTGGTGGT	60.1	chr11:22724816-22725124
Gor38-R	TTTAAGTGAATAATGAAGGCACAAA	59	
Gor39-F	GGAAGCCAAAAGATTGGACA	60	chr3:2457123-2457394
Gor39-R	CGATGCCATCTCATACAGACA	59	
Gor40-F	TTTGATTTGCAACATCCTGA	57	chr5:64532656-64532967

Gor40-R	GCAACTCAAACAAATCAGCAA	59	
Gor41-F	CTCACCTCTGACCCACATCA	59	chr13:34504629-34504938
Gor41-R	TCAGAGACTTCAATGCAACTGG	60	
Gor42-F	TCTTTTGGGAGTTAACCTTTGC	59	chr7:31163413-31163715
Gor42-R	TGGGCAATCATTTCATTCAC	59	
Gor43-F	TCAAGCAACTCACTAACAACCTG	59	chr18:63462410-63462703
Gor43-R	TGTTTGACTTTAGCGAAACAGA	58	
Gor44-F	TGGGCAAAGGTAAACACACA	60	chr7:76893271-76893575
Gor44-R	CATCATTAGGGTGCTTGATGG	60	
Gor45-F	TCAGAAAGATCATTTGGGGATA	58	chr20:60038959-60039268
Gor45-R	TGGACAGTTCTTTGTGATAAGGAT	59	
Gor46-F	TTTTCATGTAACCCCAAAAATTC	59	chr6:166730008-166730313
Gor46-R	TGGTGACATGATCACACACAG	58	
Gor47-F	TCTTAGGGCAAGAGGAAGAGG	60	chr3:12313627-12313937
Gor47-R	TCATTGAGGTCCTTAGATTCCA	58	
Gor48-F	TGTTGGTTCCATTTGTCTTGA	59	chr4:122963632-122963942
Gor48-R	TGCAGAATTTATTCTATTCCACCA	59.9	
Gor49-F	GCCCATCTTGTGCAAATTGA	62	chr10:57062272-57062573
Gor49-R	TTTCTCAAGATGCGAGTTCTTT	58	
Gor50-F	TCTTCCATTTGGCTGTTCC	60	chr11:8869340-8869674
Gor50-R	CCCCTCCTCAGGTTTGGTA	59	
Gor51-F	CTCCCACACCACTAGGTCCA	61	chr10:4399315-4399625
Gor51-R	TGGGCTTTGAAACTGTCAGG	61.2	
Gor52-F	TCTGCTGAAGGGAGATGCTT	60	chr9:56204543-

			56204846
Gor52-R	TTCAATTGCTTCAGTGTGTAAGAA	59	
Gor53-F	TCCCTCATAGAGAATGGAAAATG	59	chr14:33569252-33569554
Gor53-R	CCTGCATTTAAGGTAAGCAGTG	59	
Gor54-F	CCCGAGTAGCTGGGACTACA	60	chr20:31776579-31776886
Gor54-R	AACATGGTGAAACCCTGTCTC	58	
Gor55-F	TCTGCCTGTGTGATCATGTG	59.2	chr12:5281601-5281903
Gor55-R	TCAGGGTTCTCCAAAGAAACA	59.7	
Gor56-F	GCATGCCTTGGAGATATTGAG	59	chr4:178573661-178573968
Gor56-R	CATCCACCAGCCAAAAGATT	59	
Gor57-F	CCTTCAAAGTACCACAGAGTCATC	59	chr4:162666713-162667003
Gor57-R	TTGAGAAAGTTCCTCCCCTCT	59	
Gor58-F	TGTCAGATAACCGTGAAGTAAAAC	57	chr9:17217290-17217586
Gor58-R	TCCCGCTGATCTTTACAAGG	60	
Gor59-F	TCTTGTTATCCTGTTGCTGTCG	60	chr12:20819750-20820038
Gor59-R	CCAGCAACCACACAAACTCA	60	
Gor60-F	GATGGAGACCATCCTGGCTA	60	chr7:133914713-133915011
Gor60-R	CTCGCTGAAACCTCCACCT	60	
Gor61-F	TGAAGCATATGCCATCTTCAC	58	chr3:11931683-11931996
Gor61-R	TCACAGGGCTGCATTAAGAA	59	
Gor62-F	TGATCAGTTGTTGGCTGTCC	59	chr5:99932363-99932663
Gor62-R	CTGAGGCAGGGGAATCTCTT	60	
Gor63-F	CACTGAGAGCAGAAATCCAAAA	59	chr5:17649079-17649390
Gor63-R	TGGAGACCTATTTAACTCAAACCCTA	60	
Gor64-F	TCTAGCAAGTGAAAGTCCTTTATCA	58.7	chr7:94698875-94699186
Gor64-R	GCACCTGGCCAAACTCTTAG	59	
Gor65-F	TTCACTTTTATGCATTCTATT	57	chr11:88624056-

	CCA		88624363
Gor65-R	TCCCATACCCATCTTGCATT	60	
Gor66-F	TGCAAGAGGGATAACAGACCT	58	chr3:99686701-99686987
Gor66-R	TGGTAGGAAAATCCGAGCTCTA	60	
Gor67-F	TGAAAAGGCTGAAAAAGTAAATACA	58	chr8:112413990-112414292
Gor67-R	GCAGTTTCTCATTCTGCATGG	60.8	
Gor68-F	ACCATTGCACCCAGATTAGC	60	chr11:104840865-104841162
Gor68-R	TGGAAGAACTGCTTCAGGAT	57.5	
Gor69-F	CCTGCACACACATACCCAGA	60	chr11:38356201-38356496
Gor69-R	TGGTTGGGGCAACTACATTT	60.2	
Gor70-F	CATACACACAGCCAAGAAGCA	59.9	chr5:84930313-84930622
Gor70-R	CCCTTTTCTCCACAACCTCA	60	
Gor71-F	TCCCCAATCAGATGAAGAGC	60	chr7:82645952-82646258
Gor71-R	ATGTGGAGGAAGTTGCATGG	60	
Gor72-F	AGCCCCCGTCTTCTATGTTT	60	chr9:97862133-97862429
Gor72-R	GACATGAATGGATATCAAAAATGC	60	
Gor73-F	AAGGATGAGCAGGATTTGGT	58	chr1:91601670-91601931
Gor73-R	TTGCAGGGGACAAAGACTTC	60	
Gor74-F	CTGACCTTGTGATCCACCTG	59	chr6:6337556-6338869
Gor74-R	TGCCACTGTATTCCTCTAGCC	59	
Gor75-F	GGGAGCTGCTAAGAATGCAC	60	chr11:48176812-48177104
Gor75-R	CCTGGGATGCTCGTTTCTAA	60	
Gor76-F	CATCACAGACCCGTACCTCA	60	chr3:11919992-11920280
Gor76-R	AACTAGCTGGGTGTGGTGGT	60	
Gor77-F	GTGAAAACTTTGTGTCATGTTGC	59	chr15:19924516-19924818
Gor77-R	CCCATGGGACTCACTCTTTG	59	
Gor78-F	CCCGTGGACATGAAAAAGAG	60	chr6:137468192-

			137468492
Gor78-R	GGATGAGACTTTGGAACGTGA	60	
Gor79-F	TGGCGTAGCTACTAAGGAAAATG	59	chrX:89680916-89681211
Gor79-R	AGAATGGTTGGGTCATATGGT	59	
Gor80-F	CCACAGGCATATGGCACTAC	59	chr1:15824206-15824499
Gor80-R	CCTCAGGTATGGTGGCTCAT	60	
Gor81-F	CTGCTGGGCTCAAGCAAT	60	chr1:40843796-40844111
Gor81-R	GCCTGGGCAACATAGTGAGA	60	
Gor82-F	CATTTGAAAGCCAGCCTGA	59	chr5:147529181-147529478
Gor82-R	GCCATTGTGTGCTTTGATAA	57	
Gor83-F	CCTAAAATCCACTGCGGATAA	59	chr12:39433053-39433349
Gor83-R	CATCATTACAATAGCAGGGCTTT	59	
Gor84-F	GAGCTTTTGCCAGTCTCAGG	60	chr10:7742039-7742331
Gor84-R	CCAAAAAGGAGAGGTTCAAGG	60	
Gor85-F	ATGCCTTAGGGAGGAACAGT	57	chr1:180818620-180818906
Gor85-R	TTTATGTACTGTGTGTGTTTTTGCTT	59	
Gor86-F	TTGCTTTGGTGTCAGGGTAA	59	chr4:109685443-109685753
Gor86-R	CCTATGAAGAAAACCCCAAGC	60	
Gor87-F	ATGGTGTAGGCCATGGAAAG	59	chr8:138485277-138485586
Gor87-R	GAGGGCTTGGGATCTGAAAT	60	
Gor88-F	GCAGAAAGGGAAGCAAACAC	59	chr16:2991494-2991805
Gor88-R	TAATTGAGTCGTGGGCAGTG	59	
Gor89-F	TGCTCCCTTCAACCTTTCTG	60	chr1:18251677-18251959
Gor89-R	CCTTGTCTCAGGCTCTGCTT	59	
Gor90-F	GGTAGCACATCAATGTGAAGG	58	chr1:192919379-192919652
Gor90-R	TGACAGGTACTTTTTCCAGC	58	

	A		
Gor91-F	GGGAAGTGTGCAAAAGGAGA	60	chr20:5201905-5202214
Gor91-R	AGCCAGCAGATTGTGGGAAT	62	
Gor92-F	TGTGAAGTCAGCAGGGTCAC	60	chr18:75338859-75339161
Gor92-R	CCCGAGCCTTCTACTCTCCT	60	
Gor93-F	TAGAAAGCAGGGAAGCCATC	59	chr14:47878338-47879644
Gor93-R	CAAGTAATGAGATCAAAGCTGTAATAA	57	
Gor94-F	GCCAAAGTGAGCGCTTTAAT	59	chr9:50448314-50449620
Gor94-R	CAAATGGCCAGAAAACTCAA	58	
Gor95-F	TGTGCTCCTTAATGTGGGTTT	59	chr6:152443003-152443309
Gor95-R	AAGCAGGGAGAACTCTTACCC	58	
Gor96-F	TGAAGTGGTATTCCTGACTCCTT	59	chr3:194177227-194177534
Gor96-R	GCCATAGCCAAAACACACCT	60	
Gor97-F	AAATTTGAAATCTGTGCATTTTTG	59	chr1:188667451-188667758
Gor97-R	ACAGTGTTGGCCTGAAGACA	59	
Gor98-F	CAATTTGGATCTAGGGAAGCA	59	chr7:35433537-35433845
Gor98-R	TCATATCAAAGGGCTGCACA	60	
Gor99-F	GTTGAGGCTGCAGTAAACCA	60	chr21:30586879-30587188
Gor99-R	TGCCACCTGACACATCCTTA	60	
Gor100-F	GCCTCCTTGTTTATTGAATCAT	57	chr12:84729091-84729400
Gor100-R	TGGGTGTTGGCTTAGAACAG	58	
Gor101-F	ACAGAAAAAGGCTGGGATGG	61	chr1:10,090,560-10,090,847
Gor101-R	TTTTCAATGGCTCAAGTTTTCA	59	
Gor102-F	CACACAGAAAATTAGGAACAGAGC	59	chr9:30423357-30423666
Gor102-R	GGAAACATACAGGAAGCAAAAA	58	
Gor103-F	CCCCAGGATGTGTCCTGTT	60	chr5:140512255-140513563
Gor103-R	TTGTCAAGATTTTGGTATCTG	59	

	GAG		
Gor104-F	AATCATGGCAAAACCACCTC	59	chr6:131770434-131771742
Gor104-R	GCTTAGTAAAATGGCAGCTCTT	57	
Gor105-F	CCCCTTCCAGATCTGAGTCC	59	chr6:133656907-133657215
Gor105-R	AAGAAAGATAGTAAGGGCCTAATTC	57	
Gor106-F	CAAACTTGCAAAAGGCACAA	59	chr7:10504797-10505105
Gor106-R	AATTCTTCTTcTTGTGGTGAGAA	57	
Gor107-F	TGTGAGACATGGAGTCAAAGG	60	chrX:70533820-70534128
Gor107-R	CCATGCAAGTCTGAAATTGTG	59	
Gor108-F	AAAAGAGTGAAACTCTGTCTCAAATA	57	chr12:2133625-2133933
Gor108-R	TAAAACCATGCTTGGCCACT	60	
Gor109-F	GAATGACGCTGCTTGTCTGA	60	chr7:122655755-122656062
Gor109-R	TCTGCCTTTGTTCTTCCAAAA	59	
Gor110-F	GGCACTGTACTGGGTGCTAAA	60	chr11:104569843-104570150
Gor110-R	TTTGAATAAAAGAAACCCATGAA	58	
Gor111-F	GCTGCTCTTGCGATAGTGAA	59	chr17:91179111-91179418
Gor111-R	GCATCACAATCATGGCAGAA	60	
Gor112-F	GCAATGGAGTTGTGGCTTCT	60	chr6:115513487-115513790
Gor112-R	GAGACTGCATTTTCCAATGTGA	60	
Gor113-F	CAAAAAGTAGAGTTCCAAGCCATT	60	chr8:19352159-19352462
Gor113-R	GACCTGTCCTACAAGAAATCCAA	59	
Gor114-F	TGACGCTGTAGTCAATGCTG	58	chr8:82767352-82767655
Gor114-R	AGTGAAAATGGGGCTGTTTG	60	
Gor115-F	CCAAAAGCCTAGACTTCACCA	59	chrX:38642951-38643254
Gor115-R	ACTGGGGCCTATTTTGAAGG	60	

Gor116-F	TGGTTTTAAAATGGACCTGGA	60	chr10:28984298-28984601
Gor116-R	CACCCCACACTCATTTCTCTT	60	
Gor117-F	AGCCATGGAGATTTGTGAGG	60	chr10:99041313-99041616
Gor117-R	TTTGCATACCTGTGCCTGAG	59	
Gor118-F	AAAGATTCAAAATTTCTCCTTGTAA	58	chr12:4060434-4060737
Gor118-R	AAAAAGTCTGAGGCAAAAAGTGA	59	
Gor119-F	GGCAAATGCTTTCCCAGTAA	60	chr12:66501882-66502185
Gor119-R	TCCTTCCTCAAAGAGCTGGA	61	
Gor120-F	TCCACACCACAAAGACAAGA	57	chr13:90506534-90506837
Gor120-R	TTTTGCTTCTTTTCTTTAATTACCTTG	59	
Gor121-F	CGGCTTCTAGGGTTGGAACT	60	chr9:35076890-35077175
Gor121-R	ACTGGATCAGGCTGAGAGGA	59	
Gor122-F	GCAGCAGTGAGCCAACATC	60.5	chr3:54134621-54134902
Gor122-R	TCTGTCTGCAAATGTATAGTCAAGC	59	
Gor123-F	GCTTGATTGAGGGGTCATGT	60	chr14:71365071-71365323
Gor123-R	TCTAGGGGCAATGTAAAATGG	58	
Gor124-F	TGACTCACAGGTGCTTCTGG	60	chr15:40503129-40503381
Gor124-R	CAGCATTTGGTGGTGTCAGT	59.6	
Gor125-F	AAGCAATTCCCTGCCTCAG	60.3	chr22:25391880-25392132
Gor125-R	AGAGTGAAGTGGCCTGGTGT	59	
Gor126-F	TGGATTCACTCTGGTCAAACA	59	chr7:116606240-116606500
Gor126-R	AATGAATTTATTCCTTTATTTGGTGT	57	
Gor127-F	TAGGCCCAGCAACTCAGG	59	chr1:56387832-56388132
Gor127-R	GCCCAGAAGTGGACAGGAT	60.1	
Gor128-F	TCCCACTGATGAACCTGACA	60	chr12:88522326+88522567

Gor128-R	CAAACTGTCCTTTTTGATCTCTAGC	60	
Gor129-F	AAATCCAGAACTTTCTCAATCCA	60	chr6:23300087+23300683
Gor129-R	CCAATTAGCTTCAGGAATCAAGA	60	
Gor130-F	GCCTTTAGAACCTTGCCAGA	60	chr20:37908239+37909219
Gor130-R	GGGCTTGGAGAGTGACAAAA	60	
Gor131-F	acccacacacatccttccat	60	chr9:86937262+86938233
Gor131-R	CATAGAATGGGCTTTATAggctct	60	
Gor132-F	GCTTTGCATTTCTCATAACATCC	60	chr20:8018594+8019191
Gor132-R	cagaagctgcattacccact	60	
Gor133-F	CCCAACACACACACACGTTTA	60	chr20:29831098+29831741
Gor133-R	TGTTTTTGGGGGATGCTATG	60	
Gor134-F	CTCCTCAGGGAAGGCTTGAT	60	chr3:38917797+38918642
Gor134-R	GAGGGGAGGAACATGTCAAC	60	
Gor135-F	ACCTCCCACTAGAGCAAGCA	60	chr3:19310863+19311456
Gor135-R	AGAGAGCACAGTGCAAGCAA	60	
Gor136-F	agggaagtggtgtgaCTAGAGA	60	chr4:48299948+48300247
Gor136-R	CCACAACTTCGTGACCCATT	60	
Gor137-F	TCCCTAATACCATCATTCTTGG	60	chr3:92933098+92933597
Gor137-R	TCAAGAAGAACTAGGGGGATAAAA	60	
Gor138-F	TGCTGTGTACCAACAAAGAACA	60	chr1:146942297+146943142
Gor139-R	GGGTCCATTTATAGCATACTGGTT	60	
Gor140-F	AAGCAGGTTAAGCTCTTATTTTATCA	60	chr4:142945303+142945767
Gor140-R	ACAAATAACCTAGGAACACATATGAA	60	
Gor141-F	TTCGGAAGCCAGGAATTTTT	60	chr4:19388401+19389038
Gor141-R	TTCACATCAAGGTGTTTACTCTCAA	60	

Gor142-F	AAACGACTCCTCCCCTCAGT	60	chr4:151202455+151203289
Gor142-R	TGACTAGTTTGGGAATTGAAGTG	60	
Gor143-F	CATTAAGCAATTAACATTTCAAGC	60	chr4:162666523+162667222
Gor143-R	CCTTTTCTTGTCTTTGTGCTTT	60	
Gor144-F	GAAGGAGATTGGCAGAAGGA	60	chr4:123122676+123123174
Gor144-R	TTTTGAGATGTACAAAAATGCAG	60	
Gor145-F	TGCTATTTGATTTCAGGCAATTT	60	chr4:25523440+25523812
Gor145-R	TGTTGCTCAATGTTGGTGCT	60	
Gor146-F	TTGCATGAAAGGGTTCCTTC	60	chr4:192717402+192718073
Gor146-R	CGGTGCTAGGAATTGGACAT	60	
Gor147-F	TTTTTCTTTGGCATATTTCCAC	60	chr5:76992448+76992937
Gor147-R	CCCATCCTTTAGCAACCAAA	60	
Gor148-F	GCGTTAATCAGGCTTTTCAA	60	chr5:120001922+120002297
Gor148-R	CCAAAGGTCATAACCAACTCAA	60	
Gor149-F	AGGAGGCAGGGGTGTTTCTA	60	chr5:136764395+136764882
Gor149-R	AAGTGGGGGAAGGTGATAGG	60	
Gor150-F	CATATTTCACAATATCACACTTTTCTT	60	chr5:124902611+124902906
Gor150-R	GGAGCAATTGTTATTTCTGTGTG	60	
Gor151-F	AATTGTTGGCCTGTGCTAGG	60	chr6:64889803+64890440
Gor151-R	CCTCTTGATGCCTTTATTGAGA	60	
Gor152-F	CAGAAAGTTCTGATGGGCAGT	60	chr6:47817914+47818405
Gor152-R	AAGCAGCAGCAGGAGTCAGT	60	
Gor153-F	GGTGTTTTGCTTGTGCTTGA	60	chr6:126708598+126708943
Gor153-R	CCTTGCAAATAACCCAACAGA	60	
Gor154-F	GCATCTTGCTCTGGATTGAA	60	chr6:68235517+68235976

Gor154-R	TTGCAAGTGGCTTTATGAAAA	60	
Gor155-F	CCTGAAAAATCAATCATAAAAATGTG	60	chr5:13684008+13684587
Gor155-R	CGGTAACTTGTTCTAGTGGTTGTC	60	
Gor156-F	TCCTGTAAAGAAATGAGTGTTGGA	60	chr6:142287991+142288462
Gor156-R	TTAGCAATGGCACAAACCAG	60	
Gor157-F	GGTTCCAATCATATTACTTTTTCAA	60	chr4:162950069+162950640
Gor157-R	CTGGAAAAACTGGGGTGTGA	60	
Gor158-F	TTGGGACCACTACTTCTGCAT	60	chr6:41235694+41236056
Gor158-R	TGAATGTTCCTTAATAGTCTCAAAGG	60	
Gor159-F	CAAGCTAAGAGAAGGATTCTAGAGGA	60	chr5:87391712+87392310
Gor159-R	TTGGAACTGAGGGAGAAATACC	60	
Gor160-F	GCTTGCAATCTTATCTACAGCTTTT	60	chr6:17244192+17244791
Gor160-R	AATCATTTTTCAGTGGAAAGCA	60	
Gor161-F	GGAGAACTCAAAGAAAATAAACTGTAG	60	chr4:165286814+165287577
Gor161-R	CAGTGAGCCCTGTGGAATTA	60	
Gor162-F	GAAACTGTAGGGAAAGGAAAACA	60	chr10:118640440+118641039
Gor162-R	AAGGAAGTTTTGTCCCATGA	60	
Gor163-F	TCTGCAAGGAAACCAAAAGG	60	chr10:74362127+74362976
Gor163-R	GGATGGAGAGAGCCCAAGG	60	
Gor164-F	GCACTGAGACATGGTGAGGA	60	chr10:1902548+1902998
Gor164-R	AGCAGGAAATCACCTTGGAT	60	
Gor165-F	GCAGAAAGGAAGAAAATTCCAA	60	chr10:120948782+120949268
Gor165-R	TCACTTGCTTTTCTGCCTAAG	60	
Gor166-F	TGTGCAACTGAAATCAACTGG	60	chr10:88746451+88746850
Gor166-R	TTTATTTGCTCACTCTCACCAC	60	
Gor167-F	CACCTTGAAGCAGTTTTCTCT	60	chr10:98567270+9856

	G		7860
Gor167-R	GAGGAATCAAACTTGACTCAAAATC	60	
Gor168-F	GTTTGTTGCCATCCTTTTGC	60	chr10:125229275+125230103
Gor168-R	ACCAGAGGAGCACATGAAGG	60	
Gor169-F	TGACATTCCCCACTCTGGA	60	chrX:138046972+138047378
Gor169-R	CCCCTGCATAAGTGGCTTTA	60	
Gor170-F	GAATGATAGGCTTGTCAGCTTG	60	chrX:74538519+74538891
Gor170-R	CTTATGGCAGGAGGCAAAAC	60	
Gor171-F	AGATTGTTTCATCGTCCAGGT	60	chr10:101741838+101742224
Gor171-R	ATACAAAACCCAGCCCTTGA	60	
Gor172-F	TCCAGTTTGGGCAAAAGTCT	60	chr10:31662484+31662954
Gor172-R	GATGTTCACACAAAATTAGAATGGA	60	
Gor173-F	GAAAATATGTCTTCAGGGTCGTG	60	chr10:125223387+125223984
Gor173-R	AGGAGATGTTAGGCCAATGAA	60	
Gor174-F	TTTCCAAATAATGCCCTATAGATAATG	60	chr10:32331998+32332490
Gor174-R	tcctgacctcaaatgatctgc	60	
Gor175-F	GGATATGCGTGGACTTCCTG	60	chr11:119568631+119569030
Gor175-R	CCTTCAATTACCTCCCTATGC	60	
Gor176-F	TCCAGCCTTCAAAAATGAGC	60	chr11:103850251+103850646
Gor176-R	CAGACAGTTCTATCAAACCTACAAGG	60	
Gor177-F	GCTGTAACAAAAATCTTTGCAAGT	60	chr11:15479508+15479996
Gor177-R	AGACAACCAGCAGCCAGAGT	60	
Gor178-F	TGTGTCTTTTGTTTTTCACTCTGG	60	chr11:96651535+96651910
Gor178-R	TGGATCTCCCACTTCAGCTT	60	
Gor179-F	GGAAACAAAGATGCAGATTCTTA	60	chr12:15338549+15339105
Gor179-R	TCATTTCACTGGCAAACCAA	60	
Gor180-F	ATTTATGTGGAATTTTCATTT	60	chr12:80009677+8001

	GC		0147
Gor180-R	CCTCTTGCCAGATACTGATGC	60	
Gor181-F	CTGATGTGTTCCTGTCCTGGT	60	chr13:7889120+7889489
Gor181-R	CTAGAGACATTTAGGAAGAAACAGC	60	
Gor182-F	GGACAAAGGGCTGGAAGC	60	chr13:16526622+16527021
Gor182-R	GGCAGCTTTTTCTCCAGTGT	60	
Gor183-F	AAACCAGTATTGACATTTTGGAG	60	chr12:92826723+92827120
Gor183-R	GCATTTAGCCTTAAGAGTGTTTCC	60	
Gor184-F	AATTCACCTATTTGGGCAACA	60	chr13:93817077+93817574
Gor184-R	CTTAGTGGCCCTGTCACACA	60	
Gor185-F	GCCCCTTTGCCTCTTAAAAT	60	chr12:53073007+53073506
Gor185-R	AGGAGACAAAAGAGACATGAAAA	60	
Gor186-F	ATTCCCATTCCCACTGATGA	60	chr12:86997128+86997889
Gor186-R	TGGTAGCTGTCAATGCAAAAG	60	
Gor187-F	TGCATTTCAGGACCTCCATT	60	chr12:25824289+25824637
Gor187-R	GGACAGACAGAATGCAACAGG	60	
Gor188-F	CAGATTACTTTGCCCCTTGC	60	chr13:19683906+19684349
Gor188-R	ACCTCTCTCAAGCCAATCCA	60	
Gor189-F	AGAAAAGACAGCCCAGGAGA	60	chr12:95398360+95399073
Gor189-R	TTTGCCATCAGTTTTTCAGG	60	
Gor190-F	AGAGATCAACTGGCTCAGCA	60	chr13:11699603+11699976
Gor190-R	TGGATGTCTGCTTTTCAGTCA	60	
Gor191-F	CATCATGAGCAAGTAAGAGTTTACAA	60	chr14:33437737+33438226
Gor191-R	TCACTGTATGTTGTAAGTTCCTTTGA	60	
Gor192-F	TCCTAAGTTCAAATCCCTGCTC	60	chr14:49333235+49333866

Gor192-R	TCAGGTATAAAAGGGGGATGAA	60	
Gor193-F	AGGGAATCACTAAACAGAAAGTATG	60	chr14:65919222+65919990
Gor193-R	TGCACATTGGTTTGTAATCTTTTT	60	
Gor194-F	ACCATATGCCAGGTGTAGGC	60	chr15:12586751+12587337
Gor194-R	ACCTGCCAAAAATCATGAGC	60	
Gor195-F	TGCCAGTAACATACAGCCTCA	60	chr15:52990569+52990946
Gor195-R	GGAAAACAAAATGAGCAATGAA	60	
Gor196-F	TGGGAATGGGGATACTCAAG	60	chr15:72960978+72961422
Gor196-R	TGTAGCAGTCCAGCTTTACCC	60	
Gor197-F	TTGGGGACTTTTCTCTCTGG	60	chr16:12976623+12977015
Gor197-R	AAGGAAAAACAGAGAAGCAGGA	60	
Gor198-F	CCCAACCCACACACAGTCTA	60	chr16:67485180+67485562
Gor198-R	GAAGAAAGGAAGGAAGGCTTAAA	60	
Gor199-F	ATGCTCTTCAGGCACCATTC	60	chr17:21117608+21118151
Gor199-R	TGTAAAATAGAGGAGGTGTGTCC	60	
Gor200-F	AACCAGACTTGACCAGTACACAA	60	chr17:34116767+34117283
Gor200-R	TTGTGGTTACTATTGAGTGGTTAGA	60	
Gor201-F	GGAAACTGGCAAGGAAGACTC	60	chr17:35215662+35216138
Gor201-R	GGGAAAAGGAGGGAGTGATT	60	
Gor202-F	AAAAAGCTGGAGATGGCTGA	60	chr17:40335651+40336096
Gor202-R	CATGAACAGCAGCAAACACA	60	
Gor203-F	TTTTATGTTTATGGGACTTTTTGTT	60	chr17:78960685+78961137
Gor203-R	GTGCAGAGGAGGGCAATG	60	
Gor204-F	GAGGTCAGATCACAAAATAGCC	60	chr18:57328604+57328706

Gor204-R	TGGAATCTGTGAATCTGCATTT	60	
Gor205-F	CCCAGGTATTCCTGCTTCAG	60	chr19:2204746+2205391
Gor205-R	AGGATTGTCAAAATGCCACA	60	
Gor206-F	CTGGCAGAGCAAGTCACACT	60	chr20:12206877+12207361
Gor206-R	GGAGGGAGAGTAAGGGATGG	60	
Gor207-F	AGCCCCACTAAGGCACAGT	60	chr20:19133712+19134087
Gor207-R	tgatcattgcactccagctt	60	
Gor208-F	TCCCAAACCTGAAAAGGACA	60	chr20:25510471+25510900
Gor208-R	TCTGGAATAAGCCCAACTTGA	60	
Gor209-F	ATCATTTGCTGGGGATCAAT	60	chr20:2856712+2857141
Gor209-R	TTGGTAGCCAGGAGTTGGAG	60	
Gor210-F	GGCATTGAGGGAAAACTGAC	60	chr20:62601213+62601696
Gor210-R	GGATCTAAGAGGTGAAGGTCTAACA	60	
Gor211-F	GGGAGGGTCAGGCTCTATGT	60	chr21:13386099+13386586
Gor211-R	AAATTTACAAGGATGAAAGGAATAGT	60	
Gor212-F	TGTCCTTAACTTATTTCTTACCTTCC	60	chr21:2490196+2490577
Gor212-R	TCAAGAAAGCTGCAAAGGAAC	60	
Gor213-F	TCCCCACTCTAAAGTTCATGG	60	chr22:11043980+11044429
Gor213-R	TGCTTGACTTCCGAATTCCT	60	
Gor214-F	TTGCTATAGACCCCCTTTTCC	60	chr2A:29047955+29048602
Gor214-R	TTGGCAAGGAAAAATATCTATGA	60	
Gor215-F	ACTGCTCAGCCACACTACCC	60	chr2A:43095269+43095682
Gor215-R	GGTAAAGCAGAATGTAAGTCAAAAA	60	
Gor216-F	CAGGATGTGTTTCATAAGATTTTCC	60	chr2A:55705974+55706308
Gor216-R	GCAGAATGCATACTAGAGGA	60	

	TATTAAG		
Gor217-F	GCATGGATAGTCTTGAAAGTCC	60	chr2B:118114685+118115181
Gor217-R	ATAGCATGGCTGTCAGAGCA	60	
Gor218-F	AAAATCTGTGCCTTAAGAAAGGT	60	chr2B:128129101+128129700
Gor218-R	GGAAGTTCAGAGGGAAATTAAAG	60	
AluYa1b4_Gorilla_1F	TTTGTGGTGACTTATTTCTCATTC	60	chr3:82546758-82547074
AluYa1b4_Gorilla_1R	CAGAGCAGGGGATATCCAGA	60	
AluYa1b4_Gorilla_2F	TCAGTAACCTCATTTGCCTATTTC	60	chr8:90568974-90569285
AluYa1b4_Gorilla_2R	TGCTATTAGTAAAAACATCTGCTTAAA	60	
AluYa1b4_Gorilla_2F	AGCACAGTGCCTGACACAAA	60	chr5:122689054-122689364
AluYa1b4_Gorilla_3R	TCGCTGCTTTTGCTTATCCT	60	
AluYa1b4_Gorilla_4F	TTGTTAATGGTCATTTGGGTTG	60	chrX:43554127-43554437
AluYa1b4_Gorilla_4R	TGCTGCCTATAAGAAACATGC	60	
AluYa1b4_Gorilla_5F	TCAGTGGATCAGCAAGAGAAGA	60	chr2A:79914361-79914671
AluYa1b4_Gorilla_5R	AATTCGCAAAGGCTGAAGAA	60	
AluYb3a2b2_Gorilla_1F	TTCCTCATCCCACTCTAGCC	60	chr5:27108508-27108852
AluYb3a2b2_Gorilla_1R	GGCTAATTTGTCCCAATCCA	60	
AluYb3a2b2_Gorilla_2F	TGTTCCATAATCCCGTTTTGA	60	chr2B:40409938-40410241
AluYb3a2b2_Gorilla_2R	AAAGTCTTACAAAGACAGCTAATCCA	60	
AluYb3a2b2_Gorilla_3F	TTGTACCGATTCTATGCTACACA	60	chr4:134326052-134326345
AluYb3a2b2_Gorilla_3R	TCCTTCCCTCCCTCTTGTCT	60	
AluYb3a2b2_Gorilla_4F	TGTTCCATAATCCCGTTTTGA	60	chr22:15752000-15752328
AluYb3a2b2_Gorilla_4R	CAGCTAATCCATTAAATGAAACTCA	60	
AluYb3a2b2_Gorilla_5F	TCAACTACAAAGGGGTGAAGC	60	chr9:82102531-82102837
AluYb3a2b2_Gorilla_5F	CAAGTGATCCTCTGGCCTCT	60	
AluYb3a2b2a2_Gorilla_1F	TTGGAAACACACTTTTTGAAGA	60	chr9:44243497-44243790

AluYb3a2b2a2_Gorilla_1R	TATGTTCCCAAACCCCTGAA	60	
AluYb3a2b2a2_Gorilla_2F	TAAGCAAGCCTGGGCAAT	60	chr4:136757835-136758112
AluYb3a2b2a2_Gorilla_2R	TACTGCGCTTTTCCGATGG	60	
AluYb3a2b2a2_Gorilla_3F	CCATTTGGCTGTTCCTGAGT	60	chr11:8869340-8869674
AluYb3a2b2a2_Gorilla_3R	CCCCTCCTCAGGTTTGGTA	60	
AluY16a4_Gorilla_1F	TGAGTTGTGACGGAGACTGC	60	chr2A:33686556-33686846
AluY16a4_Gorilla_1R	TTCTGACATCAAGGATTTACACTGA	60	
AluY16a4_Gorilla_2F	TGGGAACAAAAACCAAAGAGA	60	chr4:111551744-111552035
AluY16a4_Gorilla_2R	GAACAAACAGTAACAGCAGAATCTAA	60	
AluY16a4_Gorilla_3F	GCAGGGCTCAGAGACATAGG	60	chr15:30752309-30752581
AluY16a4_Gorilla_3R	CAGTTTTGGGGTATGTTTTCTCA	60	
AluY16a4_Gorilla_4F	TGATTTGGAAAAGGCAATTAAG	60	chr7:87000093-87000369
AluY16a4_Gorilla_4R	GAGTAGCTCTCTGATGGTCTTTCTT	60	
AluY16a4_Gorilla_5F	GGCTTTCTTAATTCGGTCCTG	60	chr5:130725484-130725760
AluY16a4_Gorilla_5R	ACCCTGCCTGGCACATAATA	60	
AluY16a4_Gorilla_6F	GCCACTAGTAAGGACAAAAATAATCA	60	chr8:95098968-95099246
AluY16a4_Gorilla_6R	CTCCTTTCCTTTAAAACCAATCT	60	
AluY16a4_Gorilla_7F	CACAGTGGAGTCCCATCTCC	60	chr20:57457092-57457372
AluY16a4_Gorilla_7R	CTTCCCCATAACCCAGCAT	60	
AluY16a4_Gorilla_8F	GGTGGCAGGCTCCTGTAAT	60	chr10:125229580-125229840
AluY16a4_Gorilla_8R	TCTCAGCTCACTGCAACCTC	60	
AluY16a4_Gorilla_9F	CAACCATATGAATGGGTGTATAGC	60	chr3:81783708-81783993
AluY16a4_Gorilla_9R	GAAGGCAAATAAGCACATGAAA	60	
AluY16a4_Gorilla_10F	CTTTTAATGGCCAAATGATGA	60	chr3:58821998-58822278
AluY16a4_Gorilla_10R	TAGCTGGAATGCCTTTCGAT	60	
AluY16a4_Gorilla_11F	TGCTCCCTTCAACCTTTCTG	60	chr1:18251677-

			18251959
AluY16a4_Gorilla_11R	CCTTGTCTCAGGCTCTGCTT	60	
AluY16a4_Gorilla_12F	TTCCTACCTAGAAGCAGGAGCTTA	60	chr15:39227974-39228250
AluY16a4_Gorilla_12R	TGGAACACTTAATGGATACAGCA	60	
AluY16a4_Gorilla_13F	TTTCAAAGCAGAAGAATGAACA	60	chr17:9327125-9327412
AluY16a4_Gorilla_13R	TGAGAAGTTTGAGGGAGAAAGC	60	
AluY16a4_Gorilla_14F	GCCAAAGCAGGATATGTTTGT	60	chr10:14559545-14559837
AluY16a4_Gorilla_14R	ACAGGGTCTCGCTATGTTGC	60	
AluY16a4_Gorilla_15F	TTTGTAGAAGCCTTCCTGTATCTTC	60	chr3:11931343-11931622
AluY16a4_Gorilla_15R	AGCTTTGCGACCTTGAGAAA	60	
AluY16a4_Gorilla_16F	GATCACGCCATTGCACTTC	60	chr19:14055848-14056119
AluY16a4_Gorilla_16R	CTGGTCTCGAACTCCTGACA	60	
AluYa1_Gorilla_1F	CACACATTAAGAAACAAGCAACC	60	chr8:118735963-118736282
AluYa1_Gorilla_1R	CAGCATTCTTGTTGGTGGTG	60	
AluYa1_Gorilla_2F	GAGACCAGCCTGGACAACAT	60	chr5:62671042-62671356
AluYa1_Gorilla_2R	GAGTGTGGTGGCGTGATCT	60	
AluYa1_Gorilla_3F	AACCCAAATGTCCAACAATGA	60	chr3:99864231-99864542
AluYa1_Gorilla_3R	ATACCCCACCCCAACAACAG	60	
AluYa1_Gorilla_4F	GCAATGTGGAGGCTATTGGT	60	chr12:104209566-104209877
AluYa1_Gorilla_4R	CCAACATGATCACACAAAAATTG	60	
AluYa1_Gorilla_5F	CCATAGCAGCTGCGTACTTC	60	chr2A:8536394-8536705
AluYa1_Gorilla_5R	TTTCTACTGGGCACAATGTAAGA	60	
AluYa1_Gorilla_6F	TTTGCAATATTTTATCGAGGA	60	chr4:109685443-109685753
AluYa1_Gorilla_6R	CCTATGAAGAAAACCCCAAGC	60	
AluYa1_Gorilla_7F	CATGGAACATTCTCCAAAATCA	60	chr6:50912642-50912952
AluYa1_Gorilla_7R	TTGCTTAGTTTCCGTGCACTT	60	
AluYa1_Gorilla_8F	GACCCCACTTCTTTCTTCCA	60	chr4:11353091-

			11353400
AluYa1_Gorilla_8R	GGCATGTTTTGACCTTTCCA	60	
AluYa1_Gorilla_9F	CCAGCTTCATCCATGTTCCT	60	chr8:127010574-127010883
AluYa1_Gorilla_9R	AACCAACCCAAATGTCCATC	60	
AluYa1_Gorilla_10F	ATGGTGTAGGCCATGGAAAG	60	chr8:138485277-138485586
AluYa1_Gorilla_10R	AGCAACCATCTGGGGTGATA	60	
AluYa1_Gorilla_11F	AAGGTTAGGTTAATGTGCTAATAGGTT	60	chr13:46248181-46248488
AluYa1_Gorilla_11R	AATAAAGTTCAAATGGCATAGTGAT	60	
AluYa1_Gorilla_12F	TGTTATTAGAATGTTCAACACCAAGA	60	chr15:2087765-2088072
AluYa1_Gorilla_12R	TGGAAAACTCTGATGCCAAA	60	
AluYa1_Gorilla_13F	GCAAGAGCTTGGAATATCTTATACCT	60	chr2B:14371713-14372020
AluYa1_Gorilla_13R	TGTCCTCAAGGATTTTATTCTATGG	60	
AluYa1_Gorilla_14F	CAAAGTTGGTCCTCAGTCGAG	60	chr7:63140404-63140710
AluYa1_Gorilla_14R	CCAAGCAGTCATTATCCCACA	60	
AluYa1_Gorilla_15F	CAAAATACAATGCGCCACAC	60	chr4:168677986-168678291
AluYa1_Gorilla_15R	GGAATTTGAACCTGGGTCTG	60	
AluYa1_Gorilla_16F	TGATTCACATGAACTTACTAATGTGCT	60	chr2A:110643863-110644168
AluYa1_Gorilla_16R	GGGAGCCACAGAATCAAAA	60	
AluYa1_Gorilla_17F	GAGGTAGAAGCACAAAATATCTCCA	60	chr2B:86106658-86106963
AluYa1_Gorilla_17R	CACCTCCAGCAGTATTCCTCA	60	
AluYa1_Gorilla_18F	GGCTGCTTCTGCCAGACC	60	chr1:65716835-65717139
AluYa1_Gorilla_18R	GGGGCTTCTAATAATTTAGTTCTGA	60	
AluYa1_Gorilla_19F	TTATGCAGGCAAGCTCAATG	60	chr18:39124190-39124494
AluYa1_Gorilla_19R	ACTGGACAGAGCGGGACTTA	60	
AluYa1_Gorilla_20F	TCTTCTGTTAGTTTTCTGAGGGTAGA	60	chr3:158404337-158404639
AluYa1_Gorilla_20R	CAAAAATATTCATCATCTGTGACCA	60	

AluYc3_Gorilla_1F	TCCCCATGGAAGCACTGTAT	60	chr5:114760753-114761063
AluYc3_Gorilla_1F	CATGCTGCACCAGGGAAG	60	
AluYc3_Gorilla_2F	AACTAAAGTTCCCTGAACCATCC	60	chr13:20968400-20968710
AluYc3_Gorilla_2R	GGGTCTATAGAGAATGCACACCA	60	
AluYc3_Gorilla_3F	TTTGGTGGGGCTAGAGAAGA	60	chr18:29325002-29325312
AluYc3_Gorilla_3R	AAGAATACTTCCGGCTTTGC	60	
AluYc3_Gorilla_4F	AAAAAGAGGAAAAAGAAAATGTTCAA	60	chr3:139407803-139408112
AluYc3_Gorilla_4R	ATGAGGCTCAGAGAGGTTGG	60	
WG-Gor3chr5:41,399,009-F	CATTTTATGGATAAGCAGACTGAGG	58	chr5:41399009-41399304
WG-Gor3chr5:41,399,009-R	ACCATGCCCAAGACAAGGTC	58	
WG-Gor3chr9:16,782,260-F	TCTTAACATATCTCAATAAACTGCCTA	57	chr9:16782260-16782557
WG-Gor3chr9:16,782,260-R	AGAACAGCATCCACTCAAGGA	57	
WG-Gor3chr3:19,310,965-F	CTCCCACTAGAGCAAGCAGA	60	chr3:19310965-19311262
WG-Gor3chr3:19,310,965-R	TGTTTCAGGAAGTCAGTATGAGC	60	
WG-Gor3chr15:23,288,171-F	GCTCTTTGGTTCCTTTTGTACTG	56	chr15:23288171-23288459
WG-Gor3chr15:23,288,171-R	AAGGCTGTCATACAGTCTGCAA	56	
WG-Gor3chr3:175,415,623-F	AATTATGCCATATTGAAAGGTGTT	60	chr3:175415623-175416516
WG-Gor3chr3:175,415,623-R	TGTTCAGAAGTCAGAAGAAGGTCA	60	
WG-Gor3chr1_56387832_56388132-F	TAGGCCCAGCAACTCAGG	60	chr1:56387832-56388132
WG-Gor3chr1_56387832_56388132-R	GAAGTGGACAGGATGGGTGT	60	
WG-Gor3chrX:76,869,626-F	AAACCATTTATTATCTCATCTTGCTG	60	chrX:76869626-76870519
WG-Gor3chrX:76,869,626-R	TCATTCTGTCCAATGCCTGT	60	
WG-Gor421 chr9-F	CCAAATAACTCATAGTTCTCCCAAA	60	chr9:39053743-39054639
WG-Gor421 chr9-R	GACTATTTGATGCAGGCACA	60	

	C		
WG-Gor423 chr11-F	GGTGTTCATAGCCACTGGTTG	60	chr11:95184193-95184484
WG-Gor423 chr11-R	GAATCATCCCTGGACATTGG	60	

Table S.5: A complete listing of 1,075 verified gorilla-specific *Alu*Y insertions.

Chr	Start	End	Subfamily
chr5	162669585	162669889	*Alu*Y_*Gorilla*
chr5	13684111	13684428	*Alu*Y_*Gorilla*
chr6	142288088	142288387	*Alu*Y_*Gorilla*
chr4	162950156	162950465	*Alu*Y_*Gorilla*
chr20	37908586	37908893	*Alu*Y_*Gorilla*
chr6	41235725	41236030	*Alu*Y_*Gorilla*
chr10	74362374	74362694	*Alu*Y_*Gorilla*
chr10	1902609	1902927	*Alu*Y_*Gorilla*
chr10	120948860	120949173	*Alu*Y_*Gorilla*
chr9	86937603	86937904	*Alu*Y_*Gorilla*
chr10	101741871	101742155	*Alu*Y_*Gorilla*
chr10	31662560	31662895	*Alu*Y_*Gorilla*
chr12	86997313	86997624	*Alu*Y_*Gorilla*
chr20	29831212	29831511	*Alu*Y_*Gorilla*
chr15	12586908	12587214	*Alu*Y_*Gorilla*
chr3	38918080	38918385	*Alu*Y_*Gorilla*
chr16	67485229	67485528	*Alu*Y_*Gorilla*
chr3	19310965	19311262	*Alu*Y_*Gorilla*
chr20	2856757	2857068	*Alu*Y_*Gorilla*
chr22	11044041	11044345	*Alu*Y_*Gorilla*
chr2A	29048246	29048555	*Alu*Y_*Gorilla*
chr2B	118114799	118115112	*Alu*Y_*Gorilla*
chr4	187734963	187735301	*Alu*Y_*Gorilla*
chr11	20025655	20025990	*Alu*Y_*Gorilla*
chr19	46352863	46353211	*Alu*Y_*Gorilla*
chr12	30477573	30477862	*Alu*Y_*Gorilla*
chr13	11335634	11335926	*Alu*Y_*Gorilla*
chr10	10441170	10441462	*Alu*Y_*Gorilla*
chr7	64946310	64946617	*Alu*Y_*Gorilla*
chr7	62428470	62428753	*Alu*Y_*Gorilla*
chr5	69864913	69865222	*Alu*Y_*Gorilla*
chr5	69865519	69865828	*Alu*Y_*Gorilla*

chr7	95912282	95912598	*Alu*Y_*Gorilla*
chrX	5135584	5135921	*Alu*Y_*Gorilla*
chr9	8841606	8841905	*Alu*Y_*Gorilla*
chrX	66995255	66995554	*Alu*Y_*Gorilla*
chr10	578883	579182	*Alu*Y_*Gorilla*
chr8	44120287	44120582	*Alu*Y_*Gorilla*
chr12	51433645	51433947	*Alu*Y_*Gorilla*
chr22	9437958	9438250	*Alu*Y_*Gorilla*
chr15	23288171	23288459	*Alu*Y_*Gorilla*
chr4	107234739	107235049	*Alu*Y_*Gorilla*
chr3	175415921	175416218	*Alu*Y_*Gorilla*
chr2B	115041713	115042004	*Alu*Y_*Gorilla*
chr2B	31235551	31235847	*Alu*Y_*Gorilla*
chr8	38015650	38015945	*Alu*Y_*Gorilla*
chrX	76869938	76870235	*Alu*Y_*Gorilla*
chr6	64893235	64893537	*Alu*Y_*Gorilla*
chr16	18270868	18271162	*Alu*Y_*Gorilla*
chr2B	36000982	36001272	*Alu*Y_*Gorilla*
chr6	161333696	161333995	*Alu*Y_*Gorilla*
chr12	85522019	85522317	*Alu*Y_*Gorilla*
chr9	39054042	39054340	*Alu*Y_*Gorilla*
chr6	12544894	12545185	*Alu*Y_*Gorilla*
chr10	118640620	118640942	*Alu*Y_*Gorilla*
chr9	30423357	30423666	*Alu*Y_*Gorilla*
chr6	131770934	131771242	*Alu*Y_*Gorilla*
chrX	70533820	70534128	*Alu*Y_*Gorilla*
chr7	122655755	122656062	*Alu*Y_*Gorilla*
chr7	73620519	73620821	*Alu*Y_*Gorilla*
chr11	104569843	104570150	*Alu*Y_*Gorilla*
chr17	91179111	91179418	*Alu*Y_*Gorilla*
chr6	115513487	115513790	*Alu*Y_*Gorilla*
chr8	82767352	82767655	*Alu*Y_*Gorilla*
chrX	38642951	38643254	*Alu*Y_*Gorilla*
chr10	28984298	28984601	*AluY_Gorilla*
chr10	99041313	99041616	*Alu*Y_*Gorilla*
chr12	66501882	66502185	*Alu*Y_*Gorilla*
chr15	28440037	28440341	*Alu*Y_*Gorilla*
chr13	90506534	90506837	*Alu*Y_*Gorilla*
chr3	54134621	54134902	*Alu*Y_*Gorilla*
chr1	56387832	56388132	*Alu*Y_*Gorilla*
chr12	106523736	106524045	*Alu*Y_*Gorilla*

chr19	34866989	34867294	*Alu*Y_*Gorilla*
chr1	204853308	204853609	*Alu*Y_*Gorilla*
chr6	60911397	60911701	*Alu*Y_*Gorilla*
chr10	88746476	88746784	*Alu*Y_*Gorilla*
chr16	18374400	18374707	*Alu*Y_*Gorilla*
chr15	61251539	61251846	*Alu*Y_*Gorilla*
chr1	132758646	132758947	*Alu*Y_*Gorilla*
chr7	77324451	77324761	*Alu*Y_*Gorilla*
chr13	69081094	69081401	*Alu*Y_*Gorilla*
chr5	43934028	43934338	*Alu*Y_*Gorilla*
chr11	88231370	88231680	*Alu*Y_*Gorilla*
chr3	5496899	5497199	*Alu*Y_*Gorilla*
chr17	880013	880323	*Alu*Y_*Gorilla*
chr3	95264819	95265125	*Alu*Y_*Gorilla*
chr12	90894351	90894651	*Alu*Y_*Gorilla*
chr6	77494827	77495135	*Alu*Y_*Gorilla*
chr3	20589548	20589852	*Alu*Y_*Gorilla*
chr15	77545627	77545929	*Alu*Y_*Gorilla*
chr18	14644863	14645173	*Alu*Y_*Gorilla*
chr14	44885847	44886174	*Alu*Y_*Gorilla*
chr11	22724816	22725124	*Alu*Y_*Gorilla*
chr5	64532656	64532967	*Alu*Y_*Gorilla*
chr13	34504629	34504938	*Alu*Y_*Gorilla*
chr7	31163413	31163715	*Alu*Y_*Gorilla*
chr7	76893271	76893575	*Alu*Y_*Gorilla*
chr20	60038959	60039268	*Alu*Y_*Gorilla*
chr6	166730008	166730313	*Alu*Y_*Gorilla*
chr3	12313627	12313937	*Alu*Y_*Gorilla*
chr4	122963632	122963942	*Alu*Y_*Gorilla*
chr10	57062272	57062573	*Alu*Y_*Gorilla*
chr12	89650093	89650410	*Alu*Y_*Gorilla*
chr10	4399315	4399625	*Alu*Y_*Gorilla*
chr9	56204543	56204846	*Alu*Y_*Gorilla*
chr14	33569252	33569554	*Alu*Y_*Gorilla*
chr20	31776579	31776886	*Alu*Y_*Gorilla*
chr12	5281601	5281903	*Alu*Y_*Gorilla*
chr4	178573661	178573968	*Alu*Y_*Gorilla*
chr3	11931683	11931996	*Alu*Y_*Gorilla*
chr9	51376948	51377242	*Alu*Y_*Gorilla*
chr5	99932363	99932663	*Alu*Y_*Gorilla*
chr11	88624056	88624363	*Alu*Y_*Gorilla*

chr8	112413990	112414292	*Alu*Y_*Gorilla*
chr11	104840865	104841162	*Alu*Y_*Gorilla*
chr11	38356201	38356496	*Alu*Y_*Gorilla*
chr11	98496017	98496320	*Alu*Y_*Gorilla*
chr5	84930313	84930622	*Alu*Y_*Gorilla*
chr8	38237333	38237643	*Alu*Y_*Gorilla*
chr1	19983001	19983330	*Alu*Y_*Gorilla*
chr9	50448814	50449120	*Alu*Y_*Gorilla*
chr6	152443003	152443309	*Alu*Y_*Gorilla*
chr3	194177227	194177534	*Alu*Y_*Gorilla*
chr1	188667451	188667758	*Alu*Y_*Gorilla*
chr7	35433537	35433845	*Alu*Y_*Gorilla*
chr6	107166128	107166435	*Alu*Y_*Gorilla*
chr4	3751940	3752247	*Alu*Y_*Gorilla*
chr16	67650170	67650490	*Alu*Y_*Gorilla*
chr18	41809973	41810272	*Alu*Y_*Gorilla*
chrX	71966419	71966745	*Alu*Y_*Gorilla*
chr7	23859012	23859313	*Alu*Y_*Gorilla*
chr7	150561766	150562080	*Alu*Y_*Gorilla*
chr7	110662045	110662357	*Alu*Y_*Gorilla*
chr7	10504797	10505105	*Alu*Y_*Gorilla*
chr22	5729765	5730075	*Alu*Y_*Gorilla*
chr6	74827586	74827887	*Alu*Y_*Gorilla*
chr6	148291907	148292199	*Alu*Y_*Gorilla*
chr4	121586626	121586893	*Alu*Y_*Gorilla*
chr1	40843796	40844111	*Alu*Y_*Gorilla*
chr5	52876832	52877146	*Alu*Y_*Gorilla*
chr18	14206927	14207238	*Alu*Y_*Gorilla*
chr1	180818620	180818906	*Alu*Y_*Gorilla*
chr11	71592992	71593269	*Alu*Y_*Gorilla*
chr19	31614980	31615282	*Alu*Y_*Gorilla*
chr10	105747949	105748250	*Alu*Y_*Gorilla*
chr2B	83465017	83465317	*Alu*Y_*Gorilla*
chr9	102727418	102727716	*Alu*Y_*Gorilla*
chr1	62675213	62675510	*Alu*Y_*Gorilla*
chr17	50515540	50515841	*Alu*Y_*Gorilla*
chr3	85874922	85875233	*Alu*Y_*Gorilla*
chr16	75031384	75031688	*Alu*Y_*Gorilla*
chr20	19133735	19134026	*Alu*Y_*Gorilla*
chr18	59872925	59873223	*Alu*Y_*Gorilla*
chr14	2227448	2227745	*Alu*Y_*Gorilla*

chr14	82186050	82186347	*Alu*Y_*Gorilla*
chr12	55513650	55513961	*Alu*Y_*Gorilla*
chr20	25510566	25510832	*Alu*Y_*Gorilla*
chr16	30349058	30349369	*Alu*Y_*Gorilla*
chr17	16237430	16237741	*Alu*Y_*Gorilla*
chr9	70309089	70309392	*Alu*Y_*Gorilla*
chr1	192919379	192919652	*Alu*Y_*Gorilla*
chr19	21235238	21235573	*Alu*Y_*Gorilla*
chr20	5201905	5202214	*Alu*Y_*Gorilla*
chr10	11637060	11637348	*Alu*Y_*Gorilla*
chr4	13881361	13881637	*Alu*Y_*Gorilla*
chr3	107929811	107930114	*Alu*Y_*Gorilla*
chrX	78613189	78613503	*Alu*Y_*Gorilla*
chr8	102910367	102910680	*Alu*Y_*Gorilla*
chr11	28109297	28109610	*Alu*Y_*Gorilla*
chr15	52990616	52990878	*Alu*Y_*Gorilla*
chr17	77908632	77908944	*Alu*Y_*Gorilla*
chr2A	51836876	51837188	*Alu*Y_*Gorilla*
chr12	25824329	25824595	*Alu*Y_*Gorilla*
chr1	31721842	31722153	*Alu*Y_*Gorilla*
chr1	73097089	73097400	*Alu*Y_*Gorilla*
chr13	7889153	7889439	*Alu*Y_*Gorilla*
chr12	115437153	115437464	*Alu*Y_*Gorilla*
chr12	80009772	80010058	*Alu*Y_*Gorilla*
chr11	119568741	119568994	*Alu*Y_*Gorilla*
chr2A	14959134	14959445	*Alu*Y_*Gorilla*
chr2B	33475785	33476096	*Alu*Y_*Gorilla*
chrX	74538593	74538859	*Alu*Y_*Gorilla*
chr4	140302595	140302905	*Alu*Y_*Gorilla*
chr5	136764594	136764851	*Alu*Y_*Gorilla*
chr5	137724371	137724681	*Alu*Y_*Gorilla*
chrX	24110077	24110387	*Alu*Y_*Gorilla*
chrX	41960925	41961235	*Alu*Y_*Gorilla*
chr5	70694823	70695132	*Alu*Y_*Gorilla*
chr5	70722799	70723108	*Alu*Y_*Gorilla*
chr5	82781125	82781434	*Alu*Y_*Gorilla*
chr5	154040963	154041272	*Alu*Y_*Gorilla*
chr7	54368489	54368798	*Alu*Y_*Gorilla*
chr12	69565417	69565726	*Alu*Y_*Gorilla*
chr2B	74299235	74299544	*Alu*Y_*Gorilla*
chr1	30341681	30341989	*Alu*Y_*Gorilla*

chr1	143126583	143126891	*Alu*Y_*Gorilla*
chr3	18432457	18432765	*Alu*Y_*Gorilla*
chr3	175297780	175298088	*Alu*Y_*Gorilla*
chr4	193869099	193869407	*Alu*Y_*Gorilla*
chr5	121628231	121628539	*Alu*Y_*Gorilla*
chr11	13600887	13601195	*Alu*Y_*Gorilla*
chr11	119648850	119649158	*Alu*Y_*Gorilla*
chr16	55084860	55085168	*Alu*Y_*Gorilla*
chr18	74463427	74463735	*Alu*Y_*Gorilla*
chr3	85917294	85917601	*Alu*Y_*Gorilla*
chrX	37213942	37214249	*Alu*Y_*Gorilla*
chr12	86268924	86269231	*Alu*Y_*Gorilla*
chr13	74816771	74817078	*Alu*Y_*Gorilla*
chr16	65667032	65667339	*Alu*Y_*Gorilla*
chr17	49034712	49035019	*Alu*Y_*Gorilla*
chr22	28025785	28026092	*Alu*Y_*Gorilla*
chr1	137976193	137976499	*Alu*Y_*Gorilla*
chr3	197584541	197584847	*Alu*Y_*Gorilla*
chr3	197592615	197592921	*Alu*Y_*Gorilla*
chr5	54422205	54422511	*Alu*Y_*Gorilla*
chr5	65376872	65377178	*Alu*Y_*Gorilla*
chr5	102541622	102541928	*Alu*Y_*Gorilla*
chr11	88201355	88201661	*Alu*Y_*Gorilla*
chr16	30907953	30908259	*Alu*Y_*Gorilla*
chr2A	81593657	81593963	*Alu*Y_*Gorilla*
chr2B	58525317	58525623	*Alu*Y_*Gorilla*
chr4	143702742	143703047	*Alu*Y_*Gorilla*
chr8	9198041	9198346	*Alu*Y_*Gorilla*
chr9	21079889	21080194	*Alu*Y_*Gorilla*
chr11	47740874	47741179	*Alu*Y_*Gorilla*
chr2B	107806034	107806339	*Alu*Y_*Gorilla*
chr1	217260108	217260412	*Alu*Y_*Gorilla*
chr13	32805307	32805611	*Alu*Y_*Gorilla*
chr14	1617138	1617442	*Alu*Y_*Gorilla*
chr19	25916843	25917147	*Alu*Y_*Gorilla*
chr20	7410922	7411226	*Alu*Y_*Gorilla*
chr2A	85157269	85157573	*Alu*Y_*Gorilla*
chr2B	100641909	100642213	*Alu*Y_*Gorilla*
chr2B	42856607	42856910	*Alu*Y_*Gorilla*
chr6	103416835	103417137	*Alu*Y_*Gorilla*
chr6	127940734	127941036	*Alu*Y_*Gorilla*

chrX	16745783	16746085	*Alu*Y_*Gorilla*
chr10	41468507	41468809	*Alu*Y_*Gorilla*
chr2B	89710637	89710939	*Alu*Y_*Gorilla*
chr3	183005831	183006132	*Alu*Y_*Gorilla*
chrX	322200	322500	*Alu*Y_*Gorilla*
chr19	46240994	46241294	*Alu*Y_*Gorilla*
chr22	27436596	27436896	*Alu*Y_*Gorilla*
chr1	105738744	105739043	*Alu*Y_*Gorilla*
chr4	127615535	127615834	*Alu*Y_*Gorilla*
chr5	16028889	16029188	*Alu*Y_*Gorilla*
chr5	105863249	105863548	*Alu*Y_*Gorilla*
chr5	143292571	143292870	*Alu*Y_*Gorilla*
chr6	53108147	53108446	*Alu*Y_*Gorilla*
chr12	25767770	25768069	*Alu*Y_*Gorilla*
chr2B	102212407	102212706	*Alu*Y_*Gorilla*
chr6	133992444	133992742	*Alu*Y_*Gorilla*
chrX	23858118	23858416	*Alu*Y_*Gorilla*
chrX	101713280	101713578	*Alu*Y_*Gorilla*
chr10	104980530	104980828	*Alu*Y_*Gorilla*
chr12	39227206	39227504	*Alu*Y_*Gorilla*
chr2B	89316444	89316742	*Alu*Y_*Gorilla*
chr3	12746762	12747059	*Alu*Y_*Gorilla*
chr3	112811554	112811851	*Alu*Y_*Gorilla*
chr9	108479262	108479559	*Alu*Y_*Gorilla*
chr12	83915563	83915860	*Alu*Y_*Gorilla*
chr13	31391780	31392077	*Alu*Y_*Gorilla*
chr13	77739911	77740208	*Alu*Y_*Gorilla*
chr17	62745053	62745350	*Alu*Y_*Gorilla*
chr19	1889748	1890045	*Alu*Y_*Gorilla*
chr19	31651793	31652090	*Alu*Y_*Gorilla*
chr2B	14410383	14410680	*Alu*Y_*Gorilla*
chr4	31910188	31910484	*Alu*Y_*Gorilla*
chr4	148652288	148652584	*Alu*Y_*Gorilla*
chr4	163629526	163629822	*Alu*Y_*Gorilla*
chr5	130714857	130715153	*Alu*Y_*Gorilla*
chr6	24219258	24219554	*Alu*Y_*Gorilla*
chr6	78466257	78466553	*Alu*Y_*Gorilla*
chr6	152592779	152593075	*Alu*Y_*Gorilla*
chr7	25378617	25378913	*Alu*Y_*Gorilla*
chr7	38311443	38311739	*Alu*Y_*Gorilla*
chr9	252503	252799	*Alu*Y_*Gorilla*

chr11	57479287	57479583	*Alu*Y_*Gorilla*
chr12	99570678	99570974	*Alu*Y_*Gorilla*
chr16	5588580	5588876	*Alu*Y_*Gorilla*
chr20	58604645	58604941	*Alu*Y_*Gorilla*
chr3	183833146	183833441	*Alu*Y_*Gorilla*
chr4	197033908	197034203	*Alu*Y_*Gorilla*
chr5	129896925	129897220	*Alu*Y_*Gorilla*
chr7	89925757	89926052	*Alu*Y_*Gorilla*
chr8	34418611	34418906	*Alu*Y_*Gorilla*
chr14	31055358	31055653	*Alu*Y_*Gorilla*
chr15	53484567	53484862	*Alu*Y_*Gorilla*
chr20	13218802	13219097	*Alu*Y_*Gorilla*
chr22	23619879	23620174	*Alu*Y_*Gorilla*
chr2B	76454443	76454738	*Alu*Y_*Gorilla*
chr3	103268923	103269217	*Alu*Y_*Gorilla*
chr4	14970633	14970927	*Alu*Y_*Gorilla*
chr4	162757326	162757620	*Alu*Y_*Gorilla*
chr6	29167862	29168156	*Alu*Y_*Gorilla*
chr7	19054556	19054850	*Alu*Y_*Gorilla*
chr7	110444126	110444420	*Alu*Y_*Gorilla*
chrX	14824879	14825173	*Alu*Y_*Gorilla*
chr10	736761	737042	*Alu*Y_*Gorilla*
chr2A	10695039	10695319	*Alu*Y_*Gorilla*
chr2B	122106667	122106947	*Alu*Y_*Gorilla*
chrX	95936646	95936950	*Alu*Y_*Gorilla*
chr1	188816443	188816734	*Alu*Y_*Gorilla*
chr1	205894136	205894439	*Alu*Y_*Gorilla*
chr1	120400240	120400542	*Alu*Y_*Gorilla*
chr1	196145179	196145481	*Alu*Y_*Gorilla*
chr1	103402000	103402298	*Alu*Y_*Gorilla*
chr3	38138472	38138777	*Alu*Y_*Gorilla*
chr22	4209498	4209805	*Alu*Y_*Gorilla*
chr3	63382091	63382397	*Alu*Y_*Gorilla*
chr6	6338056	6338369	*Alu*Y_*Gorilla*
chr3	11919992	11920280	*Alu*Y_*Gorilla*
chr15	19924516	19924818	*Alu*Y_*Gorilla*
chr1	15824206	15824499	*Alu*Y_*Gorilla*
chr2A	84264995	84265307	*Alu*Y_*Gorilla*
chr18	11127824	11128136	*Alu*Y_*Gorilla*
chr1	56388141	56388452	*Alu*Y_*Gorilla*
chr7	97815480	97815801	*Alu*Y_*Gorilla*

chr16	57065344	57065667	*Alu*Y_*Gorilla*
chr1	104954608	104954909	*Alu*Y_*Gorilla*
chr3	198537044	198537365	*Alu*Y_*Gorilla*
chr4	95187983	95188291	*Alu*Y_*Gorilla*
chr5	19319469	19319794	*Alu*Y_*Gorilla*
chr9	107092595	107092899	*Alu*Y_*Gorilla*
chr3	180617812	180618101	*Alu*Y_*Gorilla*
chr3	189350349	189350659	*Alu*Y_*Gorilla*
chr4	155080244	155080555	*Alu*Y_*Gorilla*
chr7	142080970	142081281	*Alu*Y_*Gorilla*
chr10	63130736	63131047	*Alu*Y_*Gorilla*
chr2B	128481748	128482059	*Alu*Y_*Gorilla*
chr1	128234045	128234355	*Alu*Y_*Gorilla*
chr1	211114837	211115147	*Alu*Y_*Gorilla*
chr8	85511333	85511643	*Alu*Y_*Gorilla*
chr10	133042818	133043128	*Alu*Y_*Gorilla*
chr12	23846698	23847008	*Alu*Y_*Gorilla*
chr12	42369893	42370203	*Alu*Y_*Gorilla*
chr12	96688959	96689269	*Alu*Y_*Gorilla*
chr14	81033976	81034286	*Alu*Y_*Gorilla*
chr15	34469538	34469848	*Alu*Y_*Gorilla*
chr17	5294258	5294568	*Alu*Y_*Gorilla*
chr2B	31817846	31818156	*Alu*Y_*Gorilla*
chr2B	96181940	96182250	*Alu*Y_*Gorilla*
chr2B	96210970	96211280	*Alu*Y_*Gorilla*
chr1	44491239	44491548	*Alu*Y_*Gorilla*
chr1	61695783	61696087	*Alu*Y_*Gorilla*
chrX	231908	232225	*Alu*Y_*Gorilla*
chrX	27538602	27538923	*Alu*Y_*Gorilla*
chr4	2550359	2550652	*Alu*Y_*Gorilla*
chr1	19090305	19090597	*Alu*Y_*Gorilla*
chr16	62567709	62568034	*Alu*Y_*Gorilla*
chr19	13439442	13439746	*Alu*Y_*Gorilla*
chr8	78546722	78547028	*Alu*Y_*Gorilla*
chr19	55079471	55079791	*Alu*Y_*Gorilla*
chr2B	84565521	84565819	*Alu*Y_*Gorilla*
chr7	138396440	138396737	*Alu*Y_*Gorilla*
chr2B	16622804	16623089	*Alu*Y_*Gorilla*
chr19	9871821	9872104	*Alu*Y_*Gorilla*
chr16	75033712	75033993	*Alu*Y_*Gorilla*
chr19	44070760	44071041	*Alu*Y_*Gorilla*

chr14	35677666	35677944	*Alu*Y_*Gorilla*
chr19	12216602	12216878	*Alu*Y_*Gorilla*
chr3	10140587	10140861	*Alu*Y_*Gorilla*
chr2A	83469194	83469463	*Alu*Y_*Gorilla*
chr11	72067024	72067401	*Alu*Y_*Gorilla*
chr19	50212989	50213364	*Alu*Y_*Gorilla*
chr9	47058818	47059180	*Alu*Y_*Gorilla*
chr5	41589935	41590296	*Alu*Y_*Gorilla*
chr15	53446024	53446382	*Alu*Y_*Gorilla*
chr12	125712063	125712419	*Alu*Y_*Gorilla*
chr4	163150414	163150763	*Alu*Y_*Gorilla*
chr7	57547178	57547527	*Alu*Y_*Gorilla*
chr18	38835244	38835592	*Alu*Y_*Gorilla*
chr15	53452784	53453142	*Alu*Y_*Gorilla*
chrX	54332056	54332402	*Alu*Y_*Gorilla*
chr19	20761256	20761600	*Alu*Y_*Gorilla*
chr6	34363492	34363835	*Alu*Y_*Gorilla*
chr1	193435098	193435440	*Alu*Y_*Gorilla*
chr8	137326777	137327119	*Alu*Y_*Gorilla*
chr6	1469903	1470244	*Alu*Y_*Gorilla*
chr7	135937893	135938234	*Alu*Y_*Gorilla*
chr5	10104989	10105328	*Alu*Y_*Gorilla*
chr9	47057410	47057749	*Alu*Y_*Gorilla*
chr21	4315560	4315899	*Alu*Y_*Gorilla*
chrX	54326143	54326481	*Alu*Y_*Gorilla*
chr20	54439758	54440095	*Alu*Y_*Gorilla*
chr3	103951698	103952033	*Alu*Y_*Gorilla*
chr5	17662067	17662397	*Alu*Y_*Gorilla*
chr10	16812088	16812418	*Alu*Y_*Gorilla*
chr17	238830	239160	*Alu*Y_*Gorilla*
chr20	31069083	31069413	*Alu*Y_*Gorilla*
chr2B	37905589	37905919	*Alu*Y_*Gorilla*
chr6	159766550	159766879	*Alu*Y_*Gorilla*
chr4	12124318	12124645	*Alu*Y_*Gorilla*
chr5	1882375	1882702	*Alu*Y_*Gorilla*
chr15	39476172	39476499	*Alu*Y_*Gorilla*
chr7	6047595	6047921	*Alu*Y_*Gorilla*
chr7	6066128	6066454	*Alu*Y_*Gorilla*
chrX	75067972	75068298	*Alu*Y_*Gorilla*
chr19	24308849	24309175	*Alu*Y_*Gorilla*
chr19	49831969	49832294	*Alu*Y_*Gorilla*

chr22	22935273	22935598	*Alu*Y_*Gorilla*
chr1	42121095	42121419	*Alu*Y_*Gorilla*
chr5	1441861	1442185	*Alu*Y_*Gorilla*
chr5	2060350	2060674	*Alu*Y_*Gorilla*
chr6	116780861	116781185	*Alu*Y_*Gorilla*
chr10	20290180	20290504	*Alu*Y_*Gorilla*
chr2A	76527131	76527455	*Alu*Y_*Gorilla*
chr17	20321255	20321578	*Alu*Y_*Gorilla*
chr3	180377602	180377924	*Alu*Y_*Gorilla*
chr22	9428832	9429154	*Alu*Y_*Gorilla*
chr4	24350129	24350450	*Alu*Y_*Gorilla*
chr10	95299645	95299966	*Alu*Y_*Gorilla*
chr2A	106970983	106971304	*Alu*Y_*Gorilla*
chr2A	107771675	107771996	*Alu*Y_*Gorilla*
chr2A	108191661	108191982	*Alu*Y_*Gorilla*
chr2B	71454138	71454459	*Alu*Y_*Gorilla*
chr1	20587734	20588054	*Alu*Y_*Gorilla*
chr4	109076302	109076622	*Alu*Y_*Gorilla*
chr5	1353243	1353563	*Alu*Y_*Gorilla*
chr5	61266436	61266756	*Alu*Y_*Gorilla*
chr22	8563228	8563548	*Alu*Y_*Gorilla*
chr1	114209713	114210032	*Alu*Y_*Gorilla*
chr1	134328369	134328688	*Alu*Y_*Gorilla*
chr12	1565990	1566309	*Alu*Y_*Gorilla*
chr15	73170541	73170860	*Alu*Y_*Gorilla*
chr1	192645326	192645644	*Alu*Y_*Gorilla*
chr3	170231658	170231976	*Alu*Y_*Gorilla*
chr5	11896296	11896614	*Alu*Y_*Gorilla*
chr12	109857150	109857468	*Alu*Y_*Gorilla*
chr13	45018340	45018658	*Alu*Y_*Gorilla*
chr18	11318064	11318382	*Alu*Y_*Gorilla*
chr2B	23678959	23679277	*Alu*Y_*Gorilla*
chr3	53378847	53379164	*Alu*Y_*Gorilla*
chr10	125346470	125346787	*Alu*Y_*Gorilla*
chr14	40715842	40716159	*Alu*Y_*Gorilla*
chr16	26456679	26456996	*Alu*Y_*Gorilla*
chr16	58241738	58242055	*Alu*Y_*Gorilla*
chr18	7391685	7392002	*Alu*Y_*Gorilla*
chr1	124400636	124400952	*AluY_Gorilla*
chr1	132575699	132576015	*Alu*Y_*Gorilla*
chr9	76784972	76785288	*Alu*Y_*Gorilla*

chr1	168826124	168826439	*Alu*Y_*Gorilla*
chr13	41297267	41297582	*Alu*Y_*Gorilla*
chr2A	26268181	26268496	*Alu*Y_*Gorilla*
chr4	21610252	21610566	*Alu*Y_*Gorilla*
chr4	45979303	45979617	*Alu*Y_*Gorilla*
chr19	7036992	7037305	*Alu*Y_*Gorilla*
chrX	49682325	49682637	*Alu*Y_*Gorilla*
chr4	23915972	23916283	*Alu*Y_*Gorilla*
chr4	166505230	166505541	*Alu*Y_*Gorilla*
chr8	33060924	33061235	*Alu*Y_*Gorilla*
chr16	43348674	43348985	*Alu*Y_*Gorilla*
chr4	74475610	74475920	*Alu*Y_*Gorilla*
chr4	79403608	79403918	*Alu*Y_*Gorilla*
chr7	100212460	100212770	*Alu*Y_*Gorilla*
chr18	51370215	51370525	*Alu*Y_*Gorilla*
chr21	4508872	4509182	*Alu*Y_*Gorilla*
chr4	94659990	94660299	*Alu*Y_*Gorilla*
chr11	64073859	64074168	*Alu*Y_*Gorilla*
chr13	72562148	72562457	*Alu*Y_*Gorilla*
chr17	16036171	16036480	*Alu*Y_*Gorilla*
chr14	27615259	27615567	*Alu*Y_*Gorilla*
chr16	62058405	62058713	*Alu*Y_*Gorilla*
chr2A	32976655	32976963	*Alu*Y_*Gorilla*
chr2A	65438365	65438673	*Alu*Y_*Gorilla*
chr2A	110885133	110885441	*Alu*Y_*Gorilla*
chr2B	59196826	59197134	*Alu*Y_*Gorilla*
chr2B	72308267	72308575	*Alu*Y_*Gorilla*
chr2B	95510822	95511130	*Alu*Y_*Gorilla*
chr2B	117598603	117598911	*Alu*Y_*Gorilla*
chr1	40361642	40361949	*Alu*Y_*Gorilla*
chr1	95971979	95972286	*Alu*Y_*Gorilla*
chr12	45306858	45307164	*Alu*Y_*Gorilla*
chr12	45328413	45328719	*Alu*Y_*Gorilla*
chr15	59784282	59784588	*Alu*Y_*Gorilla*
chr16	29117620	29117926	*Alu*Y_*Gorilla*
chr3	60982298	60982603	*Alu*Y_*Gorilla*
chr3	74506424	74506728	*Alu*Y_*Gorilla*
chr6	22535247	22535551	*Alu*Y_*Gorilla*
chr7	97966924	97967228	*Alu*Y_*Gorilla*
chr7	129070819	129071123	*Alu*Y_*Gorilla*
chr9	45547059	45547363	*Alu*Y_*Gorilla*

chrX	101681883	101682187	*Alu*Y_*Gorilla*
chr1	18043756	18044058	*Alu*Y_*Gorilla*
chr19	11257015	11257317	*Alu*Y_*Gorilla*
chr19	26320124	26320426	*Alu*Y_*Gorilla*
chr19	46298010	46298312	*Alu*Y_*Gorilla*
chr22	2264201	2264503	*Alu*Y_*Gorilla*
chr22	15671329	15671631	*Alu*Y_*Gorilla*
chr22	21933870	21934172	*Alu*Y_*Gorilla*
chr22	24484170	24484472	*Alu*Y_*Gorilla*
chr22	28444304	28444606	*Alu*Y_*Gorilla*
chr2A	43531917	43532219	*Alu*Y_*Gorilla*
chr2A	54979690	54979992	*Alu*Y_*Gorilla*
chr2A	94292105	94292407	*Alu*Y_*Gorilla*
chr2B	17162195	17162497	*Alu*Y_*Gorilla*
chr2B	34929591	34929893	*Alu*Y_*Gorilla*
chr1	29968791	29969092	*Alu*Y_*Gorilla*
chr1	47733922	47734223	*Alu*Y_*Gorilla*
chr1	97694308	97694609	*Alu*Y_*Gorilla*
chr1	115721389	115721690	*Alu*Y_*Gorilla*
chr3	198119954	198120255	*Alu*Y_*Gorilla*
chr4	42459352	42459653	*Alu*Y_*Gorilla*
chr4	46748437	46748738	*Alu*Y_*Gorilla*
chr4	110852199	110852500	*Alu*Y_*Gorilla*
chr4	180433855	180434156	*Alu*Y_*Gorilla*
chr5	15591033	15591334	*Alu*Y_*Gorilla*
chr5	34080001	34080302	*Alu*Y_*Gorilla*
chr5	40448542	40448843	*Alu*Y_*Gorilla*
chr6	19965293	19965594	*Alu*Y_*Gorilla*
chr6	33465183	33465484	*Alu*Y_*Gorilla*
chr6	152218406	152218707	*Alu*Y_*Gorilla*
chr6	157899567	157899868	*Alu*Y_*Gorilla*
chr7	45033437	45033738	*Alu*Y_*Gorilla*
chr7	95604067	95604368	*Alu*Y_*Gorilla*
chr8	12169597	12169898	*Alu*Y_*Gorilla*
chr9	33690042	33690343	*Alu*Y_*Gorilla*
chr9	114283341	114283642	*Alu*Y_*Gorilla*
chrX	28514818	28515119	*Alu*Y_*Gorilla*
chrX	60646410	60646711	*Alu*Y_*Gorilla*
chrX	89933191	89933492	*Alu*Y_*Gorilla*
chrX	96706846	96707147	*Alu*Y_*Gorilla*
chrX	107638355	107638656	*Alu*Y_*Gorilla*

chr10	15570098	15570399	*Alu*Y_*Gorilla*
chr10	38301316	38301617	*Alu*Y_*Gorilla*
chr10	86008221	86008522	*Alu*Y_*Gorilla*
chr11	59650367	59650668	*Alu*Y_*Gorilla*
chr11	77410160	77410461	*Alu*Y_*Gorilla*
chr11	89436929	89437230	*Alu*Y_*Gorilla*
chr12	38215834	38216135	*Alu*Y_*Gorilla*
chr12	39179599	39179900	*Alu*Y_*Gorilla*
chr12	39194073	39194374	*Alu*Y_*Gorilla*
chr12	43645227	43645528	*Alu*Y_*Gorilla*
chr13	23895371	23895672	*Alu*Y_*Gorilla*
chr13	64442178	64442479	*Alu*Y_*Gorilla*
chr14	15628066	15628367	*Alu*Y_*Gorilla*
chr14	15765707	15766008	*Alu*Y_*Gorilla*
chr14	30570896	30571197	*Alu*Y_*Gorilla*
chr12	15027349	15027650	*Alu*Y_*Gorilla*
chr15	985205	985506	*Alu*Y_*Gorilla*
chr15	18177026	18177327	*Alu*Y_*Gorilla*
chr15	20638307	20638608	*Alu*Y_*Gorilla*
chr15	26351036	26351337	*Alu*Y_*Gorilla*
chr15	36470204	36470505	*Alu*Y_*Gorilla*
chr16	60603248	60603549	*Alu*Y_*Gorilla*
chr1	160114662	160114962	*Alu*Y_*Gorilla*
chr4	18646226	18646526	*Alu*Y_*Gorilla*
chr4	122217560	122217860	*Alu*Y_*Gorilla*
chr4	122224885	122225185	*Alu*Y_*Gorilla*
chr6	142921051	142921351	*Alu*Y_*Gorilla*
chr8	24155154	24155454	*Alu*Y_*Gorilla*
chr17	4898678	4898978	*Alu*Y_*Gorilla*
chr16	19833261	19833560	*Alu*Y_*Gorilla*
chr1	113993671	113993969	*Alu*Y_*Gorilla*
chr2B	131184570	131184868	*Alu*Y_*Gorilla*
chr5	17942927	17943224	*Alu*Y_*Gorilla*
chr10	93327019	93327316	*Alu*Y_*Gorilla*
chr15	3528882	3529178	*Alu*Y_*Gorilla*
chr2A	35492199	35492495	*Alu*Y_*Gorilla*
chr19	49501138	49501432	*Alu*Y_*Gorilla*
chr22	2263759	2264053	*Alu*Y_*Gorilla*
chr2A	71502137	71502431	*Alu*Y_*Gorilla*
chr2A	87690095	87690389	*Alu*Y_*Gorilla*
chr3	112651372	112651665	*Alu*Y_*Gorilla*

chr5	37049135	37049428	*Alu*Y_*Gorilla*
chr8	96929763	96930056	*Alu*Y_*Gorilla*
chr8	101595205	101595498	*Alu*Y_*Gorilla*
chrX	129229817	129230110	*Alu*Y_*Gorilla*
chr13	82762454	82762747	*Alu*Y_*Gorilla*
chr15	23436116	23436409	*Alu*Y_*Gorilla*
chr15	56464110	56464403	*Alu*Y_*Gorilla*
chr15	72611352	72611645	*Alu*Y_*Gorilla*
chr2A	54442480	54442773	*Alu*Y_*Gorilla*
chr2A	54841638	54841931	*Alu*Y_*Gorilla*
chr1	1619159	1619451	*Alu*Y_*Gorilla*
chr1	53650812	53651104	*Alu*Y_*Gorilla*
chr3	53057944	53058236	*Alu*Y_*Gorilla*
chr3	58491050	58491342	*Alu*Y_*Gorilla*
chr7	60564180	60564472	*Alu*Y_*Gorilla*
chr8	124711109	124711401	*Alu*Y_*Gorilla*
chrX	49686197	49686489	*Alu*Y_*Gorilla*
chr10	64653926	64654218	*Alu*Y_*Gorilla*
chr11	45707419	45707711	*Alu*Y_*Gorilla*
chr15	22891560	22891852	*Alu*Y_*Gorilla*
chr15	37996567	37996859	*Alu*Y_*Gorilla*
chr19	31779306	31779598	*Alu*Y_*Gorilla*
chr2B	129984753	129985045	*Alu*Y_*Gorilla*
chr3	24193372	24193663	*Alu*Y_*Gorilla*
chr3	42396276	42396567	*Alu*Y_*Gorilla*
chr4	121067549	121067840	*Alu*Y_*Gorilla*
chr7	19696714	19697005	*Alu*Y_*Gorilla*
chr7	25739580	25739871	*Alu*Y_*Gorilla*
chr11	95184193	95184484	*Alu*Y_*Gorilla*
chr11	119463912	119464203	*Alu*Y_*Gorilla*
chr12	67245442	67245733	*Alu*Y_*Gorilla*
chr16	61955857	61956148	*Alu*Y_*Gorilla*
chr19	52369369	52369660	*Alu*Y_*Gorilla*
chr1	55571049	55571339	*Alu*Y_*Gorilla*
chr3	157258873	157259163	*Alu*Y_*Gorilla*
chr4	23818324	23818614	*Alu*Y_*Gorilla*
chr4	114784019	114784309	*Alu*Y_*Gorilla*
chr5	6129684	6129974	*Alu*Y_*Gorilla*
chr7	32917707	32917997	*Alu*Y_*Gorilla*
chr9	89604095	89604385	*Alu*Y_*Gorilla*
chr11	5840405	5840695	*Alu*Y_*Gorilla*

chr12	28068803	28069093	*Alu*Y_*Gorilla*
chr12	45331074	45331364	*Alu*Y_*Gorilla*
chr12	112922510	112922800	*Alu*Y_*Gorilla*
chr16	12525409	12525699	*Alu*Y_*Gorilla*
chr17	76980189	76980479	*Alu*Y_*Gorilla*
chr22	15476814	15477104	*Alu*Y_*Gorilla*
chr5	57927439	57927728	*Alu*Y_*Gorilla*
chr12	58017343	58017632	*Alu*Y_*Gorilla*
chr12	68889099	68889388	*Alu*Y_*Gorilla*
chr12	118946408	118946697	*Alu*Y_*Gorilla*
chr13	34562544	34562833	*Alu*Y_*Gorilla*
chr17	17987026	17987315	*Alu*Y_*Gorilla*
chr18	5040416	5040705	*Alu*Y_*Gorilla*
chr18	63131873	63132162	*Alu*Y_*Gorilla*
chr22	2843408	2843697	*Alu*Y_*Gorilla*
chr2B	125527072	125527361	*Alu*Y_*Gorilla*
chr1	202479	202767	*Alu*Y_*Gorilla*
chr3	54750646	54750934	*Alu*Y_*Gorilla*
chr3	112696471	112696759	*Alu*Y_*Gorilla*
chr5	165927021	165927309	*Alu*Y_*Gorilla*
chr8	108161504	108161792	*Alu*Y_*Gorilla*
chrX	134456788	134457076	*Alu*Y_*Gorilla*
chr11	86444600	86444888	*Alu*Y_*Gorilla*
chr12	44985764	44986052	*Alu*Y_*Gorilla*
chr15	71321360	71321648	*Alu*Y_*Gorilla*
chr17	1343019	1343307	*Alu*Y_*Gorilla*
chr1	147502873	147503160	*Alu*Y_*Gorilla*
chr3	53440874	53441161	*Alu*Y_*Gorilla*
chr3	172167818	172168105	*Alu*Y_*Gorilla*
chr7	106564813	106565100	*Alu*Y_*Gorilla*
chrX	23746690	23746977	*Alu*Y_*Gorilla*
chr12	1005965	1006252	*Alu*Y_*Gorilla*
chr14	12476611	12476898	*Alu*Y_*Gorilla*
chr19	6373956	6374243	*Alu*Y_*Gorilla*
chr1	9715750	9716036	*Alu*Y_*Gorilla*
chr4	128987816	128988102	*Alu*Y_*Gorilla*
chr9	13591328	13591614	*Alu*Y_*Gorilla*
chrX	116029031	116029317	*Alu*Y_*Gorilla*
chrX	137647333	137647619	*Alu*Y_*Gorilla*
chr2B	105641792	105642078	*Alu*Y_*Gorilla*
chr3	199502610	199502895	*Alu*Y_*Gorilla*

chr4	200878424	200878709	*Alu*Y_*Gorilla*
chr5	54690419	54690704	*Alu*Y_*Gorilla*
chr15	38274339	38274624	*Alu*Y_*Gorilla*
chr22	23644382	23644667	*Alu*Y_*Gorilla*
chr2B	13001504	13001789	*Alu*Y_*Gorilla*
chr4	98393576	98393860	*Alu*Y_*Gorilla*
chr5	9320045	9320329	*Alu*Y_*Gorilla*
chr5	95884722	95885006	*Alu*Y_*Gorilla*
chr8	103178874	103179158	*Alu*Y_*Gorilla*
chr9	111233047	111233331	*Alu*Y_*Gorilla*
chr14	84601712	84601996	*Alu*Y_*Gorilla*
chr5	97965852	97966135	*Alu*Y_*Gorilla*
chr7	87392565	87392848	*Alu*Y_*Gorilla*
chr10	126580617	126580900	*Alu*Y_*Gorilla*
chr16	31133693	31133976	*Alu*Y_*Gorilla*
chr17	1340653	1340936	*Alu*Y_*Gorilla*
chr19	4644680	4644963	*Alu*Y_*Gorilla*
chr1	2138713	2139007	*Alu*Y_*Gorilla*
chr1	15736074	15736368	*Alu*Y_*Gorilla*
chr1	125821951	125822233	*Alu*Y_*Gorilla*
chr3	15171578	15171860	*Alu*Y_*Gorilla*
chr5	38702798	38703080	*Alu*Y_*Gorilla*
chr6	53914977	53915259	*Alu*Y_*Gorilla*
chr6	167648176	167648458	*Alu*Y_*Gorilla*
chr7	6048429	6048711	*Alu*Y_*Gorilla*
chr12	31797939	31798221	*Alu*Y_*Gorilla*
chr6	16323009	16323290	*Alu*Y_*Gorilla*
chr1	131791870	131792149	*Alu*Y_*Gorilla*
chr8	104924196	104924475	*Alu*Y_*Gorilla*
chr12	69173049	69173328	*Alu*Y_*Gorilla*
chr1	79325446	79325724	*Alu*Y_*Gorilla*
chr4	29084486	29084764	*Alu*Y_*Gorilla*
chr7	124630306	124630584	*Alu*Y_*Gorilla*
chr10	133173231	133173509	*Alu*Y_*Gorilla*
chr13	66976980	66977258	*Alu*Y_*Gorilla*
chr19	49906769	49907047	*Alu*Y_*Gorilla*
chr5	101806392	101806669	*Alu*Y_*Gorilla*
chr16	65046398	65046675	*Alu*Y_*Gorilla*
chr18	12914290	12914567	*Alu*Y_*Gorilla*
chr19	24597036	24597313	*Alu*Y_*Gorilla*
chr1	215758540	215758816	*Alu*Y_*Gorilla*

chr3	54041391	54041666	*Alu*Y_*Gorilla*
chr5	57970551	57970826	*Alu*Y_*Gorilla*
chr5	150032804	150033079	*Alu*Y_*Gorilla*
chr12	4016243	4016518	*Alu*Y_*Gorilla*
chr13	21708919	21709194	*Alu*Y_*Gorilla*
chr1	60623714	60623988	*Alu*Y_*Gorilla*
chr8	109494874	109495148	*Alu*Y_*Gorilla*
chr11	19528754	19529028	*Alu*Y_*Gorilla*
chr20	25180655	25180929	*Alu*Y_*Gorilla*
chrX	106449297	106449570	*Alu*Y_*Gorilla*
chr5	141026472	141026744	*Alu*Y_*Gorilla*
chr19	12030272	12030544	*Alu*Y_*Gorilla*
chr1	40904605	40904876	*Alu*Y_*Gorilla*
chr14	5085273	5085544	*Alu*Y_*Gorilla*
chr17	70469635	70469906	*Alu*Y_*Gorilla*
chr19	4564732	4565003	*Alu*Y_*Gorilla*
chr9	12493812	12494082	*Alu*Y_*Gorilla*
chr13	44228958	44229228	*Alu*Y_*Gorilla*
chr6	17804094	17804361	*Alu*Y_*Gorilla*
chr2B	122589101	122589432	*Alu*Y_*Gorilla*
chr1	132860921	132861250	*Alu*Y_*Gorilla*
chr16	32892735	32893064	*Alu*Y_*Gorilla*
chrX	71963401	71963727	*Alu*Y_*Gorilla*
chrX	53552021	53552346	*Alu*Y_*Gorilla*
chr21	649147	649470	*Alu*Y_*Gorilla*
chr7	141072136	141072458	*Alu*Y_*Gorilla*
chr22	26094248	26094566	*Alu*Y_*Gorilla*
chr11	65236208	65236525	*Alu*Y_*Gorilla*
chr19	6368465	6368781	*Alu*Y_*Gorilla*
chr19	11150362	11150678	*Alu*Y_*Gorilla*
chr8	5420844	5421158	*Alu*Y_*Gorilla*
chr9	76653427	76653740	*Alu*Y_*Gorilla*
chr16	25199336	25199649	*Alu*Y_*Gorilla*
chr3	177244702	177245014	*Alu*Y_*Gorilla*
chrX	98065204	98065515	*Alu*Y_*Gorilla*
chr4	173569814	173570124	*Alu*Y_*Gorilla*
chr7	100116379	100116689	*Alu*Y_*Gorilla*
chr11	29958162	29958472	*Alu*Y_*Gorilla*
chr18	18678555	18678864	*Alu*Y_*Gorilla*
chr4	49087563	49087871	*Alu*Y_*Gorilla*
chr18	9963772	9964080	*Alu*Y_*Gorilla*

chr19	20474251	20474559	*Alu*Y_*Gorilla*
chr1	27835949	27836255	*Alu*Y_*Gorilla*
chr4	199896826	199897132	*Alu*Y_*Gorilla*
chr1	195180519	195180824	*Alu*Y_*Gorilla*
chr5	38872257	38872562	*Alu*Y_*Gorilla*
chr12	130974093	130974398	*Alu*Y_*Gorilla*
chr5	37561715	37562019	*Alu*Y_*Gorilla*
chr5	39745881	39746185	*Alu*Y_*Gorilla*
chr12	57768818	57769122	*Alu*Y_*Gorilla*
chr6	154614819	154615122	*Alu*Y_*Gorilla*
chr7	55059151	55059454	*Alu*Y_*Gorilla*
chr7	79328183	79328486	*Alu*Y_*Gorilla*
chr7	136305097	136305400	*Alu*Y_*Gorilla*
chr14	46472077	46472379	*Alu*Y_*Gorilla*
chr14	65821547	65821849	*Alu*Y_*Gorilla*
chr14	71680953	71681255	*Alu*Y_*Gorilla*
chr15	68805345	68805647	*Alu*Y_*Gorilla*
chr17	2505285	2505587	*Alu*Y_*Gorilla*
chr19	15001868	15002169	*Alu*Y_*Gorilla*
chr8	86680971	86681270	*Alu*Y_*Gorilla*
chr7	72886093	72886391	*Alu*Y_*Gorilla*
chr9	92512320	92512617	*Alu*Y_*Gorilla*
chr11	64580591	64580888	*Alu*Y_*Gorilla*
chr2B	76456808	76457103	*Alu*Y_*Gorilla*
chr19	31965263	31965554	*Alu*Y_*Gorilla*
chr8	119236065	119236339	*Alu*Y_*Gorilla*
chr15	58249144	58249454	*Alu*Y_*Gorilla*
chr8	79988396	79988694	*Alu*Y_*Gorilla*
chr14	84511650	84511912	*Alu*Y_*Gorilla*
chr6	92310242	92310542	*Alu*Y_*Gorilla*
chr12	126042133	126042421	*Alu*Y16_*Gorilla*
chr1	133883001	133883257	*Alu*Y16_*Gorilla*
chr9	121577036	121577327	*Alu*Y16_*Gorilla*
chr3	65263305	65263591	*Alu*Y16_*Gorilla*
chr9	108949687	108949961	*Alu*Y16_*Gorilla*
chr5	147529181	147529478	*Alu*Y16_*Gorilla*
chr12	39433053	39433349	*Alu*Y16_*Gorilla*
chr10	7742039	7742331	*Alu*Y16_*Gorilla*
chr8	130039645	130039934	*Alu*Y16_*Gorilla*
chrX	100692762	100693049	*Alu*Y16_*Gorilla*
chr12	99470325	99470611	*Alu*Y16_*Gorilla*

chrX	92009745	92010030	*Alu*Y16_*Gorilla*
chr11	122178932	122179216	*Alu*Y16_*Gorilla*
chr1	161336658	161336937	*Alu*Y16_*Gorilla*
chr5	20522596	20522874	*Alu*Y16_*Gorilla*
chr2B	89067353	89067629	*Alu*Y16_*Gorilla*
chr1	179240494	179240747	*Alu*Y16_*Gorilla*
chr1	100542814	100543087	*Alu*Y16_*Gorilla*
chr9	97862133	97862429	*Alu*Y16_*Gorilla*
chr1	91601670	91601931	*Alu*Y16_*Gorilla*
chr6	68235572	68235845	*Alu*Y16_*Gorilla*
chrX	97929843	97930141	*Alu*Y16_*Gorilla*
chr8	19352159	19352462	*Alu*Y16_*Gorilla*
chr6	137468192	137468492	*Alu*Y16_*Gorilla*
chrX	89680916	89681211	*Alu*Y16_*Gorilla*
chr1	134841996	134842277	*Alu*Y16_*Gorilla*
chrX	30183451	30183739	*Alu*Y16_*Gorilla*
chr3	23684973	23685265	*Alu*Y16_*Gorilla*
chr5	49439275	49439558	*Alu*Y16_*Gorilla*
chr17	43646197	43646468	*Alu*Y16_*Gorilla*
chr3	81783708	81783993	*Alu*Y16a4_*Gorilla*
chr10	125229580	125229840	*Alu*Y16a4_*Gorilla*
chr20	57457092	57457372	*Alu*Y16a4_*Gorilla*
chr8	95098968	95099246	*Alu*Y16a4_*Gorilla*
chr5	130725484	130725760	*Alu*Y16a4_*Gorilla*
chr7	87000093	87000369	*Alu*Y16a4_*Gorilla*
chr15	30752309	30752581	*Alu*Y16a4_*Gorilla*
chr17	9327125	9327412	*Alu*Y16a4_*Gorilla*
chr15	39227974	39228250	*Alu*Y16a4_*Gorilla*
chr1	18251677	18251959	*Alu*Y16a4_*Gorilla*
chr4	111551744	111552035	*Alu*Y16a4_*Gorilla*
chr2A	33686556	33686846	*Alu*Y16a4_*Gorilla*
chr3	58821998	58822278	*Alu*Y16a4_*Gorilla*
chr3	11931343	11931622	*Alu*Y16a4_*Gorilla*
chr10	14559545	14559837	*Alu*Y16a4_*Gorilla*
chr11	48176812	48177104	*Alu*Y16a4_*Gorilla*
chr19	14055848	14056119	*Alu*Y16a4_*Gorilla*
chrX	74544052	74544324	*Alu*Y16a4_*Gorilla*
chr3	99686701	99686987	*Alu*Ya1_*Gorilla*
chr8	118735963	118736282	*Alu*Ya1_*Gorilla*
chr5	62671042	62671356	*Alu*Ya1_*Gorilla*
chr3	99864231	99864542	*Alu*Ya1_*Gorilla*

chr12	104209566	104209877	*Alu*Ya1_*Gorilla*
chr2A	8536394	8536705	*Alu*Ya1_*Gorilla*
chr4	109685443	109685753	*Alu*Ya1_*Gorilla*
chr6	50912642	50912952	*Alu*Ya1_*Gorilla*
chr4	11353091	11353400	*Alu*Ya1_*Gorilla*
chr8	127010574	127010883	*Alu*Ya1_*Gorilla*
chr8	138485277	138485586	*Alu*Ya1_*Gorilla*
chr13	46248181	46248488	*Alu*Ya1_*Gorilla*
chr15	2087765	2088072	*Alu*Ya1_*Gorilla*
chr2B	14371713	14372020	*Alu*Ya1_*Gorilla*
chr7	63140404	63140710	*Alu*Ya1_*Gorilla*
chr4	168677986	168678291	*Alu*Ya1_*Gorilla*
chr2A	110643863	110644168	*Alu*Ya1_*Gorilla*
chr2B	86106658	86106963	*Alu*Ya1_*Gorilla*
chr1	65716835	65717139	*Alu*Ya1_*Gorilla*
chr1	215607784	215608088	*Alu*Ya1_*Gorilla*
chr3	92502746	92503050	*Alu*Ya1_*Gorilla*
chr3	92513129	92513433	*Alu*Ya1_*Gorilla*
chr18	39124190	39124494	*Alu*Ya1_*Gorilla*
chr5	38566450	38566753	*Alu*Ya1_*Gorilla*
chr2B	92306080	92306383	*Alu*Ya1_*Gorilla*
chr3	158404337	158404639	*Alu*Ya1_*Gorilla*
chr4	9967274	9967576	*Alu*Ya1_*Gorilla*
chr4	38651157	38651459	*Alu*Ya1_*Gorilla*
chr2B	60736036	60736338	*Alu*Ya1_*Gorilla*
chr13	40272689	40272990	*Alu*Ya1_*Gorilla*
chr1	55511997	55512297	*Alu*Ya1_*Gorilla*
chrX	21162933	21163232	*Alu*Ya1_*Gorilla*
chr9	25383135	25383433	*Alu*Ya1_*Gorilla*
chr12	10188146	10188444	*Alu*Ya1_*Gorilla*
chr2B	75803480	75803778	*Alu*Ya1_*Gorilla*
chr13	86881555	86881852	*Alu*Ya1_*Gorilla*
chr12	4365688	4365984	*Alu*Ya1_*Gorilla*
chr12	70905905	70906197	*Alu*Ya1_*Gorilla*
chr12	15338644	15338930	*Alu*Ya1_*Gorilla*
chr20	57687886	57688160	*Alu*Ya1_*Gorilla*
chrX	59846341	59846644	*Alu*Ya1_*Gorilla*
chr1	39184220	39184585	*Alu*Ya1_*Gorilla*
chr7	120399309	120399581	*Alu*Ya1_*Gorilla*
chr5	17649079	17649390	*Alu*Ya1b4_*Gorilla*
chr7	94698875	94699186	*Alu*Ya1b4_*Gorilla*

chr18	190504	190845	*Alu*Ya1b4_*Gorilla*
chr3	82546758	82547074	*Alu*Ya1b4_*Gorilla*
chr8	90568974	90569285	*Alu*Ya1b4_*Gorilla*
chr5	122689054	122689364	*Alu*Ya1b4_*Gorilla*
chrX	43554127	43554437	*Alu*Ya1b4_*Gorilla*
chr2A	79914361	79914671	*Alu*Ya1b4_*Gorilla*
chr14	66142482	66142791	*Alu*Ya1b4_*Gorilla*
chr2A	44402620	44402929	*Alu*Ya1b4_*Gorilla*
chr6	39011570	39011876	*Alu*Ya1b4_*Gorilla*
chr15	21465019	21465325	*Alu*Ya1b4_*Gorilla*
chr1	5236233	5236535	*Alu*Ya1b4_*Gorilla*
chr1	146942421	146942725	*Alu*Yb3a2b2_*Gorilla*
chr4	165286908	165287219	*Alu*Yb3a2b2_*Gorilla*
chr12	95398448	95398744	*Alu*Yb3a2b2_*Gorilla*
chr13	11699647	11699955	*Alu*Yb3a2b2_*Gorilla*
chr6	133656907	133657215	*Alu*Yb3a2b2_*Gorilla*
chr6	147497831	147498149	*Alu*Yb3a2b2_*Gorilla*
chr7	118933215	118933524	*Alu*Yb3a2b2_*Gorilla*
chr14	47878838	47879144	*Alu*Yb3a2b2_*Gorilla*
chr5	17221637	17221935	*Alu*Yb3a2b2_*Gorilla*
chr5	86845305	86845614	*Alu*Yb3a2b2_*Gorilla*
chr9	42299030	42299336	*Alu*Yb3a2b2_*Gorilla*
chr9	82102531	82102837	*Alu*Yb3a2b2_*Gorilla*
chr21	18179353	18179639	*Alu*Yb3a2b2_*Gorilla*
chr2B	40409938	40410241	*Alu*Yb3a2b2_*Gorilla*
chr3	197005090	197005388	*Alu*Yb3a2b2_*Gorilla*
chr10	125223495	125223796	*Alu*Yb3a2b2_*Gorilla*
chr3	145703697	145704000	*Alu*Yb3a2b2_*Gorilla*
chr12	35702878	35703188	*Alu*Yb3a2b2_*Gorilla*
chr5	27108508	27108852	*Alu*Yb3a2b2_*Gorilla*
chr22	15752000	15752328	*Alu*Yb3a2b2_*Gorilla*
chr16	30070450	30070744	*Alu*Yb3a2b2_*Gorilla*
chr4	134326052	134326345	*Alu*Yb3a2b2_*Gorilla*
chr12	45314706	45314998	*Alu*Yb3a2b2_*Gorilla*
chr6	90781042	90781333	*Alu*Yb3a2b2_*Gorilla*
chrX	136541619	136541889	*Alu*Yb3a2b2_*Gorilla*
chr3	92933242	92933549	*Alu*Yb3a2b2a2_*Gorilla*
chr6	17244417	17244718	*Alu*Yb3a2b2a2_*Gorilla*
chr13	19683998	19684293	*Alu*Yb3a2b2a2_*Gorilla*
chr14	49333378	49333687	*Alu*Yb3a2b2a2_*Gorilla*
chr4	67790814	67791122	*Alu*Yb3a2b2a2_*Gorilla*

chr11	8869340	8869674	*Alu*Yb3a2b2a2_*Gorilla*
chr8	50857823	50858130	*Alu*Yb3a2b2a2_*Gorilla*
chr1	129974274	129974580	*Alu*Yb3a2b2a2_*Gorilla*
chr14	68161786	68162092	*Alu*Yb3a2b2a2_*Gorilla*
chr5	87391874	87392179	*Alu*Yb3a2b2a2_*Gorilla*
chr10	99875551	99875859	*Alu*Yb3a2b2a2_*Gorilla*
chr10	138294433	138294734	*Alu*Yb3a2b2a2_*Gorilla*
chr9	44243497	44243790	*Alu*Yb3a2b2a2_*Gorilla*
chr12	66431339	66431622	*Alu*Yb3a2b2a2_*Gorilla*
chr4	136757835	136758112	*Alu*Yb3a2b2a2_*Gorilla*
chr4	74815762	74816072	*Alu*Yb3a2b2a2_*Gorilla*
chr16	18865105	18865410	*Alu*Yb3a2b2a2_*Gorilla*
chr20	12206968	12207270	*Alu*Yc3_*Gorilla*
chr8	72954061	72954356	*Alu*Yc3_*Gorilla*
chr12	84729091	84729400	*Alu*Yc3_*Gorilla*
chr11	76806168	76806477	*Alu*Yc3_*Gorilla*
chr5	140512755	140513063	*Alu*Yc3_*Gorilla*
chr12	2133625	2133933	*Alu*Yc3_*Gorilla*
chr12	4060434	4060737	*Alu*Yc3_*Gorilla*
chr6	121604525	121604839	*Alu*Yc3_*Gorilla*
chr14	82763727	82764022	*Alu*Yc3_*Gorilla*
chr7	82645952	82646258	*Alu*Yc3_*Gorilla*
chr18	75338859	75339161	*Alu*Yc3_*Gorilla*
chr21	30586879	30587188	*Alu*Yc3_*Gorilla*
chr14	24549834	24550117	*Alu*Yc3_*Gorilla*
chr3	88124774	88125084	*Alu*Yc3_*Gorilla*
chr5	114760753	114761063	*Alu*Yc3_*Gorilla*
chr5	163269357	163269666	*Alu*Yc3_*Gorilla*
chr2A	39302891	39303187	*Alu*Yc3_*Gorilla*
chr12	74826419	74826727	*Alu*Yc3_*Gorilla*
chr1	164891241	164891524	*Alu*Yc3_*Gorilla*
chr6	73598308	73598588	*Alu*Yc3_*Gorilla*
chr5	65649825	65650132	*Alu*Yc3_*Gorilla*
chr15	18391568	18391875	*Alu*Yc3_*Gorilla*
chr22	26519767	26520074	*Alu*Yc3_*Gorilla*
chr10	104539891	104540197	*Alu*Yc3_*Gorilla*
chr8	121990557	121990838	*Alu*Yc3_*Gorilla*
chr18	26275459	26275764	*Alu*Yc3_*Gorilla*
chr2B	22378052	22378357	*Alu*Yc3_*Gorilla*
chr2A	90579663	90579949	*Alu*Yc3_*Gorilla*
chr4	79882888	79883187	*Alu*Yc3_*Gorilla*

chr13	12192333	12192632	*Alu*Yc3_*Gorilla*
chr10	70451743	70452041	*AluYc3_Gorilla*
chr3	29082070	29082367	*AluYc3_Gorilla*
chr8	130394198	130394495	*Alu*Yc3_*Gorilla*
chr9	108314950	108315247	*AluYc3_Gorilla*
chr4	160866757	160867051	*Alu*Yc3_*Gorilla*
chr13	20968400	20968710	*Alu*Yc3_*Gorilla*
chr3	36619368	36619677	*Alu*Yc3_*Gorilla*
chr3	139407803	139408112	*Alu*Yc3_*Gorilla*
chr14	71473665	71473956	*Alu*Yc3_*Gorilla*
chr8	68537612	68537935	*Alu*Yc3_*Gorilla*
chr1	145355962	145356281	*Alu*Yc3_*Gorilla*
chrX	93615031	93615350	*Alu*Yc3_*Gorilla*
chr4	183271106	183271423	*Alu*Yc3_*Gorilla*
chr7	10943240	10943556	*Alu*Yc3_*Gorilla*
chr13	19692297	19692610	*Alu*Yc3_*Gorilla*
chr18	29325002	29325312	*Alu*Yc3_*Gorilla*
chr22	27699642	27699952	*Alu*Yc3_*Gorilla*
chr13	63645783	63646091	*Alu*Yc3_*Gorilla*
chr6	62291463	62291767	*Alu*Yc3_*Gorilla*
chr2B	116718944	116719246	*Alu*Yc3_*Gorilla*
chr1	91318108	91318409	*Alu*Yc3_*Gorilla*
chr5	34965062	34965363	*Alu*Yc3_*Gorilla*
chr8	140890854	140891155	*Alu*Yc3_*Gorilla*
chr10	30488110	30488411	*Alu*Yc3_*Gorilla*
chr11	114136513	114136814	*Alu*Yc3_*Gorilla*
chr17	28341016	28341310	*Alu*Yc3_*Gorilla*
chr21	12106659	12106953	*Alu*Yc3_*Gorilla*
chr2B	59031885	59032179	*Alu*Yc3_*Gorilla*
chr3	169028448	169028741	*Alu*Yc3_*Gorilla*
chr4	183234380	183234673	*Alu*Yc3_*Gorilla*
chr4	194738507	194738800	*Alu*Yc3_*Gorilla*
chr4	139317064	139317348	*Alu*Yc3_*Gorilla*
chr2B	36927702	36927981	*Alu*Yc3_*Gorilla*
chr10	95889075	95889353	*Alu*Yc3_*Gorilla*
chr11	36281034	36281342	*Alu*Yc3_*Gorilla*
chr3	112245755	112246060	*Alu*Yc3_*Gorilla*
chr4	200369692	200369994	*Alu*Yc3_*Gorilla*
chr14	41890377	41890679	*Alu*Yc3_*Gorilla*
chr15	65411618	65411920	*Alu*Yc3_*Gorilla*
chr5	120001963	120002251	*Alu*Yc5a3_*Gorilla*

chr6	47817987	47818282	*Alu*Yc5a3_*Gorilla*
chr10	98567511	98567805	*Alu*Yc5a3_*Gorilla*
chr11	103850307	103850588	*Alu*Yc5a3_*Gorilla*
chr11	15479633	15479922	*Alu*Yc5a3_*Gorilla*
chr12	92826801	92827092	*Alu*Yc5a3_*Gorilla*
chr17	35215720	35216008	*Alu*Yc5a3_*Gorilla*
chr17	40335715	40336009	*Alu*Yc5a3_*Gorilla*
chr21	2490263	2490549	*Alu*Yc5a3_*Gorilla*
chr2B	128129319	128129615	*Alu*Yc5a3_*Gorilla*
chr9	86656201	86656488	*Alu*Yc5a3_*Gorilla*
chr9	45471525	45471822	*Alu*Yc5a3_*Gorilla*
chr9	16782260	16782557	*Alu*Yc5a3_*Gorilla*
chr5	62650413	62650709	*Alu*Yc5a3_*Gorilla*
chr5	41399009	41399304	*Alu*Yc5a3_*Gorilla*
chr2B	50079082	50079378	*Alu*Yc5a3_*Gorilla*
chr1	76508291	76508568	*Alu*Yc5a3_*Gorilla*
chr20	62601356	62601637	*Alu*Yc5a3_*Gorilla*
chr18	63462410	63462703	*Alu*Yc5a3_*Gorilla*
chr3	2457123	2457394	*Alu*Yc5a3_*Gorilla*
chr14	65919448	65919741	*Alu*Yc5a3_*Gorilla*
chr7	83827494	83827766	*Alu*Yc5a3_*Gorilla*
chr7	116606240	116606500	*Alu*Yc5a3_*Gorilla*
chr5	151820936	151821233	*Alu*Yc5a3_*Gorilla*
chr22	25391880	25392132	*Alu*Yc5a3_*Gorilla*
chr10	144554349	144554646	*Alu*Yc5a3_*Gorilla*
chr15	40503129	40503381	*Alu*Yc5a3_*Gorilla*
chr5	44287466	44287761	*Alu*Yc5a3_*Gorilla*
chr9	16897286	16897580	*Alu*Yc5a3_*Gorilla*
chr10	28464590	28464871	*Alu*Yc5a3_*Gorilla*
chr10	120577207	120577493	*Alu*Yc5a3_*Gorilla*
chr14	71365071	71365323	*Alu*Yc5a3_*Gorilla*
chr2B	57870770	57871042	*Alu*Yc5a3_*Gorilla*
chr2B	83106386	83106708	*Alu*Yc5a3_*Gorilla*
chr1	53945851	53946126	*Alu*Yc5a3_*Gorilla*
chr8	49832649	49832948	*Alu*Yc5a3_*Gorilla*
chr1	97044721	97044997	*Alu*Yc5a3_*Gorilla*
chr18	22451258	22451551	*Alu*Yc5a3_*Gorilla*
chr17	21117755	21118013	*Alu*Yc5a3_*Gorilla*
chr11	96651578	96651870	*Alu*Yc5a3_*Gorilla*
chr20	11304297	11304588	*Alu*Yc5a3_*Gorilla*
chr3	155107893	155108183	*Alu*Yc5a3_*Gorilla*

chr9	63804483	63804773	*Alu*Yc5a3_*Gorilla*
chr11	68643588	68643877	*Alu*Yc5a3_*Gorilla*
chr18	59386448	59386737	*Alu*Yc5a3_*Gorilla*
chr17	50441849	50442137	*Alu*Yc5a3_*Gorilla*
chr4	123122771	123123058	*Alu*Yc5a3_*Gorilla*
chr5	29011995	29012281	*Alu*Yc5a3_*Gorilla*
chr10	30733839	30734124	*Alu*Yc5a3_*Gorilla*
chr15	53840856	53841141	*Alu*Yc5a3_*Gorilla*
chr1	105322391	105322685	*Alu*Yc5a3_*Gorilla*
chr5	1325224	1325506	*Alu*Yc5a3_*Gorilla*
chr9	76225977	76226255	*Alu*Yc5a3_*Gorilla*
chr4	183033924	183034194	*Alu*Yc5a3_*Gorilla*
chr3	9659411	9659678	*Alu*Yc5a3_*Gorilla*
chr4	142945403	142945701	*Alu*Yc5b2_*Gorilla*
chr4	19388645	19388940	*Alu*Yc5b2_*Gorilla*
chr4	151202658	151202951	*Alu*Yc5b2_*Gorilla*
chr4	25523473	25523755	*Alu*Yc5b2_*Gorilla*
chr5	76992527	76992823	*Alu*Yc5b2_*Gorilla*
chr6	126708625	126708905	*Alu*Yc5b2_*Gorilla*
chr13	16526653	16526941	*Alu*Yc5b2_*Gorilla*
chr13	93817208	93817499	*Alu*Yc5b2_*Gorilla*
chr14	33437804	33438102	*Alu*Yc5b2_*Gorilla*
chr15	72961038	72961324	*Alu*Yc5b2_*Gorilla*
chr16	12976677	12976972	*Alu*Yc5b2_*Gorilla*
chr17	78960783	78961068	*Alu*Yc5b2_*Gorilla*
chr19	2205069	2205354	*Alu*Yc5b2_*Gorilla*
chr2A	43095346	43095645	*Alu*Yc5b2_*Gorilla*
chr18	34634842	34635136	*Alu*Yc5b2_*Gorilla*
chr4	162666713	162667003	*Alu*Yc5b2_*Gorilla*
chr9	17217290	17217586	*Alu*Yc5b2_*Gorilla*
chr12	20819750	20820038	*Alu*Yc5b2_*Gorilla*
chr7	133914713	133915011	*Alu*Yc5b2_*Gorilla*
chr5	13038133	13038393	*Alu*Yc5b2_*Gorilla*
chr1	173185434	173185740	*Alu*Yc5b2_*Gorilla*
chr2B	72370599	72370899	*Alu*Yc5b2_*Gorilla*
chr3	63983947	63984246	*Alu*Yc5b2_*Gorilla*
chr9	7784599	7784897	*Alu*Yc5b2_*Gorilla*
chr9	17925753	17926051	*Alu*Yc5b2_*Gorilla*
chr9	27967228	27967526	*Alu*Yc5b2_*Gorilla*
chr17	64499549	64499847	*Alu*Yc5b2_*Gorilla*
chr5	53464920	53465216	*Alu*Yc5b2_*Gorilla*

chr3	107821391	107821685	*Alu*Yc5b2_*Gorilla*
chr1	63944229	63944509	*Alu*Yc5b2_*Gorilla*
chr8	102526261	102526526	*Alu*Yc5b2_*Gorilla*
chr14	28059186	28059487	*Alu*Yc5b2_*Gorilla*
chr8	498312	498607	*Alu*Yc5b2_*Gorilla*
chr16	62220651	62220945	*Alu*Yc5b2_*Gorilla*
chrX	138047030	138047323	*Alu*Yc5b2_*Gorilla*
chr10	116109563	116109856	*Alu*Yc5b2_*Gorilla*
chr1	121562382	121562674	*Alu*Yc5b2_*Gorilla*
chr8	57303298	57303590	*Alu*Yc5b2_*Gorilla*
chr2B	114014027	114014319	*Alu*Yc5b2_*Gorilla*
chr1	96324690	96324981	*Alu*Yc5b2_*Gorilla*
chr3	176665702	176665993	*Alu*Yc5b2_*Gorilla*
chr8	128794443	128794733	*Alu*Yc5b2_*Gorilla*
chr13	52209308	52209597	*Alu*Yc5b2_*Gorilla*
chr2A	95414819	95415106	*Alu*Yc5b2_*Gorilla*
chr17	34116946	34117231	*Alu*Yc5b2_*Gorilla*
chr12	39574263	39574547	*Alu*Yc5b2_*Gorilla*

Table S.6: Estimated age of gorilla-specific *Alu* subfamilies based on BEAST analysis. The BEAST program was run on each gorilla-specific subfamily with a baseline divergence age of 7 mya to determine the age of the subfamilies, the most likely progenitor or ancestral element, and the % divergence from the consensus sequence of the ancestral subfamily.

Subfamily	Master Gene	% Divergence from Consensus	Copy Number	Age of MRCA (in my)
*Alu*Y_*Gorilla*	AluY	0.0	759	6.5
*Alu*Yc5a3_*Gorilla*	AluYc	1.9	55	6.52
*Alu*Yc3_*Gorilla*	AluY	1.8	69	6.54
*Alu*Ya1_*Gorilla*	AluY	2.5	43	6.53
*Alu*Y16_*Gorilla*	AluY	1.9	30	6.55
*Alu*Yb3a2b2_*Gorilla*	AluYk4	0.7	25	6.51
*Alu*Yc5b2_*Gorilla*	AluYc	2.2	46	6.54
*Alu*Ya1b4_*Gorilla*	AluY	2.8	13	6.51
*Alu*Yb3a2b2a2_*Gorilla*	AluYk4	1.8	17	6.71
*Alu*Y16a4_*Gorilla*	AluY	3.0	18	6.54

APPENDIX B:
LETTERS OF PERMISSION

***PLoS One* Letter of Permission (For Chapter 2)**

Request:

Hello,

I published an article in PLoS ONE back in August, and I plan on using the publication as a portion of my Ph.D. dissertation. My institution, LSU, requires formal permission from journals for this. Is it possible for you to write me a short note granting that permission? I know with PLoS authors retain copyright to their work, but I think I still need some kind of dispensation from your journal that I can put in the dissertation stating I have said permission.

The article is here: PLoS ONE 7(8): e44035. doi:10.1371/journal.pone.0044035

Thanks!

Adam

--
Adam Trent McLain
Ph.D. Candidate
Board of Regents Fellow

Batzer Lab (Laboratory of Comparative Genomics)
Department of Biological Sciences
A653 Life Sciences Building
Louisiana State University
Baton Rouge, LA 70803 USA

Response:

Dear Adam,

Thank you for your enquiry and thanks for publishing with PLOS.

We are an open access publisher and as such everything we publish is freely available online throughout the world, for you to read, download, copy, distribute, and use (with attribution) any way you wish.

No permission required.

For information on the Creative Commons Attribution License please follow this link
http://www.plos.org/about/open-access/license/

Hope this is helpful.

Kind regards,

Colin Dixon

PLOS

16 November 2012

THE VITA

Adam Trent McLain was born in 1980, the son of Michael, a business executive and engineer, and Victoria, a college professor. He lived at varying times in Florida, Georgia, Indiana, Texas, and Virginia. He graduated from Broad Run High School in Ashburn, Virginia, in 1998. He graduated from Virginia Polytechnic Institute and State University ("Virginia Tech") in May, 2002, with a Bachelor of Arts degree in History. The focus of his undergraduate study was the classical Greek and Roman world. From 2002-2007 Adam worked at various times as a musician, bartender, waiter, journalist, file clerk, house painter, handyman, and camp counselor in Atlanta, Georgia, and New York City. Adam graduated from Hunter College of the City University of New York in May, 2009, with a Master of Arts degree in Anthropology. He began working toward his Doctor of Philosophy degree in biology in the laboratory of Dr. Mark Batzer in August, 2009, and expects to graduate with that degree in December, 2013. He is married to Rebecca McLain, a native of Cuba, New York, and is the proud parent of four cats named Freyja, Loki, Ella, and Persephone.